

ENVISIONING THE NATIONAL POSTSECONDARY DATA INFRASTRUCTURE IN THE 21ST CENTURY

# Leveraging What We Already Know: Linking Federal Data Systems

DAVID A. BERGERON  
CENTER FOR AMERICAN PROGRESS

MAY 2016

**David A. Bergeron** is a senior fellow for postsecondary education at the Center for American Progress.

This paper is part of the larger series *Envisioning the National Postsecondary Data Infrastructure in the 21st Century*. In August 2015, the Institute for Higher Education Policy (IHEP) first convened a working group of national postsecondary data experts to discuss ways to move forward a set of emerging options for improving the quality of the data infrastructure in order to inform state and federal policy conversations. The resulting paper series presents targeted recommendations, with explicit attention to related technical, resource, and policy considerations. This paper is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the author(s) and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or the Institute for Higher Education Policy.

# 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.<sup>1</sup> 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,<sup>2</sup> 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.

# Leveraging What We Already Know:

## Linking Federal Data Systems

### Overview of the Role of Federal Data Resources in Policy Discussions

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 the needs of individual citizens. Information on individual citizens is already subject to laws and regulations to ensure its safe and secure use, and it can be used effectively while still protecting people's privacy.

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:

- ▶ the student's repayment status, derived from the National Student Loan Data System (NSLDS);
- ▶ whether the student was the dependent or spouse of a service member who died in the line of duty, derived from the U.S. Department of Defense (DoD);
- ▶ whether a judge has placed restrictions on the receipt of federal benefits, derived from the U.S. Department of Justice (DoJ);
- ▶ whether the student is incarcerated or an undocumented alien, derived from the Social Security Administration (SSA);
- ▶ whether the student is a veteran and, therefore, independent for purposes of determining aid eligibility, derived from the U.S. Department of Veterans Affairs (VA); and
- ▶ whether the student is an eligible noncitizen or has appropriately registered for the draft with the selective service, derived from the U.S. Department of Homeland Security (DHS).

In their paper as part of this series *Using—and Improving—FSA Data Systems to Support Policy Analysis*, Matt Soldner and Colleen Campbell address how existing data resources held by Federal Student Aid could be leveraged to enhance our understanding of how effectively the higher education

system is serving aid recipients.<sup>3</sup> In “Building a Student-Level Data System,” Ben Miller makes the case that these data could be augmented with a student unit record system to understand the effectiveness of the system more broadly.<sup>4</sup> But either of these approaches can be enhanced by viewing other federal data resources, collected as a consequence of carrying out other basic government functions, as a significant part of the broader postsecondary education data ecosystem.

Among the basic functions of government is the generation of information about individual citizens that ranges from recruiting young adults for the various branches of the military to supporting the collection of child support payments from noncustodial parents. In some cases, the information collected is specific to postsecondary enrollment and attainment—like the information collected by military recruiters, which is particularly focused on specific types of education and training to target graduates for recruitment into military service. Other information, like the new jobs of recent graduates collected in the National Directory of New Hires (NDNH) and earnings for graduates available from SSA, can inform us about the outcomes of postsecondary education. These data, and all personally identifiable data held by the federal government, are already subject to laws and regulations to ensure their safe and secure use.

Taken together, the various pieces of information about individuals that the government holds beyond ED (see **Figure 1**) could help to better inform federal, state, and private postsecondary education policies; improve services to individuals pursuing a postsecondary education; and reduce the cost of responding to numerous requests by federal and state agencies for information about students and graduates.

Today, students and families are unable to get answers to very basic questions about the outcomes of investing time, energy, and money in postsecondary education:

- ▶ How likely is it that a student like me graduates from this or any institution of higher education on time?
- ▶ 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?
- ▶ How much will I likely earn if I attend this institution and major in this field?

- ▶ 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?

States and institutions are similarly unable to know what happens to their former students who move to another state to transfer to another school or enter the world of work. With the development of state longitudinal data systems, a significant part of the data vacuum on the impact of postsecondary education—at least for students who attend public colleges and universities and remain in the same state after graduation—has been substantially addressed. But states and institutions are still ill equipped to develop strategies that place appropriate weight on outcomes for students who enroll in private institutions of higher education or who leave the state where they attended college.

Finally, the federal government has only a limited set of information on outcomes on which to assess the impact of the significant investment in postsecondary institutions and the students they enroll. Historically, the only measure available to assess the impact of this investment has been default rates for cohorts of students two or three years after they leave postsecondary education. Such a limited dimensional measure has misled federal policymakers into believing that if they managed student loan defaults, they did not need to worry about graduation rates, job placement rates, or the cumulative impact of student loan debt.

The information the federal government collects on individuals is kept in silos. In part, these silos reflect the roles of the federal agencies that collect the data. For example, the U.S. Department of Health and Human Services (HHS) collects information on newly hired individuals to aid in the enforcement of child support orders, while the various branches of the military collect information on students enrolled in postsecondary education to aid in the recruitment of service members with specific knowledge, skills, and abilities.

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 it provides to citizens, and help potential postsecondary students make better choices about their education.

All personally identifiable data maintained by the federal government are governed by the provisions of the Privacy Act of 1974.<sup>5</sup> Significantly, 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 certain statutory exceptions. The statutory exceptions include such uses as statistical purposes by the Census Bureau and

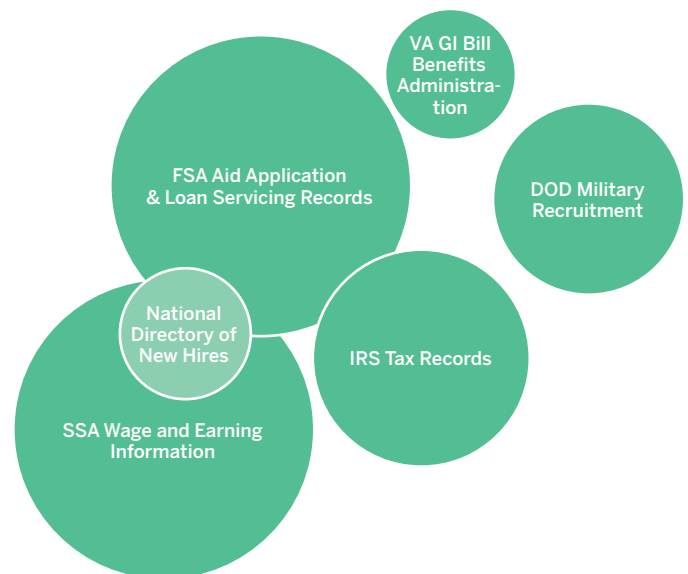
the Bureau of Labor Statistics, routine uses within a U.S. government agency, archival purposes “as a record which has sufficient historical or other value to warrant its continued preservation by the United States Government,” law enforcement purposes, congressional investigations, and other administrative purposes. While it is possible that much of the recommended data sharing could be accomplished under one of the existing statutory exceptions, it would be preferable to have an explicit exception to authorize the exchange of data on postsecondary students. The issues of data security and privacy are addressed at length in Joanna Lyn Grama’s paper, “Understanding Information Security and Privacy in Postsecondary Education Data Systems.”<sup>6</sup>

### Federal Data Resources that Could Provide Greater Transparency into the Student Experience

Much of the information that the federal government holds on individuals would help us better understand the experience of students while they are enrolled in postsecondary education. For example, the tuition statements IRS receives from nearly all institutions of higher education that could tell policymakers a tremendous amount about the unmet need of college students, when combined with information from DoD and VA, would help us understand how veterans progress through and complete a postsecondary education. These issues cannot be fully understood today because of the siloes in which information is held by the federal government for both policy and bureaucratic reasons. Leveraging these data resources would increase our understanding of the challenges facing college students without adding to the burden imposed on postsecondary education institutions.

**FIGURE 1: ROLE OF LINKING FEDERAL DATA SYSTEMS IN DATA POLICY DISCUSSIONS**

Current State: National Postsecondary Education Data Infrastructure





## Tax Records

IRS collects a tuition statement (Form 1098-T)<sup>7</sup> for each student enrolled in postsecondary education annually, but few would view this information as a part of the national postsecondary education data ecosystem—even though this potential data resource collects the most comprehensive data on tuition and fee revenues available—because nearly all postsecondary institutions must submit information to IRS on all students who pay tuition and fees. (Only those institutions that do not meet the specific section of the tax code’s statutory definition of an institution of higher education, including institutions that are predominantly or entirely online, are not required to report. This is because students enrolled at these institutions are not eligible to receive an American Opportunity, HOPE, or Lifetime Learning tax credit.<sup>8</sup>) Failing to view the IRS 1098-T and other information held by the federal government as an element of the national postsecondary education data ecosystem has significant consequences. The exclusion of these data resources from the national ecosystem results in the federal government collecting information from institutions of higher education in summary form—net price by family income level<sup>9</sup>—that it already has available to it at the student level. Even more burdensome and potentially troublesome is the requirement that institutions<sup>10</sup> manipulate these data to build a consumer disclosure tool—net price calculators<sup>11</sup>—that could be more easily and consistently built for each institution based on the information contained in the 1098-Ts that institutions already submit to IRS. To create an accurate net price for students who do not receive federal student aid, some additional information from the taxpayer’s tax return could be required. If needed, this information could be obtained in such a way as to not disclose any personally identifiable information to the institution. Most simply, this could be accomplished by releasing back to institutions only appropriately aggregated data, to avoid inadvertent release of personally identifiable information.

Each year, institutions of higher education must submit a 1098-T to IRS for each student enrolled. In addition to identifying information for both the student and the taxpayer, the data collected for each student include the following:

- ▶ payments received for tuition and related expenses,
- ▶ amounts billed for qualified tuition and related expenses,
- ▶ adjustments made for a prior year,
- ▶ scholarships and grants,
- ▶ adjustments to scholarships or grants for a prior year,
- ▶ amount of any contract reimbursement or refund, and
- ▶ whether the student was enrolled at least half time, or was a graduate student.<sup>12</sup>

Some progress has been made in viewing IRS data as an element in the national postsecondary education data ecosystem. When ED developed the most recent College Score-

card, it engaged with the U.S. Department of the Treasury to use administrative tax records to gain insight into the labor market outcomes of federally aided students who attended institutions of higher education. By linking data from the National Student Loan Data System and administrative tax records, it was possible to produce aggregated estimates of institution-level statistics, such as the mean and median of the earnings distribution of federally aided students in a particular year, along with other metrics. By using administrative tax records, it was possible to include wages and deferred compensation from all W-2 forms received for each individual as well as self-employment earnings from Schedule SE. As a result, the coverage and accuracy of the institution-level statistics is high for both employees and the self-employed and, unlike statewide systems that rely on state unemployment insurance administrative data, offers national coverage.<sup>13</sup>

If viewed as an element of the national postsecondary education data ecosystem, IRS could also offer insights into other public policy questions such as how much taxpayers with various levels of educational attainment pay in income taxes and how postsecondary education affects family structures.

## Military Recruitment

Under a provision of law commonly referred to as the Solomon Amendment, DoD—through recruiters assigned to each of 12 eligible units of the five branches of the U.S. armed forces—collects information on every student age 17 and older enrolled in postsecondary education.<sup>14</sup> An exception is made for students enrolled in an institution of higher education that has a longstanding policy of pacifism based on historical religious affiliation. A few religious denominations with which institutions of higher education are associated—Friends and Seventh Day Adventists, for example—have historically advocated pacifism. The number of institutions associated with those denominations

### Military units that can request recruiting information:

- Army
- Army Reserve
- Army National Guard
- Navy
- Navy Reserve
- Marine Corps
- Marine Corps Reserve
- Air Force
- Air Force Reserve
- Air Force National Guard
- Coast Guard
- Coast Guard Reserve

is very small; only seven institutions of higher education are affiliated with the Friends and 14 with Seventh Day Adventists.<sup>15</sup> An institution may also be able to avoid providing the required information if it can certify that it prohibits all employers from recruiting on campus, that it permits employers to recruit on campus only if students express interest, or that too few students are interested to warrant accommodating military recruiters.



Under the Solomon Amendment, institutions are required to provide to military recruiters the name, address, telephone number, age and date of birth, place of birth, level of education, academic major, and degrees received for all enrolled students. Since the Solomon Amendment serves as an extension of the Family Education Rights and Privacy Act (FERPA), the commonly held definition of a student under FERPA would apply: an individual who is enrolled in and actually attends an educational institution whether in person or by videoconference, satellite, Internet, other electronic information and telecommunications technologies, or correspondence. Individuals who are not enrolled but audit classes, or who are accepted to an educational institution but do not attend any classes, are not “students” for purposes of FERPA.<sup>16</sup> Recruiters may not ask for the student’s Social Security Number, race, ethnicity, nationality, religious affiliation, grades, or default status on loans.

The collection of the information required by the Solomon Amendment is decentralized, and the information may be solicited from institutions in many different ways. Most commonly, a recruiter from a local station will contact the institution of higher education, but some may collect the information for a region. As a result, coverage for this collection is likely not uniform. Despite the lack of uniformity in collecting this information, an institution that fails to comply loses access to federal contracts and grants from the U.S. Departments of Education, Defense, Health and Human Services, Labor, Transportation, and Homeland Security; Central Intelligence Agency; and National Nuclear Security Administration.

No provision is made in the Solomon Amendment for use of the information collected for analytic purposes, and the restrictions on what can be collected would severely limit its analytic utility. Even at the local recruiting station, the information may or may not be stored in an automated way. The primary problem with the information collected under the Solomon Amendment is that it is collected and maintained, for the most part, by individual recruiting stations. The Marine Corps alone has 48 recruiting stations, 615 recruiting substations, and 74 officer selection stations.<sup>17</sup> By contrast, the Army recruiting command has 948 recruiting centers nationwide organized under 260 recruiting companies in seven enlisted recruiting brigades, while the Navy has more than 1,500 recruiting stations. Institutions do not have to respond to multiple requests from the same unit and branch of the service but still can receive requests from up to 12 units of the five branches of military service annually. This makes the system difficult and costly to navigate for both the government and institutions. It also makes it difficult to assess how good any branch of the military’s coverage actually is. For example, the Navy indicates that it obtains information from 5,240 two-year and four-year colleges and universities,

which suggests that, at a minimum, it is not collecting from all institutions eligible under Title IV of the Higher Education Act.

More significantly, such an approach makes it impossible for this significant data resource to be used effectively as an analytic tool even to determine, for example, whether the investments made by DoD in tuition assistance for service members and their families are resulting in timely degree attainment.

Similarly, failing to view the information that military recruiters collect as a part of the national postsecondary education data system has had significant consequences. Today, this information is collected by military recruiters in a very manual way, with each recruiting station obtaining the information on paper directly from the registrar of each campus. This uncoordinated collection of personal information about students enrolled in postsecondary education places an unnecessary burden on postsecondary education institutions and results in inconsistent opportunities for students to enter the armed forces based on which institution of higher education they graduated from. Each branch of the service, both regular and reserve, is able to obtain this personal student information—name, address, telephone number, age and date of birth, place of birth, level of education, academic major, and degrees received—and institutions are required, under penalty of loss of all federal funds, to provide this information each time it is requested by the different branches of the armed forces.<sup>18</sup> This haphazard approach increases the burden on institutions and the cost to the federal government while decreasing the information’s utility as a resource for understanding how students are progressing through the postsecondary education system. Automating and routinizing this data collection would make it much easier for military recruiters to focus on their primary job responsibilities. Unlocking these data to calculate college completion rates for all students, but most significantly veterans, could fill a great void in our understanding the difficulties they face in completing their postsecondary education.

### **GI Bill Education Benefits for Veterans**

Three years ago, VA began to automate the process of certifying a veteran’s eligibility for the Post-9/11 GI Bill using a system named VA-ONCE. Significantly, the goal of this effort was not to reduce reporting burden on the part of the institution that was certifying the eligibility for benefits, but to reduce the time between when a school submits enrollment information and when the veteran receives benefits. As a result, the VA-ONCE system simply automated an existing paper process. VA-ONCE was not developed to extract data from existing enrollment management systems, nor does the system exploit any of the capacities that exist within the Title IV student aid delivery system. Before entering information about eligible veterans, the system collects information about each program of study the institution offers—the

course name, abbreviation, objective, and VA facilities code at which the program is offered. This information is different for degree programs at institutions of higher education, all programs at nondegree schools, nondegree programs at institutions of higher education, and flight training.<sup>19</sup> Also, the codes for the courses of study at higher education institutions bear no relationship to the Classification of Instructional Program codes developed by the National Center for Education Statistics and commonly used by federal agencies.

In addition to the information collected on each program, the institution reports on each enrolled student who receives Post-9/11 GI Bill benefits. Specifically, the institution reports the student's name, Social Security Number, student ID number, payee number, phone number, e-mail address, benefit chapter number, program, training type, facility code at which the student is or will be attending, period of enrollment by term, number of residential credits taken during the term, number of distance credits taken during the term, number of remedial credits taken during the term, number of clock hours taken per week (if applicable), tuition and fees charged per term, amount the institution has agreed to provide above the highest in-state rate, and charges above the in-state rate. If the student changes enrollment status after he or she has been certified, the institution must report adjustments, amendments, or terminations including the date on which the change in status occurred, the result of the change, and a reason if the student has provided one.

The information that VA collects is critical to the timely award of veterans' benefits to eligible students. However, the lack of integration into the broader postsecondary education data ecosystem results in significant inefficiencies. Consider the benefit of having access to the recruiting information potentially available to DoD under the Solomon Amendment and to data collected by VA. VA provides educational benefits to approximately 1.2 million students each year.<sup>20</sup> Until recently, VA has not attempted to calculate persistence or graduation rates for GI Bill recipients because such information was unavailable unless the student veteran continued to receive benefits throughout her or his education. When VA decided to calculate these rates in 2014, it turned to the National Student Clearinghouse<sup>21</sup> to do so even though DoD recruiting arms had access to a potentially more comprehensive dataset. In the future, VA could be able to exchange data with DoD to calculate the education attainment for beneficiaries. The data exchange could also be used to ensure that federal student aid recipients who completed a degree program were appropriately identified as a graduate in the National Student Loan Data System.

Matching records between VA, DoD, and ED would be somewhat difficult without a Social Security Number, which DoD is currently prohibited from collecting. However, VA and ED

already have other key information that can be used to match: the student's name, address, telephone number, age and date of birth, place of birth, institution of higher education, level of education, and academic major.

Building the ecosystem to allow VA and ED to share information with DoD would have an ancillary benefit in recruiting. For example, a military recruiter could highlight the loan repayment benefits available for military service to those graduating with significant amounts of student loan debt. Additionally, military recruiters could adjust their approach to recruiting if they knew that a graduate had previously served in the military. The reasons that a veteran might wish to reenlist are likely different from a recent college graduate who has not previously served in the military. However, military recruiters cannot obtain information about prior military service from the institution the graduate attended, limiting the recruiter's ability to adjust to the graduate's previous experiences. Building the ecosystem to share this information with ED would also have benefits. ED could use this information to move a borrower into repayment or to calculate an attainment rate for aid recipients who changed institutions prior to completing their studies.

Providing information to veterans on the outcomes at specific institutions for other former service members could significantly improve the decisions they make. However, because participation in the National Student Clearinghouse is voluntary and subject to disclosure agreements, it is unlikely that the Clearinghouse will be able to release graduation rates and other outcome metrics at an institutional level broadly. However, these data are still extraordinarily valuable. Like all federal agencies, the Clearinghouse is subject to Government Performance and Results Act requirements.<sup>22</sup> When VA needed to obtain outcome information on veterans, it turned to the Clearinghouse because it could analyze the data to calculate the success of those receiving veterans' benefits. Further analysis of this information could reveal the role that GI Bill benefits play in degree attainment when combined with data from a consolidated set of Solomon Amendment data available for recruiting by DoD without the additional expense or the significant limitations on use associated with the Clearinghouse.<sup>23</sup>

**The Government Performance and Results Act (Pub. L. 103-62) requires all government agencies to set goals, measure results, and report on their progress and reaching their goals. The Office of Management and Budget annually reviews and reports on each agency's success in meeting its performance goals. Working toward compliance with the act has driven much of the federal government's efforts to measure outcomes and assess results.**

As important as the data resources described above are to understanding what happens to students while they are enrolled in higher education, a number of federal data resources outside ED that are not considered a part of the national postsecondary data ecosystem could also provide valuable insight into what happens to students after they leave postsecondary education. A large number of federal data systems could inform some aspect of postsecondary education. For example, gaining insight into the labor market outcomes of graduates could help inform state investment decisions in higher education and suggest areas where loan repayment programs might be most effective for private employers wishing to recruit individuals with specialized knowledge and skill. For the purposes of this discussion, the following section focuses on systems that have the greatest perceived potential to improve the nation's understanding of the postsecondary education system.

### Federal Data Resources that Could Provide Transparency into Post-Enrollment Experience

#### National Directory of New Hires

HHS collects information from federal, state, and private sources on employment and wages, SSA benefits, military entitlements, financial accounts and balances, insurance claims, unemployment, and child support case data in the NDNH.<sup>24</sup> This information is used to help state child support programs establish and enforce fair and equitable child support orders. Each state is required to set up and maintain a state directory of new hires. Each state directory of new hires is a registry of all newly hired employees in a state, which, in turn, feeds into the NDNH.

The NDNH was implemented through a collaborative process that included the federal government's Office of Child Support Enforcement, SSA, Department of Labor, and IRS and the states' Child Support Enforcement Agencies, State Employment Security Agencies, and courts, along with employer associations.<sup>25</sup> The NDNH is augmented with data from every federal agency, which feeds new hire and quarterly wage information into the NDNH directly. The very limited use of this information is specified in law<sup>26</sup> and HHS is appropriately protective of this information so that it can meet its purpose of improving the enforcement of child support orders. Access is limited to state welfare, child and family services, foster care and adoption assistance, workforce, and supplemental nutrition assistance program agencies. At the national level, access is limited to the following:

- ▶ the Commissioner of SSA;
- ▶ the U.S. Secretary of Treasury, to administer the earned income tax credit, verify a claim of employment, and collect debts owed the federal government; and

- ▶ the U.S. Secretary of Department of Housing and Urban Development (HUD), to verify the employment and income of individuals participating in housing programs and to conduct analyses of the employment and income reporting of those individuals.

Under Section 453(j)(6) of the Social Security Act of 1935, as amended, the Secretary of Education has access to these data to improve the collection of student loan debts for defaulted borrowers only.<sup>27</sup> The effectiveness of this access as a collection tool has been mixed because of limitations on ED's ability to share the data with business partners including guaranty agencies that help collect defaulted federal family education loans before they are assigned to ED.

Also, curiously, the Secretary of HHS does not seem to be authorized to use the information in the NDNH to conduct analyses, for example, of the impact of postsecondary education on individuals who are receiving support under Temporary Assistance for Needy Families (TANF), although researchers may obtain the data without personal identifiers for research purposes the Secretary of HHS finds to be likely to contribute to achieving the purposes of Part A or Part D of the Social Security Act.<sup>28</sup> Also notable is the lack of authority for the Secretary of Education to conduct analyses of the employment and income of financial aid recipients as the Secretary of HUD can for individuals participating in certain housing programs.

#### Wage and Earnings Information

In its role as administrator of two large national entitlement programs, SSA collects a wealth of data about the employment and earnings of those working in the United States. Today, this information is collected directly from employers on the W-2. Among the data that SSA collects are information about the employment patterns and annual earnings of each working adult in the United States. This information is used to keep track of qualifying employment toward benefits that citizens are entitled to once they reach a specific age. SSA also determines the known disability status of individuals seeking benefits. This information is critical to ensuring that SSA makes the correct benefit payments, but is also necessary to appropriately estimate the long-term costs of providing those

John Corson, the Director of the Bureau of Old-Age Insurance in 1938, correctly predicted:

**“(A)s a byproduct of its operations, the records of the Bureau of Old-Age Insurance will in [the] future provide a wealth of new sources of information regarding the working population of the United States.”**

Source: Olsen, A., & Hudson, R. (2009). Social Security Administration's Master Earnings File: Background Information. *Social Security Bulletin*, 69(3). Retrieved from <https://www.ssa.gov/policy/docs/ssb/v69n3/v69n3p29.html>

benefits and assess the condition of the various Social Security trust funds into which tax payments made by employers and individuals are deposited and from which these benefits are paid.

Although SSA is not a statistical agency, it does produce public-use data files that are made available to outside researchers. In addition, SSA will share personally identifiable data with some federal employees and outside researchers, including those with Special Sworn Status, to access the data for a particular purpose—but only if the personally identifiable data are necessary to accomplish the specific research or statistical purpose.<sup>29</sup> The limited release of statistical data was used by Federal Statistical Research Data Centers to permit secure access to Census data by researchers, including many graduate students working on their Ph.D. dissertations. In 2014, 85 graduate students from 30 different universities were engaged in research, including 69 who were using Census microdata.<sup>30</sup>

ED successfully used SSA data to calculate the mean and median incomes and debt-to-earnings ratios for graduates under the initial work on gainful employment regulations but without gaining access to personally identifiable data held by SSA. It is possible that the same outcome could have been achieved by using the Research Data Centers but with restrictions on the level at which the data can be released.<sup>31</sup>

The information that SSA holds, used to augment the information collected by Census in the Current Population Survey,<sup>32</sup> allows us to know, for example, how much college graduates make over their working lifetime in comparison to those with more and less education. If more broadly considered part of the national postsecondary education data ecosystem, it would be possible to specifically authorize the use of the information held by SSA to compute, for example, earnings outcomes or employment rates for completers and noncompleters from specific institutions and programs as well as determine whether financial aid receipt plays a role in the amount that people earn.

The NDNH is housed at SSA's National Computer Center, and the collaboration led to an expedited implementation of the NDNH. Housing the database at SSA's National Computer Center also ensured that it would be maintained in a world-class computer facility with state-of-the-art security standards.

For a summary of all data resources that could be used to understand the outcomes of postsecondary education in the United States, see **Appendix A**.

## Technical Enhancements Needed to Improve Federal Data Resources

One of the critical problems facing federal agencies when attempting to share information is the lack of a common numbering system. This problem even exists within ED, where several types of identification approaches are used. While this problem was theoretically addressed by ED two decades ago, the problem still plagues ED and will only be compounded when data from other agencies that use different identification systems become part of the national postsecondary data ecosystem.

### *Recommendation to facilitate data exchanges*

ED should begin working with the Office of Management and Budget to fully implement provisions of existing law and assign each institution participating in Title IV programs a single ED identification number to uniquely identify each postsecondary education institution in all federal programs,<sup>33</sup> so that information about institutions can be readily shared and compared to help identify effective best practices.

This is a necessary step for improving the postsecondary education data ecosystem. Without this systemic reform, it is difficult to imagine how the various data systems could be used in a coordinated way to increase our understanding of the outcomes of the nation's investments in postsecondary education.

### **Data Collection**

Much of the data that should be part of the national postsecondary education data ecosystem comes from outside the postsecondary education system itself. For example, information on the income of graduates, noncompleters, and those who never enrolled comes from one of three sources: tax records collected and maintained by IRS, wage and earnings information obtained by SSA, NDNH data obtained from states by HHS, and data held by the DoD and VA (see **Table 1**).

### *Recommendations for improving data collection*

- ▶ Fully automate the process of collecting information to determine veterans' eligibility for education benefits, and streamline institutional reporting to make it easier for veterans to access their benefits while ensuring proper use of scarce resources. This would greatly reduce burden on the institution by permitting it to extract information from existing enrollment management systems without the current manual input.
- ▶ Automate and systematize data collection for military recruiters as quickly as possible to reduce cost and burden, and improve coverage. Given that this system would likely be a simple file upload facility, minimal edits would be required because the data collected are standard directory information. Immediately taking steps to automate the process of collecting the information used by military recruit-

**TABLE 1: SOURCES OF STUDENT DATA OUTSIDE OF ED**

Tax Records	Wage and Earnings Information	National Directory of New Hires	Military Recruitment	GI Bill Education Benefits for Veterans
Earnings data are collected from employers by way of SSA.	Earnings data are collected directly from employers.	Earnings information comes from employers through states as part of the unemployment insurance system.	DoD recruiters collect information on college students.	VA collects information about the institutions where veterans use benefits to pay the cost of attending.
IRS collects information on the 1098-T directly from institutions of higher education.				

ers from institutions of higher education would reduce burden and cost and increase utility of the information.

- Consolidate the multiple data collections into two groups: (1) from employers, and (2) from institutions (see **Figures 2 and 3**).

This approach to bring together these diverse collections into an integrated system could reduce the cost to the government and the burden on institutions while increasing consistency and comprehensiveness. This could be accomplished by requiring institutions to transmit to IRS the information needed to fulfill these needs. 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. DoD would then work with VA to assess outcomes for veterans’ benefits recipients and annotate the records to identify individuals who already served in the armed forces.

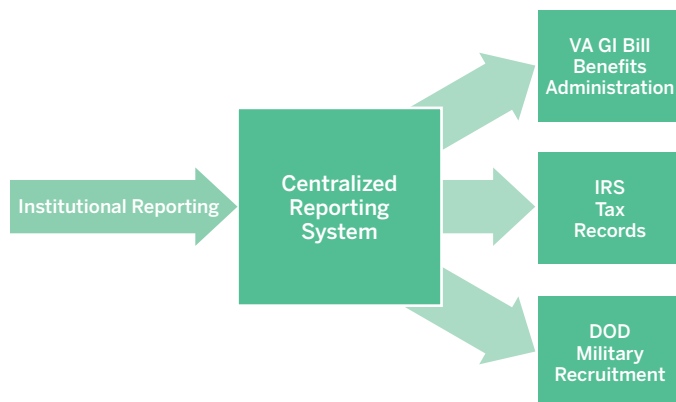
ED should immediately begin working with DoD, VA, and IRS to develop a system to exchange data with an eye toward

obtaining information on completion for aided students. If appropriate authorization is provided in law and the ban on a federal student unit record system is repealed, the entirety of the student record could be used to augment the data already available to ED. This would reduce the burden on institutions and significantly reduce the cost of the redundant data collections used by the various branches of the armed forces and individual recruiting stations. It could also result in the Integrated Postsecondary Education Data System (IPEDS) being populated with graduation rate statistics, which would include policy-significant subpopulations (e.g., veterans, service members, and tax-benefit recipients), derived from data exchange.

Wage and earnings information is submitted to SSA by employers annually. It likely accurately reflects the wages and earnings of individuals in the regular U.S. labor market. However, wages and earnings not reported to SSA are not included. Therefore, wages and earnings for jobs in the underground economy might not be included, nor would unreported tips and other unreported cash payments. A review of recent semiannual reports to Congress by SSA’s Inspector General offers no indication of concern about the underreporting of income. However, when ED was working on regulations to

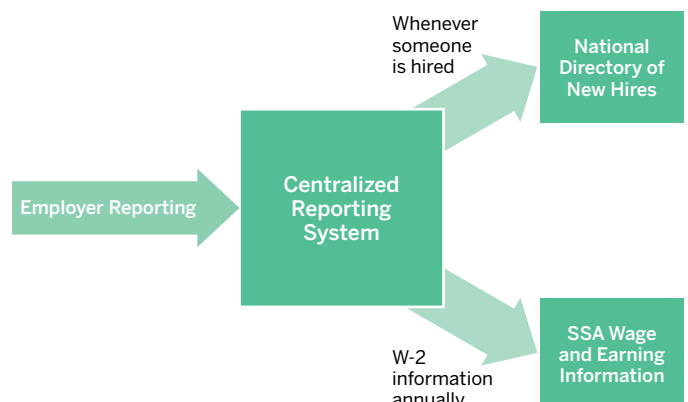
**FIGURE 2: ROLE OF LINKING FEDERAL DATA SYSTEMS IN DATA POLICY DISCUSSIONS**

Future State: National Postsecondary Education Data Infrastructure



**FIGURE 3: ROLE OF LINKING FEDERAL DATA SYSTEMS IN DATA POLICY DISCUSSIONS**

Future State: National Postsecondary Education Data Infrastructure





define gainful employment in a recognized occupation, it heard from representatives of cosmetology and culinary arts programs about underreporting of income for individuals completing those programs. Also, the data do not account for the earnings of individuals who leave the United States after leaving institutions of higher education.

Each state maintains a registry of newly hired employees that, in turn, feeds into the NDNH. The NDNH is augmented with data from every federal agency, which feed new hire and quarterly wage information into the NDNH directly. As such, the system has similar flaws to the wage and earnings information, with the additional weakness that it relies on employers reporting to each state appropriately as well as successful transmittal of the information into the NDNH by the state. Additionally, it is unlikely that the NDNH includes wages and earnings for jobs in the underground economy. Also, the NDNH would not include the earnings of individuals who leave the United States after leaving institutions of higher education because they are not employed in a state. The data are updated quarterly, but it is unclear how consistently or quickly each state reports to the NDNH.

Ultimately, the NDNH should be consolidated with SSA's wage and earnings reporting, to eliminate duplicative employer reporting of income.

**Data interoperability**  
(e.g., linkages with other data systems)

The various data resources are not designed to be interoperable (see Table 2). However, most federal data systems have names, dates of birth, and Social Security Numbers available in them. A key remaining issue is the development and use of common identifiers for institutions.

**Data governance (e.g., privacy and security)**

It should not be surprising that the issue of governance is at the core of why the totality of data resources is not brought together to build strong and credible information about how the U.S. higher education system is performing. All federal data resources are governed by specific laws and principally by the Privacy Act of 1974. The issues of data security and privacy are addressed at length in Joanna Lyn Grama's paper, "Understanding Information Security and Privacy in Postsecondary Education Data Systems." In some cases, such as taxpayer information or Census data, there are additional privacy and security requirements. Some of these additional requirements are out of genuine concern that disclosure of a taxpayer's information will be harmful to the individual or will impact voluntary compliance with tax collection requirements. In other cases, it is simply because the data resource has never been viewed as critical to the nation's postsecondary education data ecosystem (see Table 3).

**Data Use**

Each possible source of data on postsecondary education has its own limitation on use, given that the data are collected for a specific purpose. However, each data resource has the potential to unlock our understanding of the productivity and outcomes of our nation's higher education system (Table 4).

**Recommendations for data use**

- ▶ Create a mechanism to routinely generate summary statistical information from IRS, VA, and SSA records. The exact characteristics of the information to be generated should be carefully explored but should include the calculation of net price using the data from IRS and the development of a standardized net price calculator. Institutional efforts to build and control net price calculators make it difficult for

**TABLE 2: INTEROPERABILITY OF FEDERAL DATA SYSTEMS**

Tax Records	Wage and Earnings Information	National Directory of New Hires	Military Recruitment	GI Bill Education Benefits for Veterans
Tax records are specifically not designed to be interoperable with other data systems. However, since key identifiers like name, date of birth, and Social Security Number are available, it is easy to imagine that the data could be linked with other systems containing those variables. The tax records rely on the institution's federal identification number.	Like the tax records, the wage and earnings information obtained by SSA include the same key variables along with information about the employer, making it likely that the information could be readily merged with other federal postsecondary education data resources. No information about the institution(s) of higher education the employee attended is collected.	Like the tax records and wage and earnings information, the NDNH includes the same key variables along with information about the employer, making it likely that the information could be readily merged with other federal postsecondary education data resources. No information about the institution(s) of higher education the employee attended is collected.	Military recruiters are specifically prohibited from asking for Social Security Numbers; a student's race, ethnicity, nationality, or religious affiliation; or a student's grades or default status on loans. This limitation, particularly the lack of Social Security Numbers, makes it difficult to create an interoperable system. No institutional identifier is explicitly collected beyond the name and address of the institution.	VA has the key data necessary to link its records with tax information, wage and earnings information, and the NDNH. And because it has information about where GI Bill education benefits are used, VA could, with a high degree of confidence, match its data with military recruitment information if the data were placed in a single database. This would have the added benefit of assisting military recruiters by letting them know which students and graduates have already served in the military. VA uses a system of facility codes to identify where the student is attending.

**TABLE 3: DATA GOVERNANCE IN FEDERAL DATA SYSTEMS**

Tax Records	Wage and Earnings Information	National Directory of New Hires	Military Recruitment	GI Bill Education Benefits for Veterans
Information obtained by IRS is among the most carefully protected data resources held by the federal government. This information can be used only for the purpose of determining a taxpayer’s tax liability or to aid in the collection of a tax debt. However, it might be possible to find ways to use these data in a manner that does not jeopardize the privacy of the taxpayer or result in the disclosure of confidential information. This could be done by de-identifying the data or by producing summary statistics for analysis purposes.	SSA collects wage and earnings information to keep track of eligibility for future Social Security benefits. Unlike tax records, a mechanism exists for releasing information so long as the identity of the wage earner is protected. So, summary statistical information has been disclosed under the gainful employment regime and it is entirely possible that additional uses could be made of these data under current privacy and confidentiality rules. <sup>34</sup> As such, SSA must safeguard the information it maintains in administrative files against an invasion of an individual’s personal privacy. Additional legal restrictions are imposed once the data are merged with Census and IRS data. SSA has established a policy to share identifiable data only with those who have the legal authority to access data for a particular purpose, and only if identifiable data are required to accomplish a research or statistical purpose. <sup>35</sup>	The NDNH is used to help state child support programs establish and enforce fair and equitable child support orders. As noted above, access to the NDNH is very limited and governed explicitly in law.	The military recruiting information is obtained and held by recruiters in military recruiting stations around the country. No centralized repository has been created, making any use of the data extremely limited. More significantly, it is unclear how these individual recruiting stations hold and protect these data against inappropriate disclosure.	The data on veterans are held by VA. Until recently, VA had not made public any analyses of this information. However, the Student Veterans of America persuaded the National Student Clearinghouse and VA to work together to fill the gaps in knowledge about veteran success. <sup>36</sup>

**TABLE 4: DATA USE IN FEDERAL DATA SYSTEMS**

Tax Records	Wage and Earnings Information	National Directory of New Hires	Military Recruitment	GI Bill Education Benefits for Veterans
Information obtained by IRS is among the most carefully protected data resources held by the federal government. The 1098-T information is used only for the purpose of determining a taxpayer’s tax liability or to aid in the collection of a tax debt. The IRS 1098-T could be used to calculate a broad range of policy-relevant information net price by family income level and to populate consumer choice tools like the net price calculators that are more easily and consistently built for each institution. Doing so would eliminate the need for each institution to build and maintain its own net price calculator. This could be readily accomplished without disclosing any personally identifiable information.	SSA collects wage and earnings information for the purpose of keeping track of eligibility for future Social Security benefits. Summary statistical information has been generated from this information including the initial median and mean income of program completers’ debt-to-earnings ratios under the gainful employment regime. It is entirely possible that additional uses could be made of these data under current privacy and confidentiality rules. For example, accreditors might be able to calculate the earnings, debt-to-earnings, and broader measures of return on investment for the institutions or programs that they accredit as a resource to member institutions. Because SSA has already established procedures for public use of the data it collects, this data resource could be used to a much greater extent.	The NDNH is used to help state child support programs establish and enforce fair and equitable child support orders. Access to the NDNH is very limited. However, there is precedent for using the data in the NDNH to conduct analyses of the employment and income reporting of individuals served by certain HUD programs. Given this precedent on use, the permitted uses could be expanded to provide the Secretary of HHS the authority to use the information in the NDNH to conduct analyses of the impact of postsecondary education on individuals who are receiving support under TANF, or the Secretary of Education to conduct analyses of the employment and income of financial aid recipients as the Secretary of HUD can for individuals participating in certain housing programs.	The information collected under the Solomon Amendment is used exclusively by military recruiters at the local level. No national database has been created from this information, severely limiting its potential for other uses. For example, the information collected by DoD could be used to compute the postsecondary completion rates, time to completion, level of education, and degree fields for veterans without relying on the National Student Clearinghouse. This would also allow for the calculation of completion rates, time to completion, level of education, and degree fields of veterans at the institutional level, which, while technically possible, is not permitted by the National Student Clearinghouse.	Only recently has VA’s data on GI Bill Education Benefit recipients been used to assess the success of veterans in postsecondary education. This is the result of a public-private partnership that involved the Student Veterans of America, National Student Clearinghouse, and VA. The data released during the first wave of analysis of postsecondary academic outcomes included student veteran postsecondary completion rates, time to completion, level of education, and degree fields of nearly 1 million student veterans who initially used Montgomery and Post-9/11 GI Bill benefits between 2002 and 2010. <sup>37</sup>



students to make apples-to-apples comparisons of what they can expect to pay out of pocket without borrowing. Creating a standardized net price calculator would eliminate the existing inconsistencies.

- ▶ Create a mechanism to routinely calculate estimates of labor market outcomes at an institutional level (and, potentially, at a program within an institution), and combine this information with the net price to develop preliminary estimates of a return on investment for each institution of higher education. Currently, these kinds of data are available through the College Scorecard only for aided students, and such information has only been generated once, without an ongoing commitment to produce similar data in the future.
- ▶ Explore the possibility of computing additional variables to populate IPEDS with outcomes information derived from other federal data resources and share the results with the public and institutions of higher education to improve the quality and utility of data on the outcomes of participation in postsecondary education.
- ▶ Replace the current requirement that institutions develop and release a net price calculator with a requirement that IRS, in consultation with the Secretary of Education, develop net price calculators for institutions to reduce burden, increase accuracy, and improve comparability.
- ▶ Replace the reporting of net price by income group with data derived from the information that IRS receives from institutions of higher education to reduce burden, increase accuracy, and improve comparability.

### Resources Needed to Improve Federal Data Resources

The time horizon to improve federal postsecondary education data resources varies by the level of sophistication of the agency holding the data. Like the time horizon to improve federal postsecondary education data resources, the cost will vary by the level of sophistication of the agency holding the data as well (see **Table 5**).

### Recommendations for Data Use

OMB should allocate from funds already available for cross-agency enhancements to federal data resources to a pilot program to enhance the national postsecondary education data ecosystem beginning with the automation of the military recruitment collection including the capacity to identify veterans who are enrolled in postsecondary institutions.

### Public support

At present, there is very little public support for rationalizing the national postsecondary education data ecosystem. Indeed, most people do not recognize the gaps in our knowledge about higher education except for during the brief period of time that they are looking for postsecondary education programs for themselves or their families.

At the national level, none of the federal agencies that are deeply invested in higher education—IRS, SSA, HHS, VA, or DoD—view themselves as key elements of the postsecondary education data ecosystem despite the key roles they have played in recent years in describing the outcomes of that system. For this reason, 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 those agencies.

Clearly, the Social Security system is strengthened when more well-educated workers pay into the system for long periods of time. That is perhaps why SSA has, for many years, worked with the Census Bureau to produce estimates of the lifetime value of postsecondary education, and this work could be greatly enhanced if better integrated into the national postsecondary education data ecosystem.

IRS has become a primary funder of postsecondary education for students from middle- and upper-middle-income families through the American Opportunity Tax Credit. As such, it has begun to capture extremely powerful information about how postsecondary education is financed. Similarly, DoD, HHS, and SSA have come to hold powerful data resources on the outcomes of postsecondary education that have not been developed and exploited.

DoD needs to be able to more effectively recruit leaders into the many branches of the U.S. armed forces, which can only be accomplished through more thoughtful and strategic recruitment efforts. Such efforts require better data.

And VA must be able to make the case that the funding being provided to our nation's veterans is helping them transition to meaningful and rewarding careers in the civilian economy.

As a result, a key challenge is to build support among elected officials and the general public for the information necessary to measure the impact of postsecondary education. This will expand the view of what constitutes the national postsecondary education data ecosystem to include data resources that already exist or could be developed 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.

For students and families, new resources could be developed that would help them make better informed choices about which educational program at which institution to invest their time and money in.

**TABLE 5: RESOURCES NEEDED TO IMPROVE FEDERAL DATA RESOURCES**

Tax Records	Wage and Earnings Information	National Directory of New Hires	Military Recruitment	GI Bill Education Benefits for Veterans
<b>Time Horizon</b>				
<p>IRS recently demonstrated that it can conduct analyses similar to SSA when it supported ED's efforts to update and improve the College Scorecard. In that effort, IRS produced aggregated and de-identified estimates of institution-level statistics, such as the mean and median of the earnings distribution of federally aided students in a particular year, along with other metrics. In doing so, however, it did not demonstrate that it could manipulate the 1098-T information in a manner to produce summary statistical information or support the creation of consumer choice tools like the net price calculator. Therefore, it would likely take IRS more than six months but less than a year to build out such a capacity.</p>	<p>SSA already has an ecosystem for use of the data it collects. The experience with gainful employment is instructive. It took SSA approximately six months to build the ecosystem necessary to match the records of students enrolled in gainful employment programs with wage and earnings records. SSA has long recognized that the data it has are extremely useful, and it has worked to find ways to release the data without breaching the trust that taxpayers place in SSA. The work that ED did with SSA on gainful employment metrics provides a clear example of how these data can be leveraged at relatively low cost.</p>	<p>The NDNH is a highly automated system, so HHS should be able to fairly quickly begin to make the data collected available for analysis. However, the statutory framework for the NDNH severely limits HHS's ability to partner with other federal agencies to conduct analyses that would inform the national postsecondary education data ecosystem. This limitation, however, is not difficult to overcome. Congress has expanded the permissible uses of the NDNH over time. However, it is unclear what the utility of this would be, given that IRS and SSA have very similar information not intermediated by states.</p>	<p>DoD currently lacks any information technology ecosystem to support the collection of information on enrolled students. Unless such an ecosystem is built on top of an existing data resource—for example, the IRS 1098-T reporting—it will take 18 months to two years to design and implement a system to collect and make data available to recruiters for the 12 elements of the five branches of the U.S. armed services. Making these data available for other purposes will likely take longer.</p>	<p>VA has recently worked with the National Student Clearinghouse to produce an analysis of postsecondary academic outcomes included student veteran postsecondary completion rates, time to completion, level of education, and degree fields of nearly 1 million student veterans who initially used Montgomery and Post-9/11 GI Bill benefits between 2002 and 2010. Therefore, it would likely take VA more than six months but less than a year to build out its capacity.</p>
<b>Funding</b>				
<p>Since ED worked with IRS to develop the metrics released in the updated College Scorecard, it is likely that the costs of leveraging this data resource would be similar. It is likely, therefore, that IRS could, for the investment of only a few hundred thousand dollars, develop other appropriate metrics based on information it holds, including the 1098-T data. This cost could be offset by savings associated with the reduction in IPEDS collections on net price and the cost of maintaining the net price calculator template.</p>	<p>Since SSA already has the ecosystem and established approach to sharing data, it likely will cost relatively little to leverage this data resource. For example, the gainful employment regime, including the cost of processing institutional appeals, was budgeted in fiscal year 2013 to cost \$3.5 million and included the updating of the interface to obtain median and mean incomes along with computed debt-to-income ratios from SSA.<sup>38</sup></p>	<p>The data in the NDNH maintained by HHS would require substantially more resources, including a national student unit record system, to integrate it with the national postsecondary education data ecosystem. Currently, the NDNH is designed to be a secure data resource that is made available for extremely limited purposes.</p>	<p>DoD needs to build an entirely new system to support military recruitment. Building such a system would cost between \$600,000 and \$1 million.<sup>39</sup> Ultimately, however, the automation of the collection should yield significant savings over the piecemeal approach to collecting data.</p>	<p>Because VA has recently worked with the National Student Clearinghouse to develop outcome measures for veterans, it seems likely that a modest investment of a few hundred thousand dollars could result in better alignment with and integration into the national postsecondary education data ecosystem.</p>
<b>Capacity (or burden)</b>				
<p>IRS has adequate capacity and the information it could produce could significantly reduce the burden on institutions by negating the need to collect summary-level data on net price and develop net price calculators.</p>	<p>SSA already has the ecosystem and established approach to sharing data. It has the greatest capacity to be integrated into the national postsecondary education data ecosystem. Since it will be leveraging data that are already collected, it will not impact the reporting burden of the employers from whom it collects information, but it could substantially reduce the burden on institutions to provide data to accreditors and others on the employment and earnings of graduates.</p>	<p>HHS has very little capacity, although it may already be providing some support to HUD to evaluate some housing programs. The NDNH would provide timely information on new entrants to the labor force, which would be critical in determining, potentially for the first time, how long graduates and noncompleters spend looking for their first post-enrollment job.</p>	<p>DoD likely has no capacity to build a data collection system for military recruitment that could be integrated into the national postsecondary education data ecosystem. However, the data that it collects—level of education, academic major, and degrees received—are only available through a highly decentralized collection. If automated and systematized, it likely would dramatically reduce the burden imposed on institutions of higher education and ensure greater coverage than the current, apparently haphazard, approach.</p>	<p>VA likely has an adequate capacity, but the information it has available is limited. It would, therefore, be unlikely to reduce the burden on institutions by including (or excluding) the information they hold from the national postsecondary education data ecosystem.</p>

For state and institutional leaders, the additional data could help improve program offerings and better align those offerings to the demands of the state and local economy.

Federal policymakers would benefit from the sharing of federal data because it would reduce the cost of data collection and institutional reporting burden while significantly improving the ability to measure the return on investment in postsecondary education.

### **Action Items for Improving Federal Data Resources** **Recommendations to build public support**

- ▶ Advocates for enhancing our knowledge of the impacts of postsecondary education should use the critical moment when students and families are making decisions about postsecondary education institutions and programs to showcase what we do not know about the outcomes of the investment.
- ▶ Business leaders and other employers should demand 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 those agencies.
- ▶ Repeal the ban on a student unit record system and consider the inclusion of or linkage to additional federal data resources like earnings information from SSA and completion information from DoD in the authorization of a student unit record system, to reduce redundancy and cost and increase access and use.
- ▶ Provide access to earnings data to the Secretaries of Education and Veterans Affairs in order to access 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.
- ▶ Consolidate the NDNH with SSA's wage and earnings reporting. Requiring states to build an ecosystem that already largely exists at the national level is unnecessarily burdensome and expensive.
- ▶ Provide a statutory exception to the Privacy Act of 1974 to specifically provide for 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**

- ▶ Explore whether the NDNH could be augmented with information from other postsecondary education data resources to enhance its primary purpose of improving the collection of 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 the ability of states in crafting and enforcing child support orders.
- ▶ Explore whether the state longitudinal data systems could be enhanced by including information from the NDNH. While the restrictions on use of the NDNH are significant, an opportunity might exist to exploit some of the permissive language on state use of the data to augment these state systems.
- ▶ 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.

### **Conclusion**

The federal government already knows a great deal about its citizens. Unfortunately, too little of what is known is turned into information that can be acted on by federal and state policymakers and institutional leaders to improve the experience of students in higher education. Leveraging data the federal government already has could augment the information collected on higher education from students and institutions to answer the vital questions about college costs and near- and long-term outcomes that millions of students and their families ask every year.

At the national level, many federal agencies are deeply invested in higher education. ED, IRS, VA, and DoD all provide significant funding to students enrolled in higher education. And all federal agencies along with nearly all private and public sector employers depend on higher education to provide adequately prepared employees.

Our nation's higher education system must improve to meet the needs of our economy, and it will only improve with more and better data on outcomes. A key aspect in answering that challenge is leveraging the data that already exist about postsecondary outcomes.

**APPENDIX A: SUMMARY OF FEDERAL DATA RESOURCES THAT COULD BE EXPLOITED TO BETTER UNDERSTAND THE OUTCOMES OF POSTSECONDARY EDUCATION IN THE UNITED STATES**

	Employer Reporting		Institutional Reporting		
	National Directory of New Hires	Wage and Earnings Information	Military Recruitment	GI Bill Education Benefits for Veterans	Tax Records
Agency	U.S. Department of Health and Human Services (HHS)	Social Security Administration (SSA)	Components of the Department of Defense (DoD)	Department of Veterans Affairs (VA)	Internal Revenue Service (IRS)
Authority	Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA)	Social Security Act	Solomon Amendment to the General Military Law	Post-9/11 GI Bill	Tax Code
Purpose	To assist state child support agencies in locating parents and enforcing child support orders	To calculate Social Security benefits for individuals. Data are also shared with IRS to support the collection of income taxes.	To assist military recruiters in identifying and recruiting individuals to serve in the U.S. armed forces	To ensure proper payment of veterans' benefits	To assist taxpayers in qualifying for various higher education tax benefits and assist IRS in ensuring taxpayer eligibility for those benefits
Where the data come from (i.e., who reports the data)	Employers through states	Employers directly to SSA, which shares the data with IRS	Institutions of higher education		
Data elements	<p>In the new hire's file:</p> <ul style="list-style-type: none"> <li>Name</li> <li>Social Security Number (SSN)</li> <li>Address</li> <li>Employer name</li> <li>Federal employer identification number</li> <li>Employer address</li> <li>Date of hire</li> </ul> <p>In addition, in the quarterly earnings file:</p> <ul style="list-style-type: none"> <li>Employee wage amount</li> <li>Reporting period</li> </ul> <p>In addition, in the Unemployment Insurance (UI) file:</p> <ul style="list-style-type: none"> <li>Gross wage amount (before any deductions)</li> <li>Reporting period</li> </ul>	<ul style="list-style-type: none"> <li>Name</li> <li>Address</li> <li>SSN</li> <li>Employer identification number</li> <li>Employer's name, address, and ZIP code</li> <li>Wages, tips, and other compensation</li> <li>Federal income tax withheld</li> <li>Social Security wages</li> <li>Social Security tax withheld</li> <li>Medicare wages and tips</li> <li>Medicare tax withheld</li> <li>Social Security tips</li> <li>Allocated tips</li> <li>Advance earned income credit payment</li> <li>Dependent care benefits</li> <li>Nonqualified plan distributions</li> <li>Codes for reporting types of deferred compensation</li> <li>Checkboxes for statutory employee status, retirement plan participation, and third-party sick pay disburser status</li> <li>State and local income tax information</li> </ul>	<ul style="list-style-type: none"> <li>Name</li> <li>Address</li> <li>Telephone number</li> <li>Age and date of birth</li> <li>Place of birth</li> <li>Level of education</li> <li>Academic major</li> <li>Degrees received</li> <li>Educational institution in which the student was most recently enrolled</li> </ul>	<ul style="list-style-type: none"> <li>Name</li> <li>SSN</li> <li>Student ID number</li> <li>Payee number</li> <li>Phone number</li> <li>E-mail address</li> <li>Benefit chapter</li> <li>Program</li> <li>Training type</li> <li>Facility code at which the student is or will be attending</li> <li>Period of enrollment by term</li> <li>Number of residential credits taken during the term</li> <li>Number of distance credits taken during the term</li> <li>Number of remedial credits taken during the term</li> <li>Number of clock hours taken per week, if applicable</li> <li>Tuition and fees charged per term</li> <li>Amount the institution has agreed to provide above the highest in-state tuition rate</li> <li>Charges above the in-state rate</li> </ul>	<ul style="list-style-type: none"> <li>Institution's name</li> <li>Address</li> <li>Telephone number</li> <li>Federal employer identification number; student's SSN</li> <li>Payments received for tuition and related expenses</li> <li>Amounts billed for qualified tuition and related expenses</li> <li>Adjustments made for prior year</li> <li>Scholarships and grants</li> <li>Adjustments to scholarships or grants for a prior year</li> <li>Amount of any contract reimbursement or refund</li> <li>Whether the student was enrolled at least half time, or was a graduate student<sup>40</sup></li> </ul>
Data dissemination	The information contained in the NDNH can be released <b>without</b> personal identifiers for research purposes found by the Secretary of HHS to be likely to contribute to achieving the purposes of Part A or Part D of the Social Security Act.	SSA collects wage and earnings information for the purpose of keeping track of eligibility for future Social Security benefits. Summary statistical information has been generated from this information SSA has established procedures for releasing data for research purposes.	The information collected under the Solomon Amendment is used exclusively by military recruiters at the local level. No national database is created from this information, severely limiting its potential for other uses.	The data are generally used only to distribute benefits to eligible veterans. Only recently has VA's data on GI Bill Education Benefit recipients been used to assess veterans' success in postsecondary education.	Information obtained by IRS is among the most carefully protected data resources held by the federal government. The 1098-T information is only used for the purpose of determining a taxpayer's tax liability or to aid in the collection of a tax debt.
Pros	<p>Because it represents quarterly reporting of income, it is more timely.</p> <p>An ecosystem exists for the use of these data for both compliance and analytic purposes.</p>	<p>It is the most inclusive source of income information.</p> <p>It has the potential to obtain information about the employer's industry.</p>	It is a potential source of completion information for all students enrolled in postsecondary education.	It is a potential source of comparison data for adult learners enrolled in postsecondary education, particularly for students who are pursuing programs via distance learning.	It is a potential source of high-quality net price information on all students enrolled in postsecondary education.
Cons	It does not capture income from self-employment.	Data are reported only annually, so there is a considerable delay in having data available for analysis.	No national database currently exists.	Data are currently used only to award benefits.	Data are currently used only to determine eligibility for tax benefits.
What questions could be answered if linked with other federal sources?	What are the earnings, debt to earnings, and return on investment for the institutions or programs?	What are the earnings, debt to earnings, and return on investment for the institutions or programs?	What is the completion rate for various groups of students including veterans and aid recipients?	Where are veterans enrolled and where do they complete?	What is the extent of unmet need? How does unmet need impact persistence and degree attainment?

## Endnotes

- 1 5 U.S.C. § 552(a).
- 2 20 U.S.C. § 1094(b).
- 3 Soldner, M. & Campbell, C. (2016). *Using—and improving—FSA data systems to support policy analysis*. American Institutes for Research and Association of Community College Trustees.
- 4 Miller, B. (2016). *Building a student-level data system*. Center for American Progress.
- 5 5 U.S.C. § 552(a).
- 6 Grama, J.L. (2016). *Understanding information security and privacy in postsecondary education data system*. Educause.
- 7 U.S. Department of the Treasury, Internal Revenue Service. (2016, April 1). *Form 1098-T*. Retrieved from <https://www.irs.gov/uac/Form-1098-T-Tuition-Statement>
- 8 26 I.R.C. § 25A(f)(2), from all institutions eligible under title IV of the Higher Education Act at a minimum, that they are not collecting them at
- 9 U.S. Department of Education, National Center for Education Statistics. (2015). *IPEDS: Average institutional net price FAQs*. Retrieved from [http://nces.ed.gov/ipeds/Section/Institutional\\_net\\_price](http://nces.ed.gov/ipeds/Section/Institutional_net_price)
- 10 20 U.S.C. § 1015a(a); 20 U.S.C. § 1015a(h).
- 11 U.S. Department of Education, National Center for Education Statistics. (2016, January 28). *IPEDS: Net price calculator information center*. Retrieved from [http://nces.ed.gov/ipeds/Section/Net\\_price\\_calculator](http://nces.ed.gov/ipeds/Section/Net_price_calculator)
- 12 U.S. Department of the Treasury, Internal Revenue Service. (2016, April 1).
- 13 U.S. Department of Education. (2015). *Using federal data to measure and improve the performance of U.S. institutions of higher education*. Retrieved from <https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf>
- 14 10 U.S.C. § 983; Buckley, B., Baker, E., Gourley, J., Kunisch, W., & Rizkallah, J. (2001). *The Solomon Amendment: A guide for recruiters and student records managers*. Washington, DC: American Association of Collegiate Registrars and Admissions Officers.
- 15 Author's calculation using "religious affiliation" variable from U.S. Department of Education, National Center for Education Statistics, *College Navigator*. Retrieved 2016, February 26 from <http://nces.ed.gov/collegenavigator/>
- 16 Hlavac, G.C. & Easterly, E.J. (2015). *FERPA primer: The basics and beyond*. National Association of Colleges and Employers. Retrieved from <http://www.nacweb.org/public/ferpa0808.htm>
- 17 U.S. Department of the Navy. (2015, March 25). *U.S. Marine Corps concepts & programs: Recruiting*. Retrieved from <https://marinecorpsconceptsandprograms.com/programs/manpower/recruiting>
- 18 Buckley et al.
- 19 U.S. Department of Veterans Affairs. (2015, November 24). *VA-ONCE quick reference users guide*. Retrieved from [http://www.benefits.va.gov/gibill/school\\_training\\_resources.asp](http://www.benefits.va.gov/gibill/school_training_resources.asp)
- 20 U.S. Department of Veterans Affairs. (2016). *Department of Veterans Affairs budget in brief 2016*.
- 21 Cate, C.A. (2014). *Million Record Project: Research from Student Veterans of America*. Student Veterans of America. Retrieved from [http://www.studentveterans.org/images/Reingold\\_Materials/mrp/download-materials/mrp\\_Full\\_report.pdf](http://www.studentveterans.org/images/Reingold_Materials/mrp/download-materials/mrp_Full_report.pdf)
- 22 Government Performance and Results Act, Public Law 103-62, 103rd Congress, (1993). Retrieved from <https://www.gpo.gov/fdsys/pkg/STATUTE-107/pdf/STATUTE-107-Pg285.pdf>
- 23 Afet Dunder and Doug Shapiro, in their paper written for this series, make the case for a continuing and expanded role for the National Student Clearinghouse in the postsecondary ecosystem. Shapiro & Dunder correctly point out that often the federal government will not be able to pass back to institutions as much data as they would need to improve their postsecondary programs. As an agent acting on behalf of institutions, the Clearinghouse can fill this role even if summary statistical information becomes available from harvesting existing federal data resources. Dunder, A., & Shapiro, D. (2016). *The National Student Clearinghouse as an integral part of the national postsecondary data infrastructure*. Herndon, VA: National Student Clearinghouse Research Center.
- 24 U.S. Department of Health & Human Services, Office of the Administration for Children & Families Office of Child Support Enforcement. (2015, July 1). *Overview of National Directory of New Hires*. Retrieved from <http://www.acf.hhs.gov/programs/css/resource/overview-of-national-directory-of-new-hires>
- 25 Hearings before the House Committee on Ways and Means, Subcommittee on Human Resources, 105th Congress (1998, May 19) (testimony of Donna Bonar). Retrieved from <http://www.hhs.gov/asl/testify/t980519c.html>
- 26 U.S. Social Security Administration. (2010, March 15). *Compilation of the social security laws: Federal parent locator service*. Retrieved from [https://www.ssa.gov/OP\\_Home/ssact/title04/0453.htm](https://www.ssa.gov/OP_Home/ssact/title04/0453.htm)
- 27 U.S. Social Security Administration. (2010, March 15).
- 28 U.S. Department of Health & Human Services, Office of the Administration for Children & Families Office of Child Support Enforcement. (2015, January 27). *Federal Parent Locator Service: A guide to the National Directory of New Hires*. Retrieved from [https://www.acf.hhs.gov/sites/default/files/programs/css/a\\_guide\\_to\\_the\\_national\\_directory\\_of\\_new\\_hires.pdf](https://www.acf.hhs.gov/sites/default/files/programs/css/a_guide_to_the_national_directory_of_new_hires.pdf)
- 29 McNabb, J., Timmons, D., Song, J., & Puckett, C. (2009). *Uses of administrative data at the Social Security Administration*, 69(1), 75-84. Retrieved from <https://www.ssa.gov/policy/docs/ssb/v69n1/ssb-v69n1.pdf>
- 30 U.S. Census Bureau. (2015). *Center for Economic Studies and Research Data Centers research report: 2014*. Retrieved from [http://www.census.gov/ces/pdf/2014\\_CES\\_Research\\_Report.pdf](http://www.census.gov/ces/pdf/2014_CES_Research_Report.pdf)
- 31 U.S. Department of Education, Office of Federal Student Aid. (2015, September 3). *Information About the Gainful Employment Downloadable Spreadsheet*. Retrieved from <https://studentaid.ed.gov/sa/about/data-center/school/ge/data>
- 32 U.S. Census Bureau. (2015, December 16). *Current Population Survey*. Retrieved from <http://www.census.gov/programs-surveys/cps.html>
- 33 20 U.S.C. § 1094(b).
- 34 Other suggestions for how labor market information could be used to inform higher education policy are provided by Rachel Zinn in her paper, *Classroom to career: Leveraging employment data to measure labor market outcomes*. For example, it might be possible to work with SSA to obtain aggregate employment and earnings information for graduates of institutions accredited by a regional or national accrediting agency. Alternatively, a professional society might work with institutions to obtain employment and earnings information for all program completers in a particular discipline. The wage and earnings data are governed by the provisions of the Privacy Act.
- 35 The entity that wishes to gain access to the data held by SSA must submit a proposal, data protection plan, and confidentiality agreements prior to being granted access to the data, and the user must guarantee to keep the data secure, not re-disclose the data, and restrict the use of the data to the approved purpose. A user can access SSA data that has been merged with Census data only for purposes that have been approved by Census and where the analysis is performed at a site approved by Census.
- 36 Cate, C.A. (2014).
- 37 Cate, C.A. (2014).
- 38 In 2017, this task was consolidated with cohort default rate appeals, making it impossible to isolate the cost of the gainful employment regime. U.S. Department of Education, Student Aid Administration. (2012). *Fiscal year 2013 request*. Retrieved from <https://www2.ed.gov/about/overview/budget/budget13/justifications/aa-saadmin.pdf>
- 39 The core frame for such a system already exists in various systems at ED including IPEDS and the Campus Security, Equity in Athletics, and Accreditation collections, but a collection tool would need to be built with appropriate safeguards to ensure that only an authorized institutional representative, likely the registrar, can upload the information. In addition, help desk support would be needed to help institutions submit the required data, which are minimal, and to help institutions resolve edits. Finally, tools would need to be built to allow recruiters to access the resulting database and allow for other appropriate uses of the data collected. A contract to support the Campus Security, Equity in Athletics, and Accreditation collections along with other work was awarded in 2015 and totaled \$2.3 million for a five-year period.
- 40 U.S. Department of the Treasury, Internal Revenue Service. (2016, April 1).









*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.

