Data Sources

This section describes the types of data used in the *Tracking Postsecondary Outcomes for High Schools* pilot and offers one model for states that wish to develop linked K–12 and postsecondary education data to help improve college readiness and success.

The five major sources of data that were used in the pilot include:

- State education records for grades 8–12,
- State postsecondary education records,
- National Student Clearinghouse (NSC) enrollments and completions data,
- K–12 school and district information from the federal Common Core of Data (CCD), and
- Postsecondary institutional information from the federal Integrated Postsecondary Education Data System (IPEDS).

Additional detail on both the potential and constraints of these data sources is provided below.

*State education records for grades 8–12.* A wide range of K–12 data were used in the pilot, including student enrollment records showing the district and school of attendance, number of days in attendance, and sometimes a withdrawal code to indicate why a student left a school (for example, due to transfer to another school in the same district, out of the district but within the same state, graduation, or dropout); assessment data, including the major state tests used for accountability under No Child Left Behind, along with other tests such as end-of-course tests in middle and high school; and high school course records. Since the pilot covered a 10-year period, the records for each year sometimes reflected different content, even though the structure of the records remained the same (for example, a test may have been replaced, codes for absences may have expanded or contracted, and courses may have been added or subtracted from the state course catalog). Significant effort was required to reconcile and standardize data across years when constructing student-level longitudinal records from annual files.

*State postsecondary education records.* Enrollment records for public postsecondary institutions were obtained from state sources. In some states, there was one postsecondary data repository that included institutions at all levels (4-year public, 2-year public, and less-than-2-year public technical institutes), while in other states requests for data had to be made separately for each sector. In addition to postsecondary enrollment, the pilot obtained information on student majors, credit attainment, and credentials awarded.
National Student Clearinghouse (NSC) enrollments and completions data. NSC data on enrollment and degree attainment were used to supplement information from state postsecondary sources to describe experiences at out-of-state and private in-state postsecondary institutions. Although the NSC database includes a large proportion of postsecondary enrollments nationwide, there is variation in coverage by type of institution and some postsecondary institutions do not participate in the Clearinghouse. During the course of the pilot we identified several postsecondary institutions that were not NSC participants at the time, but represented important destinations for students in local districts in participating states. It can be important, therefore, to compare NSC’s list of participating institutions to those known to be important postsecondary destinations for students at both the state and district levels in target locales in order to ensure that estimates of enrollment and completion are not significantly underestimated.

K–12 school and district information from the CCD. The CCD is an annual data collection by the National Center for Education Statistics (NCES) of the U.S. Department of Education. CCD collects fiscal and non-fiscal data on all public schools, public school districts, and state education agencies. CCD provides school-, district-, and state-level estimates of student demographic characteristics (such as percent minority and percent eligible for free and reduced price lunch, a measure of family poverty) as well as school enrollments and “urbanicity,” the latter indicating whether a school is located in a large metropolitan area, a suburb or a rural area. CCD data may lag one or more years behind state data sources.

Postsecondary institutional information from IPEDS. IPEDS is an annual data collection of every postsecondary institution in the country that participates in federal student aid programs. IPEDS is administered by NCES. IPEDS surveys collect information on institutional characteristics such as location, level (4-year, 2-year, less than 2-year), control (public; private, not-for-profit; private, for-profit), price of attendance, enrollment, persistence, degrees awarded, and financial aid, as well as information on faculty, credit hours and other aspects of postsecondary education. IPEDS classifications of postsecondary institutions may not match a state’s own classification of its institutions, and some adjustments were made in the pilot to reflect state practice.

Information is provided on key measures constructed from these data sources.