Imagining the Future of Postsecondary Education Data

The Role of a Federal Student-Level Data Network

Prepared for

Gates Foundation
500 Fifth Avenue North
Seattle, WA 98109

September 2025

Prepared by

Erin Thomsen
Andrew Crichton
Kaitlin Lilienthal
Harper Haynes
RTI International
3040 E. Cornwallis Road
Durham, NC 27713-2852
www.rti.org

RTI INTERNATIONAL

Imagining the Future of Postsecondary Education Data: The Role of a Federal Student-Level Data Network

Prepared for

Gates Foundation
500 Fifth Avenue North
Seattle, WA 98109

Prepared by

Erin Thomsen
Andrew Crichton
Kaitlin Lilienthal
Harper Haynes
RTI International
3040 E. Cornwallis Road
Durham, NC 27713-2852
www.rti.org

September 2025

This report was prepared for the Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Gates Foundation.

RTI Project Number 24677_1.1

Executive Summary

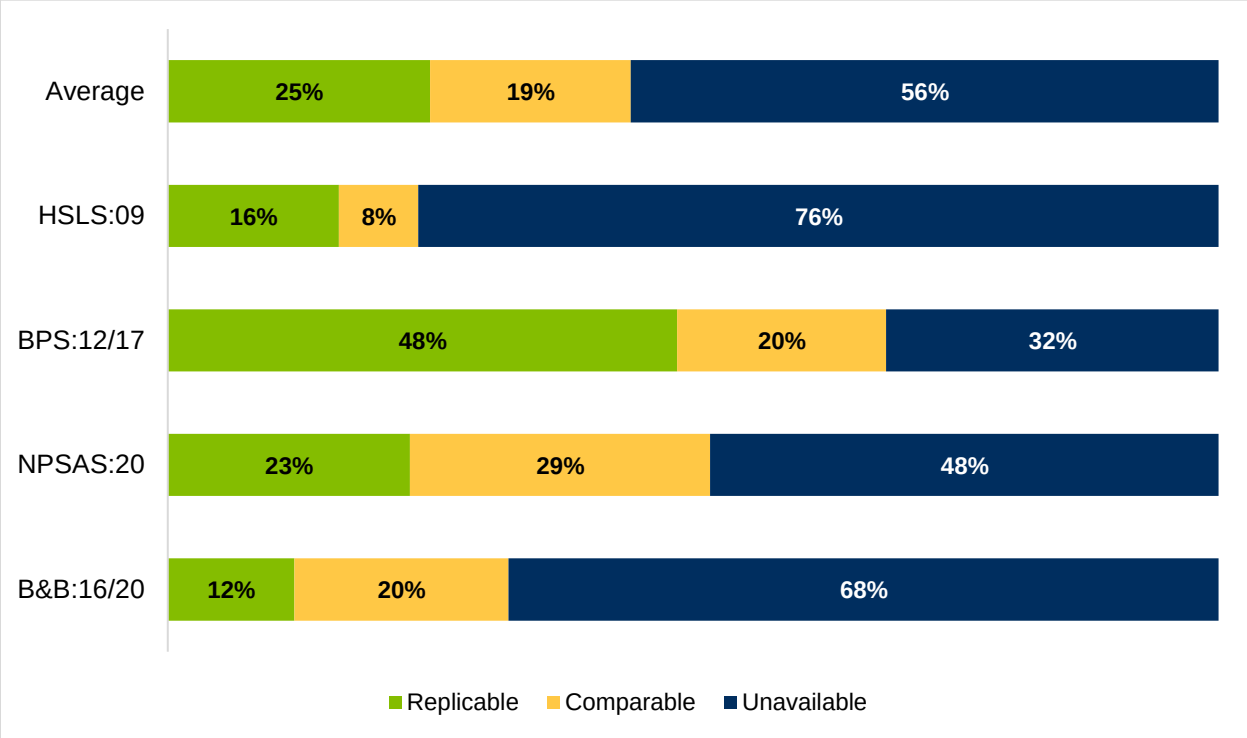
CONTEXT: In an effort to think strategically about redesigning the federal postsecondary education data landscape to maximize efficiency and data utility, this report examines how a proposed federal student-level data network (SLDN) could add efficiencies and reduce burdens in future federal sample survey research efforts.

METHODS: RTI selected four NCES sample surveys of postsecondary students: the High School Longitudinal Study of 2009 (HSLS:09), the Beginning Postsecondary Students Longitudinal Study (BPS), the National Postsecondary Student Aid Study (NPSAS), and the Baccalaureate and Beyond Longitudinal Study (B&B). We compared each analysis variable from the most recent publicly available data file to the list of potential data elements for the SLDN in recently proposed legislation. For each variable, we recorded whether one or more SLDN data elements could replicate the variable exactly (coded as *replicable*) or whether they could provide a similar measure of the variable (coded as *comparable*). If the variable was neither replicable nor comparable, it was coded as *unavailable* from an SLDN. We then calculated the percentage of variables that would be replicable, comparable, and unavailable using an SLDN, overall and within a subset of variables in a key content area for each study.

RESULTS: There is considerable variation across the four studies, with an average of 48 percent of variables coded as either replicable or comparable using SLDN data elements (see **Figure A**). The study with the highest percentage of replicable or comparable variables was BPS (68 percent), followed by NPSAS (52 percent), B&B (32 percent), and HSLS (24 percent).

DISCUSSION: The results suggest that an SLDN could add efficiency to the design of sample surveys by collecting administrative data in a centralized and standardized location, simplifying the administrative data collection procedures for institutional reporters and government researchers and contractors. An SLDN might also support shorter surveys—and thus reduce burden on survey respondents—by eliminating the need to ask some survey questions. However, data on student reasons for choosing an institution or degree program, student perceptions of support from their institutions or instructors, and graduate job satisfaction and benefits would be unavailable in an SLDN and thus need to be collected in surveys.

Figure A. Frequency distribution of variable availability in an SLDN, by study



NOTE: Variables were coded as *replicable* when the listed SLDN data elements could reproduce the operationalization of the study variable, regardless of whether one or more SLDN data elements would be needed to do so. Variables were coded as *comparable* when the current list of SLDN data elements provides a similar but not replicable measure of the study variable in question. Variables were coded as *unavailable* when no SLDN data element(s) aligned with the study measure. Coding assignments were determined through independent dual reviews and in-depth team discussions to resolve disagreements.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 second follow-up, 2012/17 Beginning Postsecondary Student Longitudinal Study, 2019–20 National Postsecondary Student Aid Study, and 2016/20 Baccalaureate and Beyond Longitudinal Study; the College Affordability Act, H.R. 4674, 116th Congress, 2019; Velez, E. D., Pretlow, J., and Roberson, A. J. (2020) Implementing a Federal Student-Level Data Network: Advice from Experts.

Contents

Tables & Figures

Introduction

As a result of recent actions to reform the Institute of Education Sciences and National Center for Education Statistics (NCES), the future landscape of federal postsecondary education research will look different from that of the past. There is still uncertainty around what will and will not be included. Some have proposed the creation of a federal student-level data network (SLDN). Since 2019, Congress has introduced three bills in the House of Representatives that outline specifications for a student-level data network: the College Affordability Act (CAA),[1] the College Transparency Act (CTA),[2] and the College Cost Reduction Act (CCRA).[3] One goal of an SLDN as outlined in these bills is to "accurately evaluate student enrollment patterns, progression, completion, and post-collegiate outcomes, and higher education costs and financial aid."[4] The legislation would mandate that higher education institutions report student-level data such that the student-related metrics currently collected in the Integrated Postsecondary Education Data System (IPEDS) could be replicated by aggregating those student-level data.[5] Another goal of the SLDN would be to reduce the reporting burden on institutions, which currently need to aggregate data themselves for reporting into IPEDS.[6] The legislation outlines required data elements for an SLDN, matching and data sharing arrangements with other federal data systems, and privacy and security requirements.

If any of these bills were to become law, it would allow access to unit-level data for all students attending a Title IV postsecondary institution, which would be a significant shift in the field of postsecondary education research. Thus, in recent years, RTI International—often in partnership with the Institute for Higher Education Policy (IHEP)—has carried out activities to explore implications, design considerations, and issues for further investigation for an SLDN. RTI and IHEP convened discussion panels of experts on institutional reporting and financial aid from institutions and higher education associations. Additionally, RTI explored the landscape of existing state- and student-level data systems in the United States as well as national SLDNs in Australia, Canada, Germany, Norway, Switzerland, and the United Kingdom to inform potential designs and options for a federal SLDN in the United States.[7]

Utilizing data collected in an SLDN could add efficiencies to sample survey research and reduce burden on both institutions and student respondents while simultaneously preserving the unique powers of survey research to capture context that administrative data cannot provide. For example, an SLDN could serve as a sampling frame for future sample studies, eliminating the burden on institutions to

Utilizing data collected in an SLDN could add efficiencies to sample survey research and reduce burden on both institutions and student respondents.

[1] The College Affordability Act (H.R. 4674) was introduced in the 116th Congress in October 2019.
[2] The College Transparency Act (H.R. 2957) was introduced in the 118th Congress in April 2023.
[3] The College Cost Reduction Act (H.R. 6951) was introduced in the 118th Congress in January 2024.
[4] The College Affordability Act, Section 1022(a)(1)(1)(A)(i).
[5] The College Affordability Act, Section 1022(a)(1)(2)(C)(i).
[6] The College Affordability Act, Section 1022(b)(2).
[7] This series of reports can be found on RTI.org: Projects | Building a Federal Student-Level Data Network.

provide student rosters for sampling and reducing the cost for government contractors and researchers to collect the rosters. There is also meaningful overlap between the proposed data elements for an SLDN and the data elements collected in the NCES sample surveys that could be leveraged to further reduce costs and burden. Historically, the NCES sample surveys combined survey data with institution records and other administrative data. In the future, an SLDN could add efficiency in administrative data matching by centralizing much of the administrative data collections for NCES studies, thus reducing the burden on institutions and other agencies who provide institution records and other administrative data separately for each sample survey. Moreover, administrative data from an SLDN could be combined with survey data, resulting in shortened surveys that eliminate questions about information stored in an SLDN and focus on questions that tell the story of students' choices, beliefs, and attitudes. Utilizing both an SLDN and focused NCES sample surveys could lead to greater efficiencies and increased opportunities to collect data and inform policy development.

This report explores the relationship between a possible future SLDN and selected sample studies, including the High School Longitudinal Study of 2009 (HSLS:09), Beginning Postsecondary Students Longitudinal Study (BPS), National Postsecondary Student Aid Study (NPSAS), and the Baccalaureate and Beyond Longitudinal Study (B&B), that collect data on postsecondary students. Specifically, it explores the data elements proposed for an SLDN that overlap with data elements collected on each of these selected sample studies, as well as the data elements that would not overlap. From this review, we will be able to examine the extent to which a future SLDN might complement and support data collections like the selected studies. Additionally, this report will discuss the key content areas that researchers and policymakers often use HSLS, BPS, NPSAS and B&B to study, and how an SLDN and survey data could be efficiently combined to address those topics.

Methods

The objective of this report is to determine the amount of overlap between the data elements collected in NCES sample studies and the data elements proposed for a future SLDN. First, we selected four NCES sample studies to review: HSLS, BPS, NPSAS, and B&B. These studies were selected because they are among the NCES surveys commonly cited in research and reporting on postsecondary education. Their data elements cover a wide range of postsecondary education topics, from the transition from high school to postsecondary education, persistence and degree attainment, the cost of postsecondary education and how students and families pay for it, and postbaccalaureate education and workforce outcomes. Additionally, on behalf of NCES, RTI has managed the data collection and data processing for these studies for decades and is deeply familiar with their data elements. **Table 1** summarizes the characteristics of each of the four studies reviewed for this report.

Table 1. Overview of study characteristics

Study	Sample Characteristics	Primary Content Focus	Data Sources
HSLS:09 second follow-up	Students in 9th grade in the fall of 2009 who attended regular public or private schools that provided instruction in both 9th and 11th grade in the 50 states or the District of Columbia. The sample for the second follow-up in 2016 included sample members who responded in at least one prior round and had not withdrawn from the study.	High school educational experiences, high school academic performance, and the transitions that students make from high school to adult roles. The second follow-up focused on the transition to postsecondary education and the workforce, including access to higher education and choice of postsecondary institution.	Respondent survey
BPS:12/17	First-time postsecondary students in the 2011–12 academic year enrolled at Title IV–eligible institutions.[1] The final follow-up (BPS:12/17) contacted the sample 6 years after first starting postsecondary education.	Postsecondary persistence and degree attainment, transition to employment, academic and work experiences.	Respondent survey, institutional records, FAFSA, NSC, CPS, NSLDS, IPEDS, VBA, ACT Test, College Board
NPSAS:20	Students enrolled in postsecondary courses at Title IV–eligible institutions[1] during the 2019–20 academic year. First-time beginning students were oversampled as the basis group for BPS:20 follow-up studies.	Financial aid provided by the federal government, the states, postsecondary institutions, and private agencies; student demographic and enrollment data.	Student survey, institutional records, FAFSA, NSC, CPS, NSLDS, IPEDS, VBA, ACT Test, College Board
B&B:16/20	Graduates who completed the degree requirements for a bachelor's degree at a Title IV–eligible institution[1] in the 2015–16 academic year. The second follow-up, B&B:16/20 in 2020, was 4 years after bachelor's degree completion.	Postbaccalaureate outcomes: employment, continued enrollment, student debt and repayment, and financial well-being.	Respondent survey, NSLDS, NSC

[1] Title IV–eligible institutions are postsecondary institutions that participate in federal student aid programs under Title IV of the Higher Education Act of 1965.

NOTE: B&B:16/20 = 2016/20 Baccalaureate and Beyond Longitudinal Study; BPS:12/17 = 2012/17 Beginning Postsecondary Student Longitudinal Study; CCD = Common Core of Data; CIP = Classification of Instructional Programs; CPS = Central Processing System; FAFSA = Free Application for Federal Student Aid; HSLS:09 = High School Longitudinal Study of 2009; IPEDS = Integrated Postsecondary Education Data System; NPSAS:20 = 2019–20 National Postsecondary Student Aid Study; NSC = National Student Clearinghouse; NSLDS = National Student Loan Data System; PSS = Private School Universe Survey; VBA = Veterans Benefits Administration

For each study, we compared each analysis variable from the most recent publicly available round of data collection to the list of data elements proposed in the CTA, the CAA, and the CCRA for a potential SLDN. A working interpretation of the legislation for the required data elements in an SLDN was first published in *Implementing a Federal Student-Level Data Network: Advice from Experts* (Velez, Pretlow, & Roberson, 2020). The table published in that brief summarizing the required data elements for an SLDN can also be found in this report's appendix. The comparison review was to determine whether each variable from the selected sample studies had similarities with a proposed data element for an SLDN and to what extent the data elements overlapped. For each variable, the following was recorded:

1. **Whether the sample study variable would be replicable, comparable, or unavailable** in a future SLDN according to the list of data elements specified in the CAA/CTA/CCRA (see appendix).

    a. Variables were coded as ***replicable*** when one or more of the listed SLDN data elements could *reproduce* the operationalization of the study variable. For example, *military status*, a variable in several sample surveys, would be available directly in an SLDN. Similarly, *ever attended a 2-year public institution*, another variable in several sample surveys, could be calculated by combining *institution level* and *institution control* from the SLDN.

    b. Variables were coded as ***comparable*** when the list of SLDN data elements provided a *similar but not replicable* measure of the study variable in question. For example, *annualized total pay from all current jobs* in B&B is derived at the individual level directly from the respondent survey; *aggregate earnings by institution and program* from the list of SLDN data elements would offer analysts a similar but less individually specific measure.

    c. Variables were coded as ***unavailable*** when no SLDN data element(s) aligned with the study variable. For example, *hours worked weekly* from BPS:12/17, a variable constructed from the respondent survey only, would not be available from the current list of SLDN variables.

2. **Variable source in origin study.** This information provided important context and a critical check for replicability because variables that were solely sourced from surveys are not likely to be replicable in an SLDN. Some exceptions were made for those measures that are typically standardized across sources (e.g., sex, race).[8]

3. **Relevant SLDN data element(s).** This information was used for mapping study variables onto SLDN data elements when replicable or comparable and to support justifications for coding decisions.

4. **Whether the variable would be required to be collected in a future SLDN** according to the legislation as it is written in the CAA/CTA/CCRA. Each variable was coded as either *definitely required*, *possibly required*, or *not required*. The three codes were selected to allow for some ambiguity and room for interpretation in the legislative text. A field to identify the reference language from the CAA/CTA/CCRA when indicating that a data element would be definitely or possibly required was also documented. The information considered in this field should not be confused with the existing statues governing the postsecondary studies, such as 20 U.S.C. Chapter 28 § 1015(k), which mandates a survey of recipients of federal student financial aid (i.e., NPSAS). The requirements of the existing statutes were not analyzed as part of this review.

[8] In the case of the HSLS:09 second follow-up, all variables were sourced from the student survey, as administrative source matching did not take place in that round of the study. Variables were still marked as *replicable* if they corresponded to an SLDN data element.

Study weights and other methodological variables, like imputation flags, were excluded from the variables reviewed. For the postsecondary longitudinal studies, prior round variables were also excluded because many variables are redundant across rounds. For example, in B&B, several variables were derived to measure the same concepts at the time of data collection for both the 1-year follow-up and the 4-year follow-up, such as *current or most recent occupation* and *current monthly loan payment amount.* Only the 4-year follow-up iteration of these measures was included in the review. Likewise, only the 6-year follow-up versions of repeated BPS:12/17 measures were included. Because many of the excluded prior-round variables are the same measure at a different point in time, the proportion of variables identified as replicable, comparable, and unavailable in the latest round of the study is commensurate with the study as a whole. Prior-round variables were also excluded from the analysis of HSLS:09. In the case of HSLS:09, the prior rounds occurred when sample members were still in high school, and the proposed legislation prohibits an SLDN from collecting elementary and secondary education data.[9]

A team of four analysts conducted the review of variables. Each analyst served as the primary reviewer on one study and the secondary reviewer on another study. This procedure provided a dual review for each study and cross-study awareness for consistency in decisions. First, primary reviewers coded each variable. Then, secondary reviewers independently reviewed each variable and the coding decisions of the primary reviewer and documented any disagreements with the primary review. Finally, all four analysts met to discuss findings and study overlap, raise questions, and decide on resolutions for each coding disagreement. Once all coding disagreements had been discussed and coding was complete, summary statistics were calculated.

In addition to the overall distribution of variables from each study that would be replicable, comparable, or unavailable with an SLDN, we investigated the availability of a subset of variables that corresponded with the study's primary research questions and were emphasized in each study's design. The key content areas of a study are informed by theory and utility. These content areas demonstrate a study's unique contribution to the public, including taxpayers, researchers, and policymakers. Each study's publicly available codebooks include subject labels for each variable. Each study's primary reviewer used the subject labels to select the variables most aligned with their study's key content area. Each study's secondary reviewer reviewed the selected variables and resolved any disagreements with the primary reviewer. By examining the availability of variables within each study's key content area compared to the availability of all variables in the study, we are able to provide a more nuanced picture of how an SLDN could support future research in each of the key postsecondary topic areas.

There are a few limitations to the methods used in this review. Although the current text of the CTA, CAA, and CCRA provides the best available understanding of which data elements a future SLDN might contain, none of these bills have become law. The authors acknowledge that

[9] The College Affordability Act, section 1022(a)(I)(2)(F).

future revisions to any of these pieces of legislation may result in a different list of data elements, which could have implications for the results of this review.

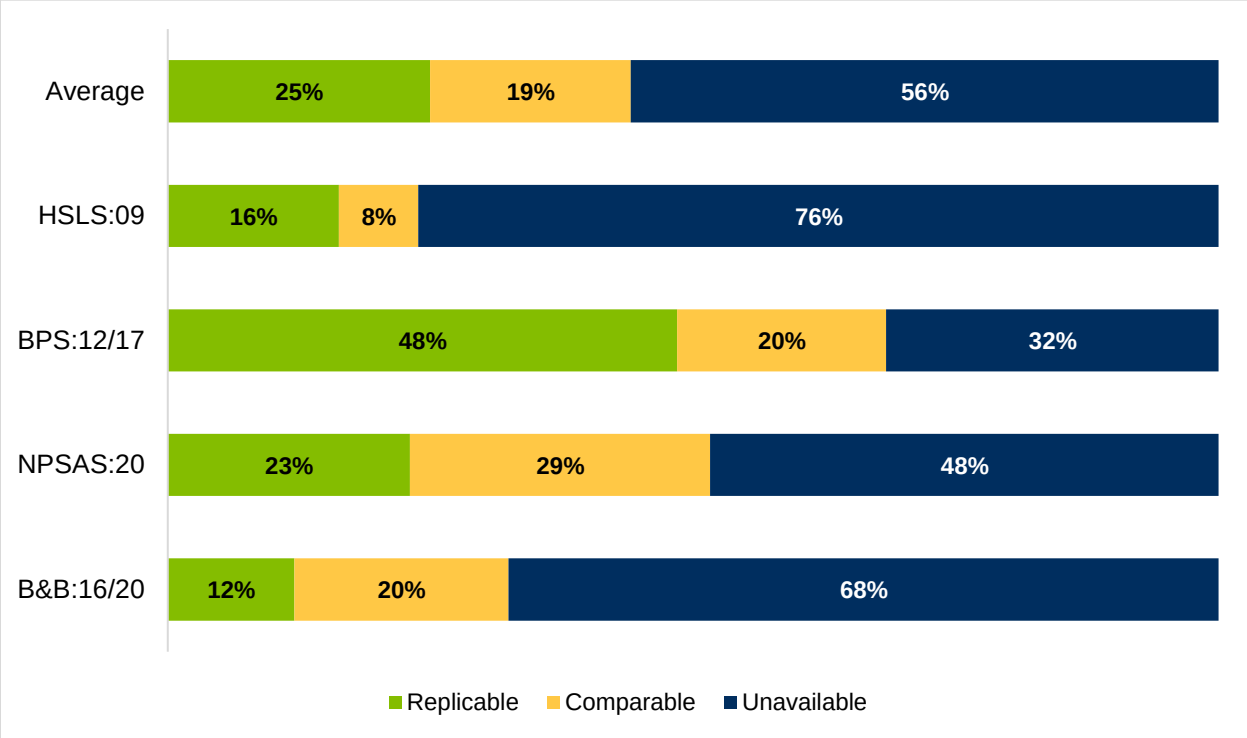
Additionally, there is a level of ambiguity in the text of the CTA, CAA, and CCRA that leaves room for interpretation. The current drafts of the legislation do not provide precise definitions for each data element. The list of SLDN data elements that was used as the basis for this review includes a variety of postsecondary data and policy experts' perspectives on how an SLDN could collect or define different data elements. The list of data elements generated by these experts contains ideas for both minimum and better options for collecting the data, as well as questions related to the construction, level of detail, and sources for different data elements (see appendix). This level of ambiguity impacted the review presented in this paper as well, particularly when deciding whether a variable would be replicable or comparable in a future SLDN based on the information available in the current drafts of the legislation. Efforts were made to agree upon coding decisions to the greatest extent possible, but a small level of error in the application of the codes *replicable* and *comparable* was accepted given the various interpretations available in the legislation and in the list of data elements prepared by the experts.

Results

The analysis indicates that on average, across all four studies reviewed, 48 percent of a study's variables would be either replicable or comparable within an SLDN collection. Of all the studies, HSLS:09 had the smallest percentage of data elements replicable or comparable with an SLDN, at 24 percent (16 percent replicable and 8 percent comparable). BPS:12/17 had the largest percentage of data elements replicable or comparable with an SLDN at 68 percent (48 percent replicable and 20 percent comparable). **Figure 1** displays the estimated distribution of replicable and comparable variables by study.

The analysis indicates that on average, across all four studies reviewed, 48 percent of a study's variables would be either replicable or comparable within an SLDN collection.

Figure 1. Frequency distribution of variable availability in an SLDN, by study



NOTE: Variables were coded as *replicable* when the listed SLDN data elements could *reproduce* the operationalization of the study variable, regardless of whether one or more SLDN data elements would be needed to do so. Variables were coded as *comparable* when the current list of SLDN data elements provides a *similar but not replicable* measure of the study variable in question. Variables were coded as *unavailable* when no SLDN data element(s) aligned with the study measure. Coding assignments were determined through independent dual reviews and in-depth team discussions to resolve disagreements.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 second follow-up, 2012/17 Beginning Postsecondary Student Longitudinal Study, 2019–20 National Postsecondary Student Aid Study, and 2016/20 Baccalaureate and Beyond Longitudinal Study; the College Affordability Act, H.R. 4674, 116th Congress, 2019; Velez, E. D., Pretlow, J., and Roberson, A. J. (2020) *Implementing a Federal Student-Level Data Network: Advice from Experts*.

The following sections provide details about the estimation of data availability overall and by key content area for each study.

High School Longitudinal Study of 2009 Second Follow-Up

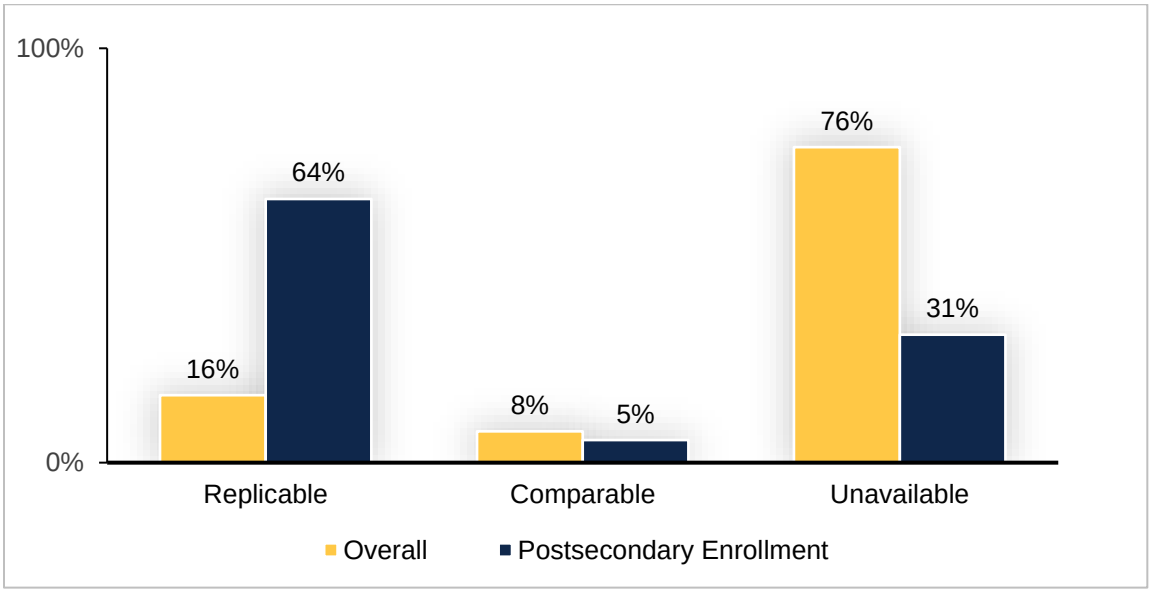
Of the 344 variables reviewed for the HSLS:09 second follow-up, an estimated 24 percent would be either replicable (16 percent) or comparable (8 percent) in an SLDN, and 76 percent of all variables would be unavailable in an SLDN.

The overarching focus of HSLS:09 was the relationship between high school experiences and academic performance and the pathways students take after high school. For the second follow-up, which took place when most sample members were 3 years out of high school, the primary focus areas were access to postsecondary education, choice of postsecondary institution, and attainment of subbaccalaureate credentials. Therefore, the key content area analyzed was

postsecondary enrollment.[10] Of the 344 HSLS:09 variables considered for this report, 16 percent (*n* = 55) relate to the key content area of postsecondary enrollment. As shown in **Figure 2**, about 64 percent (*n* = 35) of the variables related to postsecondary enrollment would be replicable with an SLDN, and 5 percent (*n* = 3) would be comparable. Policymakers and researchers would largely be able to answer questions about postsecondary enrollment with an SLDN, including questions about enrollment status, retention and transfer, characteristics of the chosen institution, major field of study, and enrollment in any remedial course. However, 31 percent (*n* = 17) of the variables related to postsecondary enrollment from the HSLS:09 second follow-up would be unavailable. These 17 unavailable variables pertain to subject-specific postsecondary course taking, the major field of study students considered at first enrollment compared with the field of study of their completed credential, and the characteristics they considered when choosing an institution to attend.

For the HSLS:09 second follow-up, an estimated 24 percent of the variables would be either replicable (16 percent) or comparable (8 percent) using an SLDN.

Figure 2. Estimated availability of data elements in an SLDN, overall and by key content area (postsecondary enrollment): HSLS:09 second follow-up



NOTE: Of the 344 variables reviewed, 16 percent (*n* = 55) were categorized within the key content area of *Postsecondary Enrollment*. Variables were coded as *replicable* when the listed SLDN data elements could *reproduce* the operationalization of the study variable, regardless of whether one or more SLDN data element would be needed to do so. Variables were coded as *comparable* when the current list of SLDN data elements provides a *similar but not replicable* measure of the study variable in question. Variables were coded as *unavailable* when no SLDN data element(s) aligned with the study measure. Coding assignments were determined through independent dual reviews and in-depth team discussions to resolve disagreements.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 second follow-up; the College Affordability Act, H.R. 4674, 116th Congress, 2019; Velez, E. D., Pretlow, J., and Roberson, A. J. (2020), *Implementing a Federal Student-Level Data Network: Advice from Experts*.

[10] HSLS:09 second follow-up variables were considered *postsecondary enrollment* variables if they were assigned one of the following subject labels in the study codebook: *Postsecondary Enrollment*, *Postsecondary Important Characteristics*, *Institution Characteristics*, *Postsecondary Coursetaking*, *Transfer*, or *Education: Majors*.

Many of the data elements in the HSLS:09 second follow-up are about students' choices (e.g., why they chose their postsecondary institution, field of study, or career) and students' perceptions (e.g., perceptions of their own abilities and perceptions of their instructors). These data elements can only be captured by a survey of students. Additionally, some data elements in the HSLS:09 second follow-up referred to the students' high school (e.g., high school location, high school credential, courses, grades). The legislation, as currently drafted, prohibits an SLDN from including elementary and secondary school data.[11] The results of the HSLS:09 review in this report emphasize the importance of survey research in addition to administrative data collections—information about students' choices, perceptions, and beliefs would be lost without survey research. Moreover, the longitudinal data on students' high school experiences as they correlate with their decisions to pursue postsecondary education would be unavailable in an SLDN due to the prohibition on collecting K-12 data in the legislation. Variables from the base year and first follow-up of HSLS:09, which took place while sample members were in high school, were excluded from this analysis, since none would be available in an SLDN; but it is worth noting that without these surveys, research questions pertaining to high school experiences that influence students' decisions to pursue postsecondary education would be challenging to answer.

Nevertheless, the results of this analysis also show that basic information such as enrollment status, major field of study for completed degrees, and transfer status could be obtained from an SLDN. Those questions could perhaps be eliminated from future iterations of a survey, thereby reducing the burden on student respondents.

2012/17 Beginning Postsecondary Student Longitudinal Study

The analysis for BPS:12/17 included 553 variables. An estimated 68 percent of variables would be either replicable (48 percent) or comparable (20 percent) in an SLDN, and 32 percent of all variables would be unavailable in an SLDN.

For BPS:12/17, an estimated 68 percent of the variables would be either replicable (48 percent) or comparable (20 percent) using an SLDN.

A focus of BPS was tracking first-time beginning postsecondary students over time, specifically their persistence and degree attainment, transition to employment, and school and work experiences. The key content area examined for this report was persistence and attainment.[12] Of the 553 BPS: 12/17 variables analyzed for this study, 32 percent (*n* = 178) relate to the key content area of persistence and attainment. As shown in **Figure 3**, 44 percent (*n* = 79) of the variables related to persistence and attainment would be replicable with an SLDN, and an additional 49 percent (*n* = 87) would be comparable. Policymakers and researchers would be able to use an SLDN to replicate measures of degree attainment and yearly attendance and transfer status over time. Only 7 percent (*n* = 12) would be unavailable within an SLDN, and

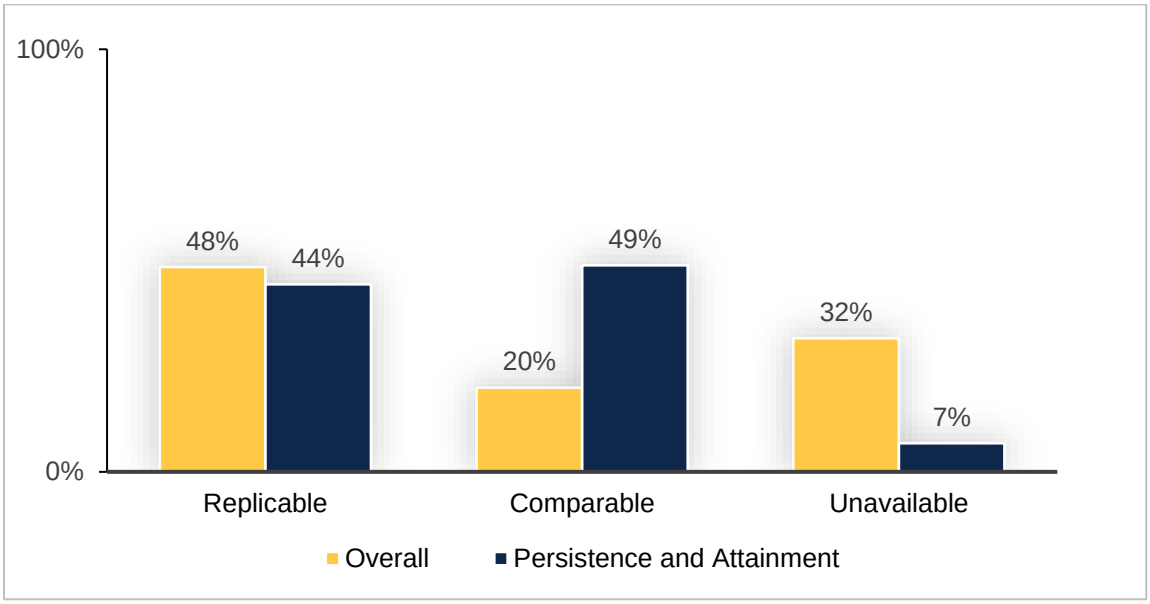
[11] The College Affordability Act, section 1022(a)(2)(F)

[12] BPS:12/17 variables were considered *persistence and attainment* variables if they were assigned one of the following subject labels in the study codebook: *Education Experiences*, *Attainment*, *Attendance*, *Persistence*, *Stopout*, or *Transfer*.

these variables pertain to enrollment plans, the accuracy of students' expected degree attainment, and attempts to transfer credits between institutions.

Many of the variables that track a student's postsecondary persistence and attainment over time rely on administrative data sources for derivation, which would likely also be available in an SLDN. For example, the priority attainment variable *first year attained degree through June 2017* is derived using survey sources along with student records and National Student Clearinghouse data and aligns to the SLDN data elements *credential conferred* and *when award was conferred*. Although administrative data effectively track student progress while enrolled, detailed information describing respondent employment and finances after leaving the institution would be lost without the BPS student interview, since detailed information about those post-graduation outcomes are not currently slated for inclusion in an SLDN.

Figure 3. Estimated availability of data elements in an SLDN, overall and by key content area (persistence and degree attainment): BPS:12/17



NOTE: Of the 553 variables reviewed, 32 percent (*n* = 178) were categorized within the key content area of persistence and degree attainment. Variables were coded as *replicable* when the listed SLDN data elements could *reproduce* the operationalization of the study variable, regardless of whether one or more SLDN data element would be needed to do so. Variables were coded as *comparable* when the current list of SLDN data elements provides a *similar but not replicable* measure of the study variable in question. Variables were coded as *unavailable* when no SLDN data element(s) aligned with the study measure. Coding assignments were determined through independent dual reviews and in-depth team discussions to resolve disagreements.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study; the College Affordability Act, H.R. 4674, 116th Congress, 2019; Velez, E. D., Pretlow, J., and Roberson, A. J. (2020), *Implementing a Federal Student-Level Data Network: Advice from Experts*.

2019–20 National Postsecondary Student Aid Study

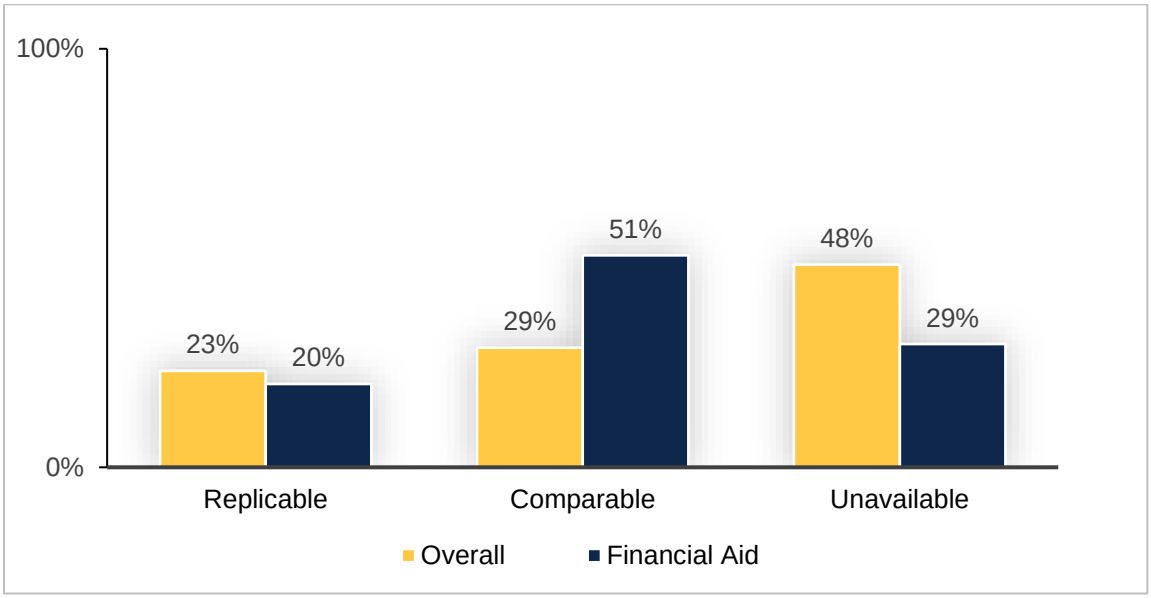
The NPSAS:20 analysis included 560 variables to be reviewed for this report. In addition to methodological variables, this analysis excluded a series of one-time variables added to measure financial aid and other assistance institutions provided during the COVID-19 pandemic, which began during the data collection period of this study. Of the 560 variables

For NPSAS:20, an estimated 52 percent of the variables would be either replicable (23 percent) or comparable (29 percent) using an SLDN.

reviewed for NPSAS:20, an estimated 52 percent would be either replicable (23 percent) or comparable (29 percent) in an SLDN.

The overarching focus of NPSAS was how students finance their postsecondary education. Therefore, the key content area selected for the NPSAS:20 analysis was financial aid.[13] Of the 560 NPSAS:20 variables analyzed, 41 percent (*n* = 231) related to the key content area of financial aid. As shown in **Figure 4**, 20 percent (*n* = 45) of the variables related to financial aid would be replicable with an SLDN, and 51 percent (*n* = 117) would be comparable. The higher percentage of variables coded as *comparable* rather than *replicable* is due to the complex calculations used to produce many of the financial aid variables in the NPSAS:20 data, using multiple data points and sources that may not become part of an eventual SLDN. It is uncertain whether many financial aid estimates could be created in a manner exactly consistent with prior editions of NPSAS.

Figure 4. Estimated availability of data elements in an SLDN, overall and by key content area (financial aid): NPSAS:20



NOTE: Of the 560 variables reviewed, 41 percent (*n* = 231) were categorized within the key content area of financial aid. Variables were coded as *replicable* when the listed SLDN data elements could *reproduce* the operationalization of the study variable, regardless of whether one or more SLDN data element would be needed to do so. Variables were coded as *comparable* when the current list of SLDN data elements provides a *similar but not replicable* measure of the study variable in question. Variables were coded as *unavailable* when no SLDN data element(s) aligned with the study measure. Coding assignments were determined through independent dual reviews and in-depth team discussions to resolve disagreements.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2019-2020 National Postsecondary Student Aid Study; the College Affordability Act, H.R. 4674, 116th Congress, 2019; Velez, E. D., Pretlow, J., and Roberson, A. J. (2020), *Implementing a Federal Student-Level Data Network: Advice from Experts*.

[13] NPSAS:20 variables were considered *financial aid* variables if they were assigned one of the following subject labels in the study codebook: *Financial Aid: Application*, *Financial Aid: Borrowed Cumulative*, *Financial Aid: Federal Grants*, *Financial Aid: Federal Loans*, *Financial Aid: Grants*, *Financial Aid: Institutional*, *Financial Aid: Loans*, *Financial Aid: Need*, *Financial Aid: Net Price*, *Financial Aid: Other*, *Financial Aid: Package*, *Financial Aid: Ratios*, *Financial Aid: State*, or *Financial Aid: Total*.

Traditionally, NPSAS collected dollar amount, program name, aid source (e.g., federal, state, institution), and aid type (e.g., merit- or aid-based, assistantship) for each award received by each sampled student. The CAA, CTA, and CCRA outline financial aid data in broad terms, so the eventual level of granularity is uncertain. For example, the legislation explicitly requires that the SLDN capture the amount of Pell Grant funding received by each student, but other federal grant programs (e.g., the Federal Supplemental Education Opportunity Grant) are not explicitly named in the bill, so it cannot be assumed that these aid sources would be collected or integrated from NSLDS. Twenty-nine percent (*n* = 68) of financial aid variables would be unavailable in a future SLDN. These 68 variables largely relate to financial aid applications and sources of financial aid such as work study, graduate assistantships, and employer assistance. Because these types of financial aid are not currently slated to be collected in an SLDN, total aid received, net cost at attendance, and related variables would consequently also be unavailable with an SLDN alone.

Because NPSAS was a repeated cross-sectional study that emphasized replicability of results over time, these potential inconsistencies in the formulation of financial aid estimates mean that the SLDN may not be an adequate source for all NPSAS measures. This underscores the importance of using survey research to complement administrative data collections.

2016/20 Baccalaureate and Beyond Longitudinal Study

Of the 432 analysis variables reviewed for B&B:16/20, an estimated 32 percent of the variables would be either replicable (12 percent) or comparable (20 percent) in an SLDN, and 68 percent would be unavailable in an SLDN.

For B&B:16/20, an estimated 32 percent of the variables would be either replicable (12 percent) or comparable (20 percent) using an SLDN.

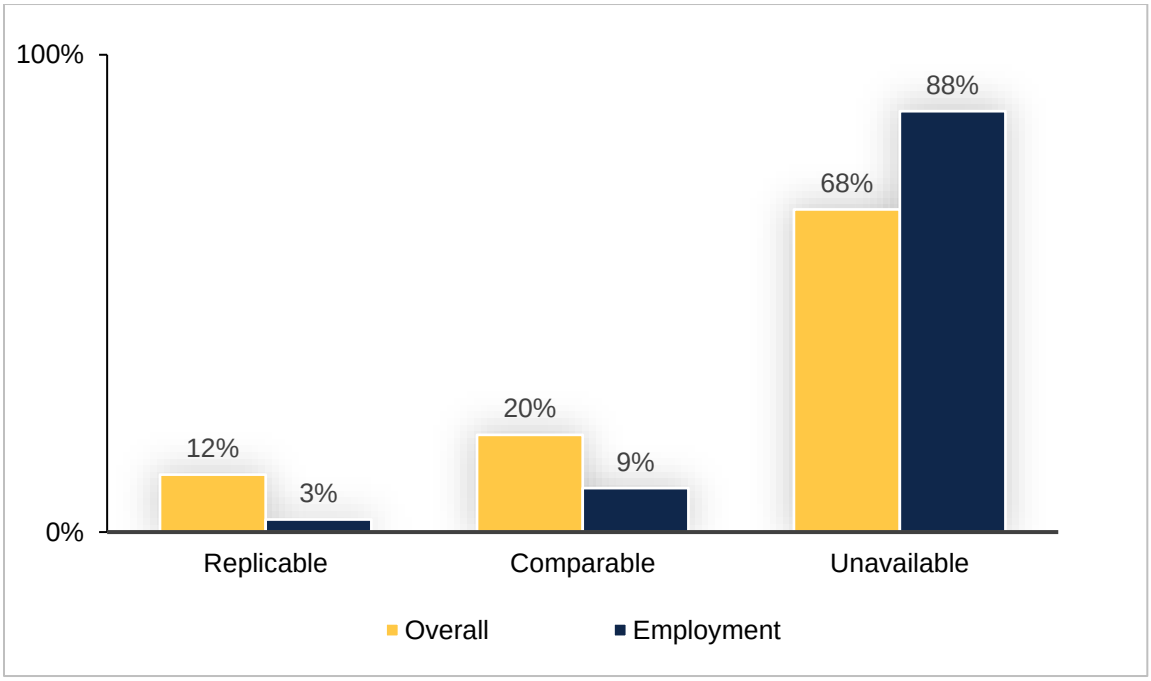
B&B focused on postbaccalaureate outcomes with a special emphasis on employment. Thus, the key content area examined for B&B:16/20 was employment.[14] Of the 432 B&B:16/20 variables analyzed for this report, 18 percent (*n* = 76) relate to the key content area of employment. As shown in **Figure 5**, 3 percent (*n* = 2) of the variables related to employment would be replicable with an SLDN, and 9 percent (*n* = 7) would be comparable. Policymakers and researchers would be able to collect basic information about employment status, occupation title, and aggregate earnings from an SLDN. However, 88 percent (*n* = 67) of employment variables would be unavailable in an SLDN. These 67 variables pertain to dates of employment and time spent in a career, pay and pay changes over time, employer-provided benefits, type of employer, and job satisfaction.

Many B&B data elements pertain to postsecondary graduates' experiences at work, family formation, household finances, and financial well-being, which are data elements that are not typically found in administrative postsecondary education data sources and would only be available with a respondent survey. The results of the B&B:16/20 analysis again emphasize how an SLDN and a survey could complement each other: an SLDN could lower the burden on

[14] B&B:16/20 variables were considered *employment* variables if they were assigned one of the following subject labels in the study codebook: *Employment: Description*, *Employment: Employer*, or *Employment: History*.

survey respondents by eliminating questions that could be found in the SLDN, such as postbaccalaureate education details. However, survey data would also be needed to provide a complete picture of postsecondary graduates' experiences and outcomes, which was the key contribution of the B&B study.

Figure 5. Estimated availability of data elements in an SLDN, overall and by key content area (employment): B&B:16/20



NOTE: Of the 432 variables reviewed for the overall study, 18 percent (n=76) were categorized within the key content area of employment. Variables were coded as *replicable* when the listed SLDN data elements could *reproduce* the operationalization of the study variable, regardless of whether one or more SLDN data element would be needed to do so. Variables were coded as *comparable* when the current list of SLDN data elements provides a *similar but not replicable* measure of the study variable in question. Variables were coded as *unavailable* when no SLDN data element(s) aligned with the study measure. Coding assignments were determined through independent dual reviews and in-depth team discussions to resolve disagreements.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2016/20 Baccalaureate and Beyond Longitudinal Study; the College Affordability Act, H.R. 4674, 116th Congress, 2019; Velez, E. D., Pretlow, J., and Roberson, A. J. (2020), *Implementing a Federal Student-Level Data Network: Advice from Experts*.

Discussion

For decades, NCES sample surveys have provided insights into education pathways, financial aid, and student outcomes across the country, informing student needs and policy development. Although an SLDN would require additional effort from postsecondary institutions at its inception, over time it could reduce burden by streamlining the reporting required of each Title IV institution and supplementing parts of the student surveys. The SLDN could act as a sampling frame, thus reducing the burden of student list collection requests from each study. Additionally, the review conducted for this report demonstrates that researchers could obtain a portion of the data elements (on average, about 48 percent) collected by the sample surveys from an SLDN, which could reduce the length of the surveys and the burden for institution staff and students who respond to them.

Although an SLDN has the potential to reduce burden and increase efficiencies in data reporting, administrative data does not replace the rich context provided by the student survey portion of the NCES sample studies. Survey questions are the sole source for many key measures of a student's educational pathway, including their choices, their reasoning for making those choices, their postsecondary and postgraduate life experiences, and their beliefs and attitudes about their education. For example, while administrative data are limited to reporting the academic program chosen by a sampled student, a survey reveals a student's reasons for selecting the program, whether they felt supported by their institution, and whether the program led to a fulfilling career after completion. Though student-level administrative data could have positive impacts on the postsecondary data ecosystem, an SLDN is not a substitute for the power of student perspective, and a great deal of context on postsecondary experiences and transitions to the labor market would be lost without the student surveys.

References

H.R.4674 - 116th Congress (2019-2020): College Affordability Act. (2025, July 16). https://www.congress.gov/bill/116th-congress/house-bill/4674

U.S. Department of Education. (2025, March 11). *U.S. Department of Education Initiates Reduction in Force* [Press release]. Last updated June 13, 2025. https://www.ed.gov/about/news/press-release/us-department-of-education-initiates-reduction-force

Velez, E. D., Pretlow, J., & Roberson, A. J. (2020). *Implementing a federal student-level data network: Advice from experts*. RTI International. https://www.rti.org/publication/implementing-federal-student-level-data-network-advice-experts/fulltext.pdf

Appendix

Data Elements that Sec. 1022 of the College Affordability Act (H.R. 4674) Requires to be Included in a Postsecondary Student-level Data Network: 2019-2020

Ref Num.	Data Element	Minimum to Meet Legislation	Better Option to Meet Legislation	Current IPEDS Component	Ref. in Legislation	Source	Questions
Enrollment							
1	Whether student was enrolled	Binary indicator of enrollment in defined time period (e.g., year, terms)	Credits attempted in defined time period (e.g., year, terms); credits earned in defined time period (e.g., year, terms)	EF, E12, SFA, OM, GR/GR200	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	What is the periodicity of the data (both the increments of the data [e.g., every term] and how often institutions report data)? Should the new system collect non-credit enrollment?
2	Attendance intensity	Full-time/part-time, by term	Computed from credits attempted	EF, E12, SFA, OM, GR/GR200	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	
3	Program of study/major	CIP for major(s) 1st/2nd, once a year		EF, C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	
4	Credential-seeking status	For IPEDS: yes/no For legislation: credential sought (e.g., cert, AA, BA, grad) in defined time period (e.g., years, terms)		EF, E12, SFA, OM, C, GR/GR200	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	(Related to this item and others:) Do institutions just want to report some elements when they change, or report the same data each time?
5	Student level	Undergraduate vs. graduate vs. professional vs. non-degree, by term		EF, E12, SFA, OM, GR/GR200	Pg. 48, line 20 – pg. 49, line 7	Institutions	
6	Permanent residence	State/territory		EF	Pg. 48, line 20 – pg. 49, line 7	Institutions	
7	Graduated high school within past 12 months	Yes/no	HS graduation date	EF	Pg. 48, line 20 – pg. 49, line 7	Institutions	Should high school graduation date be collected instead? Would high school graduation date be prohibited as "secondary education data"?
8	Enrolled in distance education	Distance education all/ none/some across all courses, yearly	Distance education credits attempted, in defined time period (e.g., years, terms)	EF, C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	Should distance education credits earned be collected? Should the SLDN capture distance education that is not 100%?
9	Retention/ persistence	Derived; no collection needed		EF	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	

Ref Num.	Data Element	Minimum to Meet Legislation	Better Option to Meet Legislation	Current IPEDS Component	Ref. in Legislation	Source	Questions
10	Transfer	Derived; no collection needed		EF, GR/GR200, OM	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	
11	Enrollment status (first time, recent transfer, other non-first-time)	Derived; no collection needed		EF, GR/GR200, SFA, OM	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	
12	First time at this institution	Derived; no collection needed		EF, GR/GR200, SFA, OM	Pg. 48, line 20 – pg. 49, line 7	Institutions	
13	Participation in remedial coursework	Yes/no	Credits attempted in defined time period (e.g., years, terms)		Pg. 51, line 15 – pg. 52, line 3	Institutions	Should remedial credits earned be collected? How is "remedial course" defined?
Completion							
14	Credential conferred	Credential conferred (e.g., certificate, AA, BA, MA), spring		GR/GR200, OM, C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	
15	Completed within 100% of normal time	Yes/no	Length of program for award conferred	GR/GR200	Pg. 48, line 20 – pg. 49, line 7	Institutions	Should length of time be calculated within program, not just within institution?
16	Completed within 150% of normal time	Yes/no	Length of program for award conferred	GR/GR200	Pg. 48, line 20 – pg. 49, line 7	Institutions	
17	Completed within 200% of normal time	Yes/no	Length of program for award conferred	GR/GR200	Pg. 48, line 20 – pg. 49, line 7	Institutions	
18	When award was conferred	Academic year conferred	Month and year conferred	GR/GR200, OM, C	Pg. 48, line 20 – pg. 49, line 7	Institutions	
19	CIP of awarded major	CIP for awarded major(s)		C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	Should CIP for minors be collected (i.e., many education programs, like "math education" are technically "math major, education minor")?
Financial Aid/Financial Indicators							
20	Pell Grant	Amount awarded yearly	Additionally, amount disbursed yearly	SFA, GR, OM	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	NSLDS	
21	State/local grants	Amount awarded yearly	Additionally, amount disbursed yearly	SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	Should state and local sources be reported separately? Should whether the aid is need based or non–need based be collected? Should disbursed grant amounts be reported more frequently than once a year?

Ref Num.	Data Element	Minimum to Meet Legislation	Better Option to Meet Legislation	Current IPEDS Component	Ref. in Legislation	Source	Questions
22	Institution grants	Amount awarded yearly	Additionally, amount disbursed yearly	SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	Should whether the aid is need based or non–need based be collected? Should disbursed grant amounts be reported more frequently than once a year?
23	Grants from third parties (private)	Amount awarded yearly	Additionally, amount disbursed yearly	SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	Should whether the aid is need based or non–need based be collected? Should disbursed grant amounts be reported more frequently than once a year?
24	Federal loans	Amount disbursed yearly	Amount disbursed, by loan type (Sub Stafford, Unsub Stafford, Perkins, Grad PLUS), yearly	SFA, GR, OM	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	NSLDS	Should Parent PLUS loans be reported?
25	Non-federal loans	Amount disbursed yearly		SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	Note: This is in CAA not CTA. What nonfederal loan data do institutions currently have? How comprehensive and reliable is institution data?
26	In-state/out-of-state tuition flag	In-district/in-state/out-of-state		SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	
27	Title IV flag	Title-IV yes/no		SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	Should work-study amount awarded be collected?
28	Post 9-11 GI Bill	Amount disbursed yearly		SFA	Pg. 48, line 20 – pg. 49, line 7	VBA/ Institutions	Should awarded amounts be reported?
29	DoD TAP aid	Amount disbursed yearly		SFA	Pg. 48, line 20 – pg. 49, line 7	DoD/ Institutions	Should this be collected from institutions or federal matching?
30	Military or veteran benefit status	Yes/no			Pg. 49, line 8 – pg. 50, line 22	VBA/DoD	How is status defined? What benefits are included? Are children/spouses using others' benefits included? Should the amount of aid disbursed be collected?
31	Cumulative student debt	Current outstanding federal and nonfederal balance (principle + interest)			Pg. 53–54	NSLDS	How is the nonfederal part of this collected? Should this element be limited to federal borrowing, which can be captured in NSLDS?
32	Loan repayment status	Loan repayment status (in repayment, deferment/forbearance, default, all loans paid off)			Pg. 53–54	NSLDS	What are the time intervals for collection? How should different statuses across loans be combined? Should percentage of principle repaid be collected?

Ref Num.	Data Element	Minimum to Meet Legislation	Better Option to Meet Legislation	Current IPEDS Component	Ref. in Legislation	Source	Questions
33	Repayment plan	Repayment plan: standard, graduated, income based (e.g., PAYE, REPAYE), other			Pg. 53–54	NSLDS	What are the time intervals for collection? How should different plans across loans be collected?
Demographics							
38	Age	Categorical age range	DOB	EF, C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	Would DOB be needed for matching? General demographic question: Should dependent/independent status be collected for FAFSA filers (or everyone)?
39	Gender	Male/female		EF, E12, C, GR/GR200, OM, SFA	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	CAA includes a pilot study for how to measure gender.
40	Race	Race/ethnicity categories	Race with multiple additional detailed categories	EF, GR/GR200, OM, C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	Note: Needs to be reported as of Section 153(a)(3)(b) of ESRA
41	Ethnicity	Race/ethnicity categories	Hispanic yes/no	EF, GR/GR200, OM, C	Pg. 48, line 20 – pg. 49, line 7; pg. 49, line 8 – pg. 50, line 22	Institutions	
42	First generation college student status	First-gen yes/no			Pg. 51, line 15 – pg. 52, line 3	FAFSA	How should this be defined? How is this collected for non-FAFSA filers?
43	Economic status	Derived; no collection needed (derived from Pell Grant status)	Household income amount for everyone who submits a FAFSA		Pg. 51, line 15 – pg. 52, line 3	FAFSA	How should economic status be defined and reported? Should this be collected for non-FAFSA filers?
44	Veteran status	Yes/no			Pg. 53–54	VA	What level of detail do stakeholders want? Should dates of service or branch be collected?
45	Military status	Yes/no			Pg. 53–54	DOD	What level of detail is wanted by stakeholders? Should dates of service or branch be collected?
46	Household income	Household income range for aided students	Household income amount for everyone (or all FAFSA filers)	SFA	Pg. 48, line 20 – pg. 49, line 7	FAFSA	How is this collected for non-FAFSA filers?
47	Living arrangement	For aided students: on-campus, off-campus w/fam, off-campus w/o-fam, for aided students	For all students (or all FAFSA filers): on-campus, off-campus w/fam, off-campus w/o-fam, for aided students	SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	

Ref Num.	Data Element	Minimum to Meet Legislation	Better Option to Meet Legislation	Current IPEDS Component	Ref. in Legislation	Source	Questions
Post-completion outcomes							
34	Aggregate earnings by institution and program	Aggregate earnings by institution and program			Pg. 53–54	IRS, Census	What are the time intervals for collection?
35	Employment	Yes/no			Pg. 53–54	IRS	
36	Occupation				Pg. 53–54	Census	How would this be measured?
37	Further education	Derived; no collection needed			Pg. 53–54	Institutions	
Institutional characteristics							
48	Institutional level	Highest postsecondary credential awarded (e.g., 1-year certificate, AA, BA, doctorate)		IC	Pg. 49, line 8 – pg. 50, line 22	Institutions	
49	Institutional control	Public, private not-for-profit, private for-profit		IC	Pg. 49, line 8 – pg. 50, line 22	Institutions	
50	Institutional predominant degree awarded	Most frequently awarded postsecondary credential (e.g., 1-year certificate, AA, BA, doctorate)		IC	Pg. 49, line 8 – pg. 50, line 22	Institutions	
52	Cost of attendance	Published "sticker price" of annual full-time enrollment including tuition, fees, room, board, and other expenses	Cost of attendance (tuition, fees, room, board, and other expenses) calculated for each student	SFA	Pg. 48, line 20 – pg. 49, line 7	Institutions	Can institutions provide information on cost of attendance at the student level?
Components to create a student unique identifier							
53	Student ID	Fuzzy match using name, date of birth, etc.	Student ID (SSN)/TIN			Institutions	Do all schools have SSN/TIN? What information will be needed to create a federal identifier?

NOTE: AA = Associate of Arts; BA = Bachelor of Arts; C = Completions; CAA = the College Affordability Act; CIP = Classification of Instructional Programs code; CTA = the College Transparency Act; DoD = U.S. Department of Defense; EF = Fall Enrollment; E12 = 12-Month Enrollment; FAFSA = Free Application for Federal Student Aid; GR/GR200 = Graduation Rates/Graduation Rates 200%; IC = Institutional Characteristics; IPEDS = Integrated Postsecondary Education Data System; IRS = Internal Revenue Service; MA = Master of Arts; NSLDS = National Student Loan Data System; OM = Outcome Measures; PAYE = Pay As You Earn; SAVE = Saving on a Valuable Education; SFA = Student Financial Aid; SLDN = Student-Level Data Network; SSN = Social Security Number; TIN = Tax Identification Number; VA = U.S. Department of Veterans Affairs