



March 20 [Reality Check](#)

On Large Scale Student Databases

By [John V. Lombardi](#)

[Much comment has swirled](#) around a proposal to create a large federal database that can [track all college students](#) from the moment they enter any higher education institution anywhere in America.

Such a database for student tracking has many possible uses: improving the ability of institutions to support the needs of students who attend more than one institution on the way to a degree, improving the federal and state oversight of higher education effectiveness and efficiency, standardizing curricular offerings nationwide, improving the transfer of credit by ensuring that all courses have equivalent content for students wherever the students may attend college, and permitting significant, wide ranging research on college attendance patterns and the effectiveness of institutions.

A recent report sponsored by the Lumina Foundation entitled [“Critical Connections: Linking State’s Unit Record Systems to Track Student Progress”](#) (January 2007) by Peter Ewell and Marianne Boeke of the National Center for Higher Education Management Systems provides a very useful survey of the current state systems for linking student records and a helpful commentary on the challenges and opportunities of a national version of such a system.

In academic settings, the creation of large databases can become a substitute for recognizing fundamental issues. While good information is always welcome, the fundamental problems in American higher education are probably not the result of inadequate information about the migration patterns of students or the peculiarities in the college transfer processes, although both are important issues. Most challenges in higher education come from a mismatch between the expectations of students, parents and employers and the investment in the institutions designed to meet those expectations.

While very wealthy private and elite public institutions can often mobilize sufficient resources to meet most of the expectations of their many constituencies (evidenced by the high demand for admission and significant success in placing graduates), most higher education institutions struggle with inadequate funding to serve populations of students that may include significant numbers of economically challenged or academically underprepared participants. The continuing decline in public funding and the mirror image increase in tuition and fees cannot be solved by better data.

National level databases have two major difficulties: expense and accuracy. The money part is easy to understand, especially for those who have participated in the last decade’s explosion of cost associated with complex computerized accounting and personnel systems on university campuses. The money will need to be spent by two categories of entities: individual higher education institutions and the federal government. The federal government will need to create a system like the [Integrated Postsecondary Education Data System](#) that can handle the massive amounts of data anticipated from the student tracking system. This is no small enterprise, and we have no real estimates of the cost. Some cynics

might wonder if the initial start-up and the continuing operating costs might not produce more benefits invested in financial aid of some kind.

The higher education institutions, for their part, will need to invest in personnel and systems to capture the data required by the federal system, verify it, and send it forward. Many states have similar systems now for tracking students through their state higher education institutions, but it is almost certain that these will have to be revised to conform to federal standards or else institutions will need to maintain two parallel systems with different standards. The National Center for Higher Education Management Systems maintains [a useful site with information](#) on all the various state systems.

Adding to this complexity, most federal data systems associated with higher education that do not involve money rely on the good will and expertise of the institutions to verify the accuracy of the data. Experience with IPEDS, and especially experience with various ranking systems that rely on institutional submissions, demonstrates the difficulty of acquiring reliable, valid and consistent data across the universe of American higher educational institutions.

Many forms of error enter these systems. Sometimes institutions interpret data definitions differently (we're still discussing how to define instructional employees as "faculty" for the various national data sets). Sometimes institutions choose different ways of reporting what appears to be similar data (on what day in the first semester do we count the freshman class and which individuals do we include when we count the average SAT of entering students). In other circumstances we report numbers that may or may not reflect what their name implies (we report instructional expenditures using an accounting definition that may not accurately separate out research from service from instructional costs).

If, as anticipated, a large database of linked student records is to be used for accountability and comparative purposes, there will be many incentives to fudge the data. Unless the federal government institutes some variety of audit function to ensure that the data are accurate and comparable among institutions, the utility of the information will be low but the temptation to use it will be high.

None of this is to say that such a system might not have considerable value for some purposes. The issue is more complicated. We like binary contrasts: bad and good. But this is not a binary issue. Accurate data are good, inaccurate or incomplete data are bad, expensive data may be good but not as good an investment as something else that contributes to student success or academic quality.

When particular categories of data are used for accountability purposes, institutions will change what they do, because institutional behavior tends to match whatever is measured. If we measure SAT scores, institutions work to increase the average SAT scores; if we measure graduation rates, institutions will do what it takes to graduate students; if we measure sports success, everyone wants a successful sports program. For this reason the quality, characteristics and type of data collected and used in any student unit record system on a national basis assume fundamental significance.

In this conversation, the reality check is to avoid minimizing the difficulties and expense of creating federal artifacts whose relative utility, compared to the real problems affecting American higher education, may be quite low.

We have many challenges in American higher education, but we should not assume that large scale data collection offers a reasonable substitute for actual investment.

John V. Lombardi, chancellor and a professor of history at the University of Massachusetts Amherst, writes *Reality Check* occasionally.

*The original story and user comments can be viewed online at
<http://insidehighered.com/views/2007/03/20/lombardi>.*

© Copyright 2006 *Inside Higher Ed*