

The Ideal State Postsecondary Data System 15 Essential Characteristics and Required Functionality

Written by: Peter Ewell and Hans L'Orange

A rapidly changing global economy, shifting demographics, and concerns about our ability to maintain a competitive workforce have focused national attention on the educational systems of America's states, highlighting their critical role in ensuring a productive and creative future for our country. As a result, America's colleges and universities are expected to perform better than ever to prepare all students to meet internationally competitive standards of achievement, the knowledge and skills needed for a vital democracy, and the escalating demands of the workforce.

Central in addressing these needs are robust and aligned data systems able to document the achievement of students, schools and colleges, as well as respond to questions about a state's investment in higher education. Questions that can be fruitfully examined using data drawn from such systems include the following:

- What does the overall “flow” of students through the “educational pipeline” look like?
- What experiences (curricular or environmental) affect student success in making progress through the “educational pipeline”?
- What facilitates successful student transitions across specific boundaries—for example from high school to college, from two-year colleges to four-year colleges, or from either of these to and/or from the workplace? How are these transitions different for different types of students?
- What role does geographic mobility (e.g., transfer) play in inhibiting or enhancing educational credentialing or attainment?

There is a growing policy imperative to address these educational pipeline issues. Attention must also be paid to data systems able to monitor student progress across the K-16 continuum, enable early assessment leading to effective intervention strategies, and locate barriers within systems. Most of the important public policy and research questions related to this imperative can *only* be addressed through the use of Student Unit Record (SUR) systems. These systems contain an individual electronic record for each student enrolled in an institution or school for each term or year. Current SUR databases contain millions of individual student records covering the majority of educational activity in the nation. Such longitudinal systems collect a range of demographic and performance data at regular intervals. These records can be merged with other longitudinal files to investigate student success and behavior over time, and across settings and treatments. They also can be analyzed to examine progress made and impacts on particular populations.

Although each state's higher education system is unique, certain characteristics and resulting functionality are **essential** in an effective longitudinal data system. Like the "Ten Essential Elements" of state longitudinal data systems in K-12 education proposed by the Data Quality Campaign (DQC), these characteristics are intended to promote educational progress and alignment among and between each state's K-12 and postsecondary data resources.

The characteristics and functions are grouped into four broad categories: student data, course data, operational characteristics, and data governance. The "state of the art," or ideal state system, is optimized around all four categories that reinforce one another. Course data have limited value without student data; the presence of student and course data enables the recommended characteristics of operational functionality. Without data governance, the data and functionality have limited value. There is no implied sequence or hierarchy represented by the four categories, but each state needs to address each of them to create a system that meets its goals. Developing the ideal state system is certainly challenging but the characteristics of such a system are presented here as a model that can and should be reached.

STUDENT DATA:

1. ***A Unique Statewide Student Identifier.*** A unique statewide student identifier is a single, non-duplicated number that is assigned to and remains with a student throughout his or her educational career. A student identifier allows the state to follow the enrollment of each student over time across multiple academic programs and institutions. It also makes it possible to identify information about a single student across various data sources (e.g. enrollment, employment, financial aid, etc.) in order to evaluate relationships between program participation and performance, and to study student mobility patterns. A unique identifier also makes it more feasible for a student to retrieve data about their own educational activity from multiple sources.

While the Social Security Number (SSN) is currently used by most states and can function effectively as a unique identifier, states should take steps to ensure that it is protected, encrypted, and that an alternative identifier is eventually developed. Alternatives should be considered; the IPEDS Feasibility Study commissioned in 2004 suggested an identifier cryptically created from characteristics unique to each student (e.g. gender, race, ethnicity, data and location of birth). Even when a new student identifier is created, the SSN should continue to be collected for use in matching student records to other databases. However, SSNs must be encrypted to protect privacy.

States should also work with the state's K-12 systems to determine whether and how to incorporate K-12 unique student identifiers into the higher education data system. The ability to develop interoperability between a state's K-12 and postsecondary data systems is critical to addressing the educational needs that transcend both systems.

2. ***Student-Level Enrollment, Degree Completion, and Demographic Data for All Public Colleges and Universities.*** Accurate data on student enrollments and degree completion is essential to evaluate the effectiveness of the higher education system and the institutions that comprise it. Correct student demographic information is also critical for accurate disaggregation of performance measures to determine what is working for whom.

At a minimum, the following data elements should be included:

Demographics:

- Gender
- Race/ethnicity
- Date of birth
- Citizenship
- Geographic origin (state, county, zip code, etc.)
- Residency

Student Enrollment/Completion Data:

- Degree-seeking status
- Credits attempted (full-time/part-time status)
- Credits completed
- Program major
- Degree awarded
- Degree field of study

Definitions for these core data elements should be consistent with definitions used in the Integrated Postsecondary Data System (IPEDS) or, if IPEDS definitions are not available, with definitions used in the Common Data Set.

The data system should contain at least one record per student per enrollment period (term, quarter, semester) for every public institution in the state.

3. ***Student-Level Financial Aid Data.*** Access to or loss of financial aid has been shown to be a powerful determinant of student persistence and academic success. As a result, data about levels of student support and how they change over time can be important in improving student progression and degree completion. At the same time, many states operate financial aid systems and it will be important to them to learn how effectively their funds are being invested. Finally, cost of attendance has become a major issue for consumers as well as policy-makers and an area of concern for institutional accountability. For all these reasons, including student-level financial aid data in a state postsecondary longitudinal database is a wise investment.

This will not be easy to accomplish; financial aid data can be difficult to access and use. Because they contain sensitive information about such things as family income, financial aid records require unusual attention to privacy and security. The manner in which financial aid records are kept also poses challenges because the typical record

structure involves a student unit record for each financial aid source a student taps. For this reason, many state databases simply carry flags that indicate whether or not a student is receiving aid from a given source, or the total amount of support (or percentage of need) addressed by that source. Finally, in order to fully reflect the cost of attendance question, data will be needed about federal, state, and institutional aid.

4. ***Student-Level Transfer Data.*** Enrollment “swirl” across institutions has become a major phenomenon in postsecondary education. According to NCES, more than two thirds of students who earn a baccalaureate degree have attended more than one institution and almost one in five have attended three or more. And these patterns of inter-institutional transfer are becoming increasingly non-traditional—from four-year to four-year and from four-year to two year, as well as the more standard two-year to four-year. As a result, the state postsecondary student database needs to support calculations of transfer rates to and from different kinds of institutions, as well as support analyses of the impact of transfer on student progression, academic performance, and eventual degree completion. It should also be capable of disaggregation to identify which populations are affected.
5. ***Student-Level Persistence and Graduation Data.*** Linking student-level records across time periods (terms, years) allows the progress of individual students to be followed over time. This, in turn, enables the calculation of such measures as degree-completion and persistence rates. Performance measures like these have become basic indicators of institutional and program effectiveness. They also provide a measure of consumer information as mandated, for example, by “Student Right-to-Know.” Established definitions for compiling such statistics are provided by the IPEDS Graduation Rate Survey (GRS) and, at minimum, a state’s postsecondary longitudinal data system should be able to calculate these statistics. It should also be capable of disaggregation to identify populations at risk. The ability to report on the degree completion and persistence rate for all students, regardless of where they started and their part-time/full-time status, should be a minimum requirement.

COURSE DATA:

6. ***Student-Level (1) Remediation Data and (2) Developmental Education Participation and Success Data.*** Large numbers of students now enter colleges and universities unprepared for college level work. This constitutes an enormous challenge and raises difficult issues of public policy because of the reluctance of public officials to “pay for the same education twice.”

As a result, state postsecondary longitudinal data systems should include the following additional data elements:

- Initial placement level in reading, writing and mathematics (at minimum “college-level” and “below college-level”). A student’s need for remediation can then be analyzed.
- Participation in developmental reading, writing and mathematics (at minimum one level or more below college level)

- Successful completion of developmental reading, writing and mathematics (at minimum one level or more below college level).

7. *Student-Level Course/Transcript-Level Data.* Data elements on enrollment and degree completion will allow states to construct essential measures of student progression and success that will be useful in developing sound policy. But longitudinal records will be even more useful if states can determine the details of academic performance for different kinds of students. Including course-level detail in a state postsecondary longitudinal data system will enable states to answer questions like:

- How important is passing “gatekeeper” courses in such subjects as English and mathematics to student progression and success?
- How well do students who participated in developmental mathematics do in subsequent college-level courses that require mathematics?
- Do students who transfer from one institution to another suffer a penalty in terms of academic performance in particular fields of study?

Adding course-level detail to a statewide database is a significant step because it greatly increases the number of unit records and complicates data governance requirements. Instead of maintaining one record per student per enrollment period, the database must contain one record per individual student course enrollment per enrollment period. Minimum data elements required for each course include:

- Course/Section identifier (which can map to subject, department, etc.)
- Credits enrolled
- Credits completed
- Grade (alternatively, pass/fail indicator)
- Mode of instruction (e.g. on-line course)
- Credit / Non-credit status

8. *Student-Level Data on Assessed Academic Achievement.* The issue of student academic achievement in the form of demonstrated competence is of growing salience for postsecondary education policy. Many states want data of this kind for accountability purposes and it can be an important part of consumer information in choosing which college to attend. Initiatives like the Voluntary System of Accountability (VSA), through which colleges and universities disclose aggregate performance on quality measures like standardized examinations has called further attention to this matter. States also want to know how participation in different courses of study affects academic outcomes and how these vary across different kinds of students. For all these reasons, consideration should be given to incorporating data on student learning outcomes into state postsecondary databases. In the near future, this will constitute a challenge because there is no single list of such outcomes or measures.

OPERATIONAL CHARACTERISTICS:

- 9. Privacy Protection for All Individually-Identifiable Student Records.*** The Family Educational Rights and Privacy Act (FERPA) and other federal/state privacy protections obligate those who use student-level educational records to keep them secure. This is particularly critical when the unique identifier is the SSN, which can be used to mark an individual identity. States should keep all individually-identifiable records in secure locations, accessible by only a few individuals with appropriate clearances. When all record matching has been accomplished, analytical files should be stripped of their unique identifiers. Finally, when analyses are undertaken, states should ensure that cell size restrictions are imposed so that individuals cannot be identified by selecting on specific combinations of data elements.

At the same time, privacy protection rules allow and encourage the use of student records in the aggregate to support research programs directed at improving instruction. This valuable resource should be utilized to the fullest extent possible. Policies including data sharing agreements, data use policies and system auditing practices must be in place that allow the appropriate use of unit record data at the same time states make the assurance of confidentiality a priority.

- 10. The Ability to Match Student Records with Data on K-12 Educational Activities.*** As state higher education systems and school systems work to align expectations in high school with the demands of postsecondary education, they need better information on school preparation and success in college-level work. Most states today do not have data systems that enable this two-way communication; in many cases, there are two separate data systems and they rarely can exchange information.

Addressing this condition does not necessarily demand that states create a single data system containing both secondary and postsecondary student records. But, if they continue to maintain separate data systems, states must move toward greater integration and alignment. Postsecondary institutions should provide annual feedback reports to high schools that indicate the success of their former students in such areas as college placement, persistence and graduation, and performance in key courses. In turn, postsecondary institutions need to know such things as whether or not a high school graduate has completed a rigorous college-preparatory curriculum. The exchange process should also include a mechanism for accommodating unique student identifiers in a consistent fashion.

Technical interoperability agreements between aligned but separate systems are a critical component. While the data issues between K-12 and postsecondary systems are frequently based on cultural differences and turf battles (e.g. who “owns” the data), agreement on technical standards and a common analytical culture provide a way to facilitate integration.

11. *The Ability to Match Student Records with Data on Employment.* Labor market outcomes for all students participating in postsecondary education are of considerable interest to state policymakers. State leaders need to be able to align enrollment and degree-granting patterns with regional job markets and meet areas of high occupational demand. Linking student records with existing state-level databases containing individual employment records allows states to construct measures of job placement and earnings by field of study and/or field of employment. It can also help state leaders determine how long it takes graduates or former students to gain employment and the individual return on investment associated with further education. Finally, it can help students determine their odds of gaining successful employment in the field(s) of their choice.

The most common data source for employment records are the Unemployment Insurance (UI) wage records maintained by all states. These contain quarterly data on individual earnings by industry, with individual records identified by SSN. The state postsecondary data system should be capable of matching enrollment records with these employment records in a secure environment, using the SSN as a key link.

Accomplishing this will be challenging for many states, but the ideal system will incorporate the ability to demonstrate the relationship between student achievement and state labor needs.

12. *Inclusion of Independent and For-Profit Institutions of Higher Education.* Most established state postsecondary databases began with public institutions only. This is because they were originally constructed to manage such things as enrollment-driven funding formulas, which required states to have accurate enrollment counts. Moving to longitudinal tracking, however, there are substantial advantages to including all institutions in a state regardless of control. In many states, private not-for-profit and for-profit institutions may garner substantial enrollment shares, so tracking students into and through these institutions has many advantages. The advantages for independent institutions also can be considerable, because they can follow their students as they enroll elsewhere. More importantly, they can demonstrate their return on investment to state officials, especially if they participate in state financial aid programs. For all these reasons, states should make every effort to include as many institutions as possible in their state postsecondary databases. This will prove exceptionally challenging but the most comprehensive state analyses will come from the most inclusive state data systems.

13. *A Single State-Level Student Unit Record (SUR) System for All Public Institutions.* Several states maintain separate SUR databases for different systems of postsecondary institutions (for example, one for four-year institutions and one for community colleges). While these may be effectively linked, the ideal system is designed around a single database environment to aid consistent data manipulation and analysis. This also obviates security and confidentiality concerns associated with transferring records from one place to another.

DATA GOVERNANCE:

14. *Data Audit System Assessing Data Quality, Validity, and Reliability.* The decisions made in higher education are only as good as the data on which they are based. A poorly organized data governance system, limited staff, and unclear policies for data entry and data manipulation all lead to less-than-valid information for decision-making. Accurate reporting by institutions to state systems is critical, and might not be attained without checks on the completeness and accuracy of the data submitted. Without a well designed and implemented data audit system, the public cannot have confidence in the quality of the information produced.

Accordingly, states need to ensure that the data elements they request from institutions are clearly and unambiguously defined, as are any rules or interpretations concerning the entry or reporting of these data. Definitions and reporting rules should be developed in consultation with institutions. Similarly, states should provide materials and professional development opportunities for institutional staff to ensure that they understand state rules, regulations, definitions, and protocols. Deliberate, collaborative, and on-going interaction must occur between the state and the institutions. Finally, states need to establish regular data checking or error-identification routines to audit the validity of submitted data.

15. *Alignment with Broader State Goals, Demonstrated Usability and Sustainability.* State postsecondary systems should not exist in isolation; they must be aligned with a state's long-term development plans and goals for its citizenry. Only by becoming an integral component of a state's overall plan, will a system be sustained over the long term. Financial and political support, regardless of current politics, is necessary to ensure that these systems become part of the basic state infrastructure. This support is also critical to address cultural and other challenges that may arise from competing state interests. Furthermore, alignment should be constantly reviewed and updated as priorities change.

A system that collects but never uses data to meet state needs is useless; the only data of value are those that are used. Data in use are transparent and continue to be improved. A mature state system uses data and associated analytics to create information and generate shared knowledge. Sustainability is developed as generated information is used to address state needs.

Achieving an ideal state postsecondary data system with all 15 of these characteristics is a challenging, but attainable, goal. This description of the characteristics is designed to set a high standard without dictating how this standard is to be met. Each state must make its own decisions to develop the ideal system. But by working toward a common set of contents, characteristics, and functions, a level of commonality can be achieved that allows consistent analysis of education progress within states and across state lines.