Institutional Research Capacity:

Foundations of Federal Data Quality

RANDY L. SWING
ASSOCIATION FOR INSTITUTIONAL RESEARCH

MAY 2016

AIR Association for Institutional Research
Data and Decisions for Higher Education

POSTSEC DATA

Randy L. Swing, Ph.D., served as executive director of the Association for Institutional Research from December 2007 through January 2016.

This paper is part of the larger series *Envisioning the National Postsecondary Data Infrastructure in the 21st Century.* In August 2015, the Institute for Higher Education Policy (IHEP) first convened a working group of national postsecondary data experts to discuss ways to move forward a set of emerging options for improving the quality of the data infrastructure in order to inform state and federal policy conversations. The resulting paper series presents targeted recommendations, with explicit attention to related technical, resource, and policy considerations. This paper is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the author(s) and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or the Institute for Higher Education Policy.

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 modeling; 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.

Institutional Research Capacity:
Foundations of Federal Data Quality

The field of institutional research (IR) was formalized more than 50 years ago[1] with the establishment of a professional organization, the Association for Institutional Research (AIR). Over the ensuing five decades, the techniques and methods for converting data to information in support of operational, strategic, and policy decisions have created a high demand for IR capacity at the institutional level. Demand has also increased due to the emphasis on data by state legislatures, the U.S. Congress, and governmental agencies to inform state and national policies and regulations. Common across these demand drivers is that the most useful and frequently needed data come directly from institutions themselves. Any options for improving national data on postsecondary education must consider the data infrastructures, resources, and capacities of each of the colleges and universities that are the primary source of these data and acknowledge that data capacity includes contributions from multiple administrative units within an institution, including offices of information technology (IT), business affairs, student records, IR, and others.

Burden and costs to the institution must be considered in efforts to improve national and state data structures. Although data management and analytics require the consumption of fiscal and human resources that could otherwise be used in direct support of the institution's mission, support for these data initiatives and their underlying infrastructures are not always a burden. They are investments when they are useful in informing policy development, organizational management decisions, or consumer decisions.

While nearly every college and university has IR capacity, both formal and informal reports point to the field falling short of the quality, quantity, and timeliness of data-informed decision support needed by institutional, state, and national leaders. At the same time, disruptive innovations have arisen, including new technologies, commercial collection of personal data, and new models of postsecondary delivery. It is readily apparent that the disruptive innovations in IR have already created and expanded demand for data from a host of new consumers. Data are desired by students, faculty, administrators, staff, cabinet-level decision makers at institutions, leaders of state agencies, and federal policymakers.

Bower and Christensen's[2] model of disruptive innovation foretells that the outcome will be a new business model for the disrupted field, even as the old model continues to exist. How will new models impact IR at colleges and universities, and how might state and federal data collections and distributions also be affected?

Lessons From Prior Disruptive Innovations

IR is not the first field of higher education to experience disruptive innovation.[3] Before the spread of personal computers and desktop publishing software, most colleges operated print shops with managers who served as gateways for publications of all kinds. It was not unusual for a newsletter or simple printing job to take several weeks to complete in the capable, if highly controlling, hands of the print shop manager. The process produced consistent and high-quality publications. However, the printing field changed quickly when desktop publishing turned personal computers into personal printing presses. In the hands of unskilled "designers," a lot of substandard newsletters were produced. After attempts to enforce printing standards failed, savvy print shop managers converted to coaching the new army of newsletter producers, and understood that some decline in professionalism was overcome by the quantity of communications that institutions were able to create. A grassroots, diffused model of printing supported by coaching from design experts has become the dominant model for campus printing offices—even while allowing for the occasional substandard newsletter from a novice user.[4]

Affordable personal computing technologies similarly disrupted the mainframe computing center, resulting in a systems approach to IT. The disparate solutions purchased by departments and offices quickly created an unmanageable array of unique technologies. Decentralization and grassroots decisions were not the right starting point for establishing a functional networked computing capacity. The addition of senior-level leadership brought order to network infrastructures, professional development for faculty and staff, and a shared vision for a networked computing system. Because this was a new concept for most employees, coordination and planning assistance was needed from a centralized source. Over time, it was possible to reduce top-down control as knowledgeable employees developed the required capacities to make increasingly refined decisions about technologies that best fit their unique needs.

IR shares aspects of the disruption of print shop management and mainframe computing. Some IR activities can successfully develop as unique, stand-alone products, but others require the development of infrastructures and skills that are unlikely without institutional investment of resources and

leadership from a senior administrator. Likewise, there is considerable need for professional development of data and analytics skills for faculty, staff, and administrators, who may not have current skills in using and interpreting data. It is highly likely that conditions will get messy during the early stages of creating a network of IR functions and a decentralized process; however, the result of far more production and greater distribution of data for decision support will far exceed the negative reactions to some data studies that do not represent best practices.

Campus Capacities Are Foundational to Federal Data Quality

The indisputable fact that higher education in the United States is large, complex, highly segmented, and unevenly resourced makes it difficult to acquire quality data needed to inform federal and state policy. Unquestionably, state and federal governments have legal authority to collect information on postsecondary education, and colleges and universities have a long history of compliance with mandated reporting. Still, the value of the collected information is highly dependent on the quality of the data gathered, which varies based on the skills and knowledge of the individuals who produce and submit the information on behalf of each postsecondary institution.

Federal reliance on institutional self-reports date to 1869–70, when "a federal education agency collected data on enrollment, earned degrees conferred and faculty."[5] The resulting collection established the first federal data on postsecondary education. This action is noteworthy because it established principles of trust and dependency between postsecondary institutions and the federal government for exchanging information on the status of higher education in the nation. In the years since the establishment of the Higher Education General Information Survey and its successor, the Integrated Postsecondary Education Data System (IPEDS), Congress and the U.S. Department of Education have continued the path of relying on individual colleges and universities as the main suppliers of higher education–related information for use by federal policymakers. These actions are founded on a core belief that no one is better prepared to supply data on postsecondary education than the administrators who lead colleges and universities.

While such data are certainly not error free, they are widely accepted as trustworthy, especially in the aggregate depiction of the national condition of higher education. This is no small feat given that higher education is so diverse that virtually no data definition fits all institutions. The job of making local sense of federal rules and transforming institutional data to conform to data submission criteria largely falls to IR professionals. The foundation of higher education data includes millions of interpretations made during the collection, cleaning, analysis, and submission of information by IR officers working at, or for, each institution.

The handcrafted nature of these data affords a level of individual attention and quality control. It also highlights the potential for data inconsistencies due to variations in individual judgments by data providers. Several federal agencies recognize the need to train the individuals who report these data; the National Center for Education Statistics (NCES) is a leader in providing this professional development. Yet the demand is always greater than governmental or institutional budgets can fully accommodate. Training will continue to be an ongoing need given the fluid environment of congressional mandates and agency regulations, as well as the natural transitions of staff at each institution. Certainly laws, regulations, guidance, and the threat of federal fines for noncompliance are part of the quality control structures undergirding federal data, but the human resource capacity that is needed to ensure accuracy should not be overlooked.

Motivation to produce quality data for submission to federal agencies is influenced by actual and perceived challenges and opportunities, including the following:

Demand that exceeds capacity. Decision makers value data that inform their decisions.[6] Access to IR was once primarily organized to support college and university senior leaders, primarily the president/chief executive officer and the provost/chief academic officer (CAO). Today, mid- and lower-level college administrators are seeking similar access to data to assist in making decisions at the department, unit, and program levels. Most IR offices report being significantly underresourced to meet the demand for services from a broader group of customers.

Burden of mandated compliance reporting. The number of annual mandatory reports constitutes a significant portion of the workload of IR offices.[7] Additionally, the fixed deadlines associated with mandatory reports limit flexibility for scheduling and prioritizing other important work.

Burden of changing reporting requirements. Many IR offices use written procedures, statistical syntax files, or homegrown computer programs to assist with data reporting. Adding new data variables requires time-consuming changes to existing processes and software. The same is true for deleting data elements. In essence, any change requires reworking existing procedures and processes. Much of the data reporting effort is in interpreting definitions and applying them to the data as collected by an individual institution. As such, changing definitions, even simplifying them, requires investment in reworking the institutional procedures and processes.

Reporting that requires linkages to external data sources. Information about students after they drop out, stop out, transfer out, graduate, or enter the workforce frequently requires linkages to external data sources that often carry participation fees and additional staff time for matching multiple data sources. Even when linking between government data sources, there can be significant time investments in establishing interagency agreements for sharing data and programming to align data across different systems.

Reporting that mixes data from different internal management systems. College and university data systems that support specific administrative tasks (e.g., human resources, payroll, student financial aid, student records, and finance) often lack easy ways to merge data with other campus data systems. Combining data from different systems can complicate reporting by requiring time-consuming manual data merges from disparate systems and databases.

Longitudinal/cohort tracking. Upgrades to computer systems and changes in system vendors are frequent occurrences as higher education abandons homegrown systems in favor of more robust commercial software. These changes have expanded data capacities at the institutional level and opened up new opportunities to use data to inform decisions. The increased opportunities come at a cost, however. It is not uncommon for student records to cover 10 or more years, and to require querying of multiple record systems with different variable names, which requires significant staff time.

In spite of the challenges in federal data reporting, most institutions successfully submit their mandated data. For example, of the 7,389 Title IV entities expected to report IPEDS data on completions and 12-month enrollments, responses were not submitted by only two entities for each survey.[8] Overall, institutional reporting of data to the federal government is producing information that is used widely by governmental agencies, researchers, and postsecondary institutions.

Although IPEDS is widely used by institutions because it is readily accessible, it is rarely nuanced enough to be useful in supporting decisions within a postsecondary institution. Institutional averages provide little information about the variance within an institution. For federal policymaking, IPEDS has provided the needed broad look at higher education nationwide, but institutions, and only recently federal and state policymakers, are making decisions about specific groups of students (e.g., low income, first generation, federal loan recipients) and majors (e.g., STEM, teacher education), which can be significantly different than the average across the whole institution. As Jamey Rorison and Mamie Voight note in "Putting the 'Integrated' Back Into IPEDS: Improving the Integrated Postsecondary Education Data System to Meet Contemporary Data Needs," improvements can be made to IPEDS data to enable them to become a more important foundation to institution-level decision making.

It is noteworthy that regional accreditors have not built significant portions of their quality assurance measures on federal or state data even as their expectations for institutions to have and use data continue to increase. Institutional quality, as defined by regional accreditors, appears to require different data elements and a greater ability to disaggregate findings than is currently collected by federal and state agencies. Accreditation teams routinely review institutional data and IR capacities using the professional judgment of the review team, but there are no standards for evaluating these capacities or defining best practices.

Institutions are the foundation of higher education data in the United States, and have been since the federal government began national data collections. The results are that institutions have built capacity to produce trustworthy data and they successfully comply with mandated federal reporting, even as mandates from state agencies are increasing. Institutions supply and also use these data; however, they often have to undertake separate data and analytic reporting processes to make the results useful for decision support at the institution level. It is not uncommon for institutions to produce multiple reports that are similar, but uniquely focused, for particular stakeholders. Additionally, institutions voluntarily participate in an array of reports to ranking organizations, consortia, and other nonmandated programs. These multiple efforts contribute to reporting burden and create confusion when the data and findings vary depending on the methods and definitions used.

Technologies and Resources Needed

It is common to blame the federal government's insatiable appetite for data as the cause for institutional reporting burden. Yet technical review panel recommendations (e.g., IPEDS, NCES sample surveys) reveal that institutional representatives rarely recommend eliminating variables already in the federal collection, and frequently recommend the addition of new elements and increased complexity. These discussions highlight that even with institutional burden considered, the resulting federal data are considered worth the investment by many institutional researchers.

It is unlikely that reducing the amount of information collected would be satisfactory to federal or institutional decision makers. Certainly there are ways to trim collections—for example, it is unclear how the IPEDS Academic Libraries Survey serves federal policymakers—but doing so is unlikely to have a significant impact on overall burden.

There are numerous recommendations for using existing federal data more efficiently through cross-agency exchanges (e.g., National Student Loan Data System, Department of Defense, Internal Revenue Service, Census Bureau, and Department of Education). Other papers in this series address in depth the strengths and limitations of these recommendations. However, it is worth noting that solutions that meet the needs of federal decision makers, but do not include access to the data for use by institutions and state systems, could lead to increased burden and cost as third-party vendors establish separate systems to supply similar data directly to institutions (e.g., tracking student enrollments across institutions).

The U.S. Department of Education's College Scorecard inaugural launch is a promising model for meeting the needs of federal and institutional decision support, along with its primary purpose of providing consumer information. The Scorecard includes data from multiple agencies and an application program interface that provides easy access to link data from the Scorecard to other initiatives (individual privacy precludes the release of some data, such as employment and income records).

Another potential option for reducing reporting burden is federal collection of student unit records that could be analyzed nationally. Higher education data experts see merit in this idea,[9] and have confidence that protecting the privacy of students can be accomplished, although there is not yet agreement about which of several arrangements would be the most workable and cost effective. As with other potential solutions, this is a viable alternative only if it provides data back to institutions for use in benchmarking, peer comparisons, and outreach and support to students, as recommended by Ben Miller in his paper, "Building a Student-Level Data System."

Reducing the burden of federal reporting is the wrong focus if doing so reduces access to data that postsecondary institutions regularly use. When burden is measured only by the time required to produce reports, the offset of the value that accrues to the institution is lost. A true cost–benefit analysis would include both the cost of production and the value of the data created. Simply put, if federal data were suddenly unavailable, colleges and universities would be purchasing comparative data from commercial sources rather than ending the practice of peer benchmarking.

Federal solutions for addressing burden should be pursued, but perhaps the fastest and most practical ways to improve federal reporting and control burden are improvements in IR capacity by institutions themselves. Burden is not equally distributed across all higher education institutions.[10] Large, well-resourced institutions report lower burden for reporting because they automate processes and use resources to improve reporting efficiencies (e.g., submitting IPEDS data). Burden is highest for institutions that have limited IR capacity. A 2016 national report of staffing and resources of IR offices[11] documented the small staff, limited resources, and large sets of tasks assigned to these offices. Most IR offices operate with three or fewer staff members as shown in **Table 1**.[12]

TABLE 1: FULL TIME EQUIVALENT (FTE) STAFF IN IR OFFICES

Director and Professional IR Staff	Two-Year Institutions	Four-Year Institutions
Fewer than 1 FTE staff	1%	1%
1 FTE to fewer than 2 FTE	17%	18%
2 FTE to fewer than 3 FTE	41%	35%
3 FTE to fewer than 5 FTE	28%	26%
5 FTE to fewer than 10 FTE	12%	17%
10 FTE or more	1%	3%

Call to Action: A Vision for the Future of IR

Responding to disruptive innovations in data management and analytics, AIR's *Statement of Aspirational Practice for Institutional Research*[13] presents a vision for the future in which a broader range of decision makers are supported through the development of an institution-wide, networked IR function. Building a network of data producers and consumers across campus (grassroots efforts) to ensure that IR is a broad-based function breaks from the model of isolating data expertise in a single administrative office for IR. Increasing the number of individuals at the institution who are invested in using data for decision support expands the drivers for data quality. Individuals who have a stake in using data are the best advocates for institutional data quality and directly link decision makers with the production and use of data.

Currently, most institutions have a centralized, dedicated administrative unit that specializes in data management and analysis. Such offices frequently report to the institution's president/CEO or CAO. Most of the unit's capacity is consumed by requests from the CEO and CAO and by mandatory reporting. Other administrators line up for access to any left over time and capacity, but it is common for them to have to make decisions without data support, or with limited data support. **Figure 1** shows this model of IR as a service provider.[14]

The *Statement of Aspirational Practice for Institutional Research* recommends that data capacity become a broad-based, networked institutional resource. In such an arrangement, data skills are not isolated in one central office, but rather are distributed to form a federated network of data managers and consumers (see **Figure 2**).[15] This arrangement takes advantage of existing data skills of faculty and staff—many of whom have graduate-level training in statistics and research methods. Most important, the expanded capacity

means that unit-level managers, as well as senior policymakers, have access to the data needed to support decisions.

Understanding that data are valuable resources calls for the establishment of a chief institutional resource officer (CIRO) to monitor and support institution-wide data practices.[16] This chief data and analytics position is not to be confused with the senior IT officer, who has primary responsibility for the technologies that support business information systems. Unique specialized skills are needed for campus-wide management of IT systems, and the same is true for the management and governance of data that result from those systems.

A networked IR function is best achieved by integrating data skills and data literacy into human resource functions, including hiring, evaluating, and career advancement across the continuum of individuals employed by colleges and universities. As was noted earlier in this paper, word processing for newsletters and use of personal computers became common among postsecondary education employees when professional development and hiring practices established these skills as priorities. Improving federal data and the use of data in institutional management depends on investments in the data skills of "occasional data producers and consumers"[17] as well as specialists in the IR office.

The CIRO provides strategic leadership on the use of data as a valuable campus resource—ensuring access to data, data tools, data storage, and technologies that specially support decision makers' capacity for turning data into useful information. Because decisions are often time sensitive, it is essential that data and analytic tools support response times that align with decision deadlines.

FIGURE 1: INSTITUTIONAL RESEARCH AS SERVICE PROVIDER (RANKED SET OF CLIENTS)

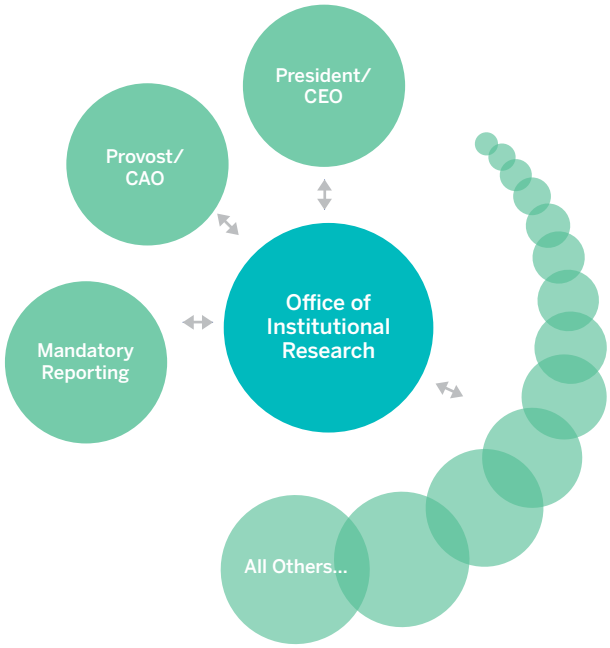
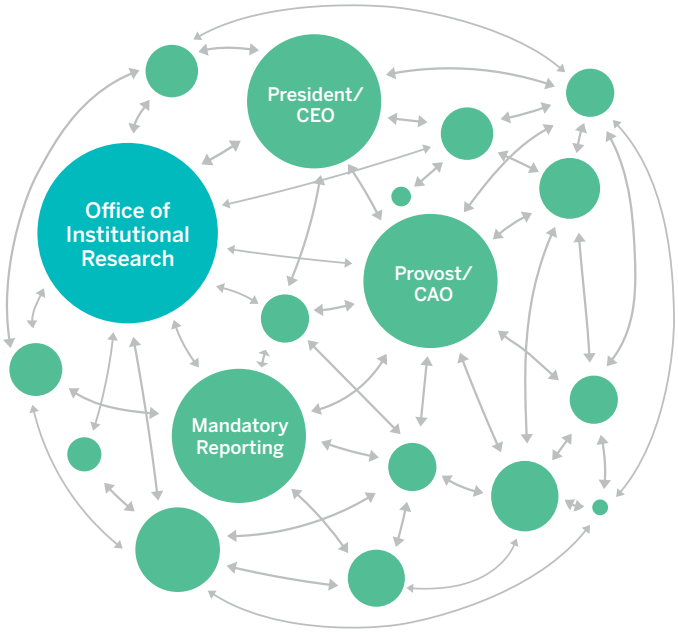


FIGURE 2: INSTITUTIONAL RESEARCH AS FEDERATED NETWORK



Under the leadership of the CIRO, the use of data to inform decisions is widely distributed across campus functions. For example, an academic advisor has data that prioritize advisees by level of risk of poor performance. A faculty member has information that helps with course planning and other information useful in committee or faculty governance work. Department chairs have data on predicted course enrollment demands and other metrics needed to efficiently manage a department or unit. And senior leaders have information to create, monitor, and modify strategic institutional goals.

Networks require attention and maintenance, and such is also true for a networked IR function. The CIRO convenes key data producers and consumers to provide space for conversations about future needs and to encourage a shared language about data initiatives. By mapping data availability, encouraging cross-unit sharing, and avoiding duplication of efforts, networking can be a significant tool for controlling the burden of analytics and reporting. These important functions differentiate the CIRO position from the director of IR position. While the CIRO needs to have a broad understanding of statistics, analytics, data management, and the context of these tools in higher education management, the CIRO must operate from a strategic view of how to accommodate the decision support needs of all faculty, staff, students, and

administrators. Additionally, that view must be clearly communicated and operationalized with understanding of change management processes. The CIRO will have access to technical experts and need not be an expert in all aspects of creating a networked decision support capacity, but will need to be a seasoned senior leader with a broad understanding of institutional context and mission.

The achievement of this new vision for IR will challenge some beliefs that are deeply embedded in traditional organizational structures. Common wisdom holds that a core mission of an office of IR is to serve as "the one source of the truth" when data are communicated. There certainly are times when an institution would be poorly served by communicating two or more different answers to the same (or similar) question, especially in cases where definitions of specific data elements are mandated. More often, however, there are different answers depending on the analytic lens and timing applied to the data. Ironically, one of higher education's greatest strengths is the capacity to recognize and honor varied viewpoints and to use critical reasoning to identify the commonalities and differences in them. Yet when data are involved, conflicting information is unwelcome. It is difficult to believe that academics would allow any one discipline to be the single source of the truth, but with data, there is a drive to have one, and only one, correct answer. The challenge for the CIRO is to encourage active dialogue about different interpretations of data as useful inquiry rather than as territorial squabbles. The CIRO will also have ultimate responsibility for the overall quality of data and analytics by providing training, data tools, and communications technologies. Such conditions are far more likely to result in improved data quality when an active grassroots approach is in place.

The CIRO must communicate the value of using data effectively. Like the print shop manager's dilemma presented earlier, the CIRO must balance enforcing rules against gains from a diffuse data ecosystem. There will remain an array of instances for which consistency and accuracy in using externally mandated data definitions are essential. Training and quality control will be most successful when a senior-level administrator has such as a specific responsibility.

In addition to growing institution-wide talent and providing senior management oversight, changes are required to the business model for establishing and resourcing a networked IR function. The number of new analytic providers entering the marketplace shows that senior decision makers are recognizing that not every institution can afford all the data and analytics talent they wish to have. Some state system offices have begun providing core IR functions, such as reporting to IPEDS, on behalf of all institutions in the system as one way to reduce institutional reporting burden that would otherwise be independently repeated by each institution in the system.[18] This burden-reducing strategy is accomplished by system-level sharing of student unit-level records, which allows one analytic process to work for data from different institutions. While system processing of basic reporting requirements holds promise, few system offices have invested in the IR capacity to bring such to scale. There are significant opportunities for the development of system support functions and other shared service models to serve private and for-profit institutions.

Building From a Strong Foundation

Guessing or philosophizing, especially in times of change and limited financial resources, is a risky management practice for all kinds of organizations, including colleges and universities. Certainly modern management practices are predicated on access to analytics that inform policy and practice. Yet the current underpinning of higher education data might best be described as a loosely coupled arrangement, locked in time, and structured for an earlier model of higher education. *The History and Origins of Survey Items for the Integrated Postsecondary Education Data System*[19] provides a historic overview of the data elements in IPEDS. Each component can be traced back to a specific congressional law or an action of a governmental agency. The historical review does not, however, trace the collection to a coherent data strategy or master plan.

The *Statement of Aspirational Practice for Institutional Research* referenced earlier is based on a core idea that institutions need a strategic data plan to align and guide an intentional design for the IR function. (See **Sidebar 1**) Ad hoc data collections, such as IPEDS, may frequently meet the needs of decision makers, but are unlikely to be effective structures if they are incrementally built without intentional design. The same foundational structure is needed for the entirety of the postsecondary data ecosystem. There is opportunity for the federal government to provide leadership in coordinating the development of a national data system designed to serve students, local and state policymakers, institutional leaders, and federal needs in a coordinated manner.

An early attempt at an intentional design for data started in 2009 when NCES launched the Common Education Data Standards initiatives under the Education Sciences Reform Act for the Institute of Education Sciences. CEDS was intended to establish a common language for data systems across early learning through postsecondary and workforce. The core work was to build a consistent data dictionary and system of "cross walks" between existing data sources to provide a coherent data framework. A natural outgrowth of that effort was finding ways to coordinate data that follow individuals through lower grades, high school, and college, and

SIDEBAR 1: STATEMENT OF ASPIRATIONAL PRACTICE FOR INSTITUTIONAL RESEARCH

An Expanded Definition of "Decision Makers"
Senior leaders have been, and will continue to be, priority consumers of data and information provided by the institutional research function. They are not, however, the only decision makers who impact an institution's achievement of its mission. Other decision makers include students shaping their own experiences, faculty shaping their teaching and interactions with students, and staff shaping program designs and direct interactions with students.

Top-down policies and structures alone do not ensure informed choices and commitments to successful pathways. Broadly engaging all stakeholders in data-informed decisions (tactical, operational, and strategic) is essential for institutional excellence. This hybrid model positions students, faculty, staff, and other decision makers as key consumers and clients of institutional research, and is foundational to a change agency vision of institutional research as a driver for institutional improvement.

Structures and Leadership for Institutional Research
The complexity of modern higher education demands investment in leadership and staffing for strategic, tactical, and operational decisions. Use of data for institutional research cannot be restricted to one office. With greater access to data sources and data tools, and increased department-specific data, institutional research products are widely dispersed across higher education institutions already, even when a strong central office of institutional research exists. An increasing number of staff and mid-level administrators are expected to use data to inform decisions, and decision makers at all levels are establishing their own data collection processes and analytics. Where institutional research once took pride in being the "one source of the truth," the reality is that the new role for institutional research is in coaching a wide array of data consumers, managing institution-wide data and analytical requirements, and orchestrating "the economics of institutional research" in balancing information supply and demand.

A Student-Focused Paradigm
In this aspirational vision of institutional research, data and analytics are transparent and are *intentionally* focused on improving the student experience. Many of the past successes in institutional research have focused on students—enrollment management, retention, engagement, and graduation rates. Yet that focus can be further enhanced by intentionally grounding institutional research initiatives and reports in a student-focused perspective. A key question to be addressed in all institutional research is "how does this exploration serve students?" An essential component of communicating these results is making clear their underlying student-centered purposes.

into the workforce. In essence, this work began the process of envisioning a data ecosystem that acknowledged the integration and codependency of data across the entirety of the education sphere and connected it with post-college outcomes.

Clearly the federal investment in postsecondary education and the connection between higher education and the country's future provides reason for federal leadership in coordination of a national data strategy that supports strong postsecondary institutional decision-making as well as information for state and national policymakers. There is need for the next higher education reauthorization act to include guidance and resources for establishing and articulating the national data strategy for postsecondary education.

Collectively, the papers in this series provide much of the raw material and options for establishing a national data strategy. A priority first step is to establish the roles and responsibilities of key stakeholders—students and their families, colleges/universities, states, and federal agencies. This step must address how each will fund their roles and what level of cost aligns with a reasonable return on investment for having these data. Such an undertaking must engage colleges and universities, and would be a fundamental element in the institution-level planning for data and analytic capacities.

Summary and Conclusion

The phrase "garbage in, garbage out" is a well-known warning that trustworthy results require quality inputs. In small datasets it is often possible to "eyeball" the numbers and check the reasonableness of the results. The colloquial saying "give it a smell test" likewise means to use your senses to determine if the results seem right, or in more technical terms, to check face validity. Datasets have grown larger and more complex, as has the case for more complete and comparable national data on postsecondary education, and it becomes increasingly difficult to sense the accuracy of a computation. Useful analyses require quality data as trustworthy inputs, and when the quality is not there, it can be difficult to discern the lack of accuracy in the output.

When data come from organizations as varied as institutions of higher education, the challenges to consistent data quality multiply rapidly. Certainly technologies that test and screen data are useful, but for the near future, human interpretations and direct manipulations of data—checking, cleaning, cross-referencing—will remain important aspects of federal higher education data collections. The foundation of quality higher education policy decisions stems from the knowledge, skills, and good will of those who supply data to the federal government. Top-down decisions about federal data collections are important; a grassroots commitment to quality data is essential.

A data strategy is foundational to improving data capacities at all levels of postsecondary education. This paper provides recommendations for capacity building and leadership for an institution-level data strategy. This strategy must be conceived as part of a larger ecosystem, with aligned and coordinated data strategies for institutions, systems, states, and federal agencies. A successful and efficient higher education network is too important to individuals and to the nation as a whole to be left to incremental, ad hoc data systems. An intentionally designed system of postsecondary data is essential and can be developed using tools and resources already available.

Endnotes

1 Association for Institutional Research. (2011). *The first 50 years*. Tallahassee, FL: Association for Institutional Research.
2 Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: Catching the wave. *Harvard Business Review 73, no.1* (January-February 1995): 43-53.
3 Swing, R. L., & Ross, L. E. (2016, forthcoming). A new vision for institutional research. *Change Magazine*.
4 Swing & Ross (2016, forthcoming).
5 Statistics, National Center for Education. (2009). *Digest of education statistics*. Retrieved from http://nces.ed.gov/programs/digest/d09/
6 Gagliardi, J. S., & Wellman, J. (2014). *Meeting demand for improvement in public system institutional research*. Washington, DC: National Association of System Heads.
7 Swing, R. L., Jones, D., & Ross, L. E. (2016). *The AIR national survey of institutional research offices*. Tallahassee, FL: Association for Institutional Research. Retrieved from http://airweb.org/aspirationalstatement
8 Ginder, S. A., Kelly-Reid, J. E., & Mann, F. B. (2014). *Postsecondary institutions and cost of attendance in 2014–15*; First look. Washington, DC: U.S. Department of Education.
9 Rorison, J., & Voight, M. (2015). *Weighing the options for improving the national postsecondary data infrastructure*. Washington, DC: Institute for Higher Education Policy.
10 National Center for Education Statistics. (2014–15). *IPEDS 2015–16 data collection system institutional burden*. Retrieved from https://surveys.nces.ed.gov/ipeds/ViewContent.aspx?contentId=12
11 Swing, R. L., Jones, D., & Ross, L. E. (2016)
12 Swing, R. L., Jones, D., & Ross, L. E. (2016)
13 Swing, R.L., and Ross, L.E. (2016). *Statement of Aspirational Practice for Institutional Research*. Retrieved from http://www.airweb.org/Resources/ImprovingAndTransformingPostsecondaryEducation/Documents/Statement%20of%20Aspirational%20Practice%20for%20IR.pdf.
14 Swing & Ross (2016, forthcoming)
15 Swing & Ross (2016, forthcoming)
16 Swing & Ross (2016, forthcoming).
17 Gagliardi & Wellman (2014).
18 Fuller, C. (2011). *The history and origins of survey items for the integrated postsecondary* (National Postsecondary Education Cooperative Report No. 2012-833). Washington, DC: U.S. Department of Education.

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.

IHEP
INSTITUTE FOR HIGHER EDUCATION POLICY