

ENVISIONING THE NATIONAL POSTSECONDARY DATA INFRASTRUCTURE IN THE 21ST CENTURY

EXECUTIVE SUMMARY

Institutional Research Capacity: Foundations of Federal Data Quality

RANDY L. SWING

ASSOCIATION FOR INSTITUTIONAL RESEARCH

MAY 2016

Executive Summary

Introduction

The need for data-informed decisions is not limited to national policy, state systems, or senior leadership of postsecondary institutions. Decisions that impact the achievement of higher education missions are also made by students, faculty, front-line staff, and program administrators—all of who deserve data and information to support their decisions. Foundational to effective decision support is the quality of data inputs and analytics provided by each college's or university's institutional research (IR) function.

There is wide agreement that variations in higher education organizations complicate efforts to collect uniform data on institutions and the students they serve. Yet federal and state policies must be informed by data that accurately describe the totality of U.S. higher education arrangements. This paper series provides highlights and details of how improvements to collection of national and state postsecondary data could be undertaken. This specific paper focuses on institution-level data capacity to prepare and report data as the foundation of existing and proposed data collections.

Nearly all colleges and universities that are accredited and participate in Title IV programs have established an IR capacity that supports mandated reporting on enrollments, resources, and student outcomes. Yet the variation in those investments creates vast differences in the capacity of IR to produce mandated reporting and to support institution-level decision support. As efforts are undertaken to improve state and national data systems, attention must be given to improving institution-level data capacities to ensure the quality of the data that enter the data ecosystem. This paper calls for the development of institution-level data strategies that are foundational to all levels of the ecosystem, including students, institutions, states, and federal agencies.

Role in the National Postsecondary Data Ecosystem

The foundation of state and federal higher education data is institution-level data, most of which are derived from IR and data functions at each postsecondary institution. Institution-level data managers and analysts are best positioned to clean and properly array data for submission to state and federal agencies. Because of the variances in postsecondary administrative arrangements and data systems, it is common

for local data expertise to map or crosswalk institutional data with external data requests. The resulting submissions are trustworthy, but come at the cost of institutional burden in human and fiscal resources needed to produce these reports.

Federal agencies are already required to monitor the burden of their regulations, but the focus on burden only as consumed time and resources is misleading. Adjustments should be included to account for the value of the data to the reporting organization. For example, many institutions make such extensive use of Integrated Postsecondary Data System data that, if IPEDS ended, they would be willing to pay a third-party source for access to similar data on their peer and competing institutions. Ultimately, burden can be managed by reducing the resources used in the production of mandated reports or by increasing the value of the collected data for the reporting institutions.

Data accuracy and quality are also functions of use and perceived value by reporting institutions. Data that can be disaggregated to align with decisions at the department, major, or program level have greater use and value than institution-level aggregated results. Although it is somewhat counterintuitive, more detailed reporting can actually result in higher-quality data with a lower burden because the data have multiple uses at the institution level and yet can be easily rolled up to create institution, state, and national data as well.

This paper acknowledges the state and federal interests in the college/university IR function—its resources and leadership—because IR is a core part of the ecosystem of postsecondary data. As such, there are roles at the institution, state, and national levels in establishing individual and overall data strategies to align and coordinate existing and future data collections.

Major Issues

Colleges and universities collect a lot of data, but converting those data into information remains a challenge for nearly all institutions. Doing so at the pace needed for decision support eludes most institutions. Real-time tactical, operational, and strategic decisions cannot wait for new data collections, nor can they be supported by elaborate research designs that may take years to produce. Yet in reality, changing processes that are intended to impact graduation rates or post-college outcomes simply cannot be tested through computer mod-

This publication is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

eling; we have to wait for actual outcomes to accrue, which can take five or more years. As such, decision makers report that data to inform decisions often fail to be current enough or specific enough to identify best choices.

In addition to general capacity shortfalls, numerous newly mandated data collections, such as campus crime data and gainful employment reporting have been added to the workload of IR offices. A common refrain is that after mandatory reporting is complete, there is little time and few resources for research on issues that are important to a specific campus. These concerns unfold in federal technical review panels and in campus-level committees where additions to existing data collections are explored and planned. Even when the higher education community predicts important topics arising, the backlog of issues already awaiting inclusion in new data collections and analyses makes it difficult to be forward-thinking about new elements.

The current trend lines for IR show a field that is growing at a slow pace while existing in a climate where desire for data to inform decisions is rapidly increasing. Simply put, there is too little capacity for IR in the current models of higher education and current structures of IR. All stakeholders, including state and federal policymakers, are negatively impacted by the lack of IR capacity.

Technical Enhancements Needed to Improve IR and Data Functions

Many of the enhancements to IR capacity are not highly dependent on new or expanded technologies. Still, technology can provide opportunities to increase efficiencies and allow maximum use of the existing investment in postsecondary education. More efficient use of existing technologies depends on advancing the technical knowledge and skills of the faculty, staff, and administrators who work at institutions as producers or consumers of postsecondary data.

Resources Needed to Improve IR and Data Functions

Even small increases in human resource capacities quickly add to a massive increase in costs when combined across thousands of colleges and universities. It is unlikely that postsecondary institutions will suddenly add numerous personnel lines to existing IR offices. Rather, capacity can be built by (1) establishing a national data strategy based on a view of a single data ecosystem, and (2) establishing leadership for data capacity at all levels of the ecosystem. While statistical agencies have an important role, they may need assistance in understanding and meeting the decision support needs of students, institutions, systems, states, and federal decision makers.

Improvements at the institutional level may require new resources, especially in establishing chief institutional research officers (CIROs), but much of the additional capacity

can be found by reassigning existing resources and operating an intentionally orchestrated data strategy. That data strategy will be best if it considers the full data ecosystem, from students as decision makers to federal policymakers.

A paradigm must first be established that data literacy for decision support is everyone's role, and then institutional commitment to professional development of staff must follow. Like other disruptive innovations in higher education, an investment by institutions in workforce skills is needed to ensure effective data literacy across all employees.

Policy Recommendations for Improving IR

The following concrete recommendations provide a roadmap for building IR capacity:

- ▶ Establish an intentional data strategy for the overall postsecondary data ecosystem and for each of the components of the ecosystem. Institutions, state systems, and state agencies should undertake this work immediately, and it should be defined and supported at the national level in the next higher education reauthorization act.
- ▶ In planning data collections, build in disaggregation capacities so that data can be useful in decisions at tactical, operational, and strategic levels. Data that inform policy decisions should also be useful in planning, implementing, and evaluating solutions that follow policy development.
- ▶ Each institution should establish a data champion at a cabinet-level position. This CIRO will have responsibility and authority to realize the data strategy for all decision makers in the institution as a decentralized IR function expands the capacities of existing IR offices.
- ▶ Each institution should develop an intentional plan for staff professional development of data literacy skills aligned with position descriptions and personnel evaluation processes.
- ▶ All federal statistical agency missions should include authority to train data providers and data consumers in their respective roles in the data ecosystem.
- ▶ Federal calculations of reporting burden should use a cost-benefit approach that acknowledges the value of data used by the reporting sources in addition to the value to the federal government.
- ▶ Automating data distribution by use of application program interfaces should be funded and required for data collections. In designing data collections, equal consideration should be given to the distribution and use of the data in addition to planning the collection.
- ▶ Institutions should rethink and remodel their data strategies to take advantage of disruptive innovations already in play and update their strategies as new technologies become available.

Envisioning the National Postsecondary Data Infrastructure in the 21st Century is a project of the Institute for Higher Education Policy and is supported by the Bill & Melinda Gates Foundation.

