

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Mathies, Charles

**Title:** Uses and Misuses of Data

**Year:** 2018

**Version:**

**Copyright:** © Springer International Publishing AG 2018

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Mathies, C. (2018). Uses and Misuses of Data. In K. L. Webber (Ed.), *Building Capacity in Institutional Research and Decision Support in Higher Education* (pp. 95-111). Springer. Knowledge Studies in Higher Education, 4. [https://doi.org/10.1007/978-3-319-71162-1\\_7](https://doi.org/10.1007/978-3-319-71162-1_7)

## Chapter 7: Uses and Misuses of Data

**Charles Mathies**

**University of Jyväskylä**

[charles.f.mathies@jyu.fi](mailto:charles.f.mathies@jyu.fi)

### **Abstract**

Higher education is awash with data with more coming every year. Much of this is due to accountability reasons and growing interest from stakeholders (internal and external). New technologies have increasingly allowed more powerful analytics and deeper dives into data. However the increasing pressures and new technologies have brought significant questions on how data is used and its relevance. This chapter examines how data is used within higher education and an extended discussion of how it has been misused. The chapter concludes with three suggestions on how institutions can do to reduce the likelihood of the misuse of data and the role of institutional research could play in facilitating its appropriate use.

### **Introduction**

“Evidence-based” or “data-driven” decision making are popular current buzz words in higher education and reflect a growing trend to collect, develop, and use data and analytics in assessments, policy-making, and planning. Stakeholders, both internal and external, are increasingly pushing for information and “want to see the data.” Higher education has become more data-driven and will continue to move in this direction in the near future (see Chapter 6 on data visualization). When a new senior executive comes to campus, they arrive with “new” ideas and often push to make a mark. In recent years, the arrival of a new senior executive is usually accompanied by distinct ideas of how data should be used. They and their team will often pour through reports and analyses looking for patterns and outliers. Seeing opportunities, they will make new policies or recommendations intending to improve student or institutional performance. But is this actually a good use of institutional data? Or could the data have been used more effectively, or in some cases, could the misuse of data been avoided?

While the use of data and analytics has allowed more insights into students, staff, and institutional achievements, it has simultaneously raised critical questions on how it is used. The above example is one small but powerful example that illustrates the desire for the collection and use of data as officials are often simply wanting to improve institutional or student success. However, the use of that data can have unintended consequences and there are significant concerns over its misuse. Specifically, there are concerns over the power, legitimacy, ideology, transparency, intentionality, and relevance of data (Calderon 2015). Whether data is used improperly or unethically, past occurrences of misuse have tended to cast a negative shadow over an institution long after an incident (see University of Oregon case from 2014 as example) (Gray 2016, New 2015, Read 2015). Institutional research (IR) has been in the middle of this data evolution as it regularly manages, synthesizes, and transforms data into useable information.

### **What is data?**

Perhaps the best place to start is to examine what is meant when the term “data” is invoked. The Oxford dictionary (Oxford University Press 2017) defines data as “facts and statistics collected together for reference or analysis.” However there are often misconceptions that data and information are one and the same as “data” often takes on

a dual meaning; in a plural form (as a fact/statistic) as well as in singular mass form (outcome from analysis - as information). While they have similarities, there are distinct differences. Paul Beynon-Davies (2011, 2013) uses the concept of a sign to differentiate between data and information; data is a series of symbols while information occurs when the symbols are used to refer to something. In other words data refers to the actual symbols (data in raw or processed form) while information refers to the interpretation or understanding of said symbols (data). However Spillane (2012) argues that simply giving data reports to colleagues is not enough for effective data-driven decision making; it requires translating data into information and actionable knowledge which can be applied to current and future problems.

Much of the core activities of IR, especially in North America, directly involve the use of data; its collection, analysis, reporting, and governance (AIR 2017). This includes the technical aspects of working with data as well as serving in roles of data stewards and educators to the larger institutional community (AIR 2017). Having multiple roles with data is something IR has been historically known for, as illustrated in the surveys of AIR membership and their duties over the years (Knight et al. 1997, Lindquist 1999, Swing et al. 2016, Volkwein 1990, 2011). The increasing complexity of data interactions though, coupled with the continuing evolution of institutional (organizational) structures, has amplified the reliance on and need for the experience and expertise with data that IR (professionals and offices) possess (Possey & Pitter 2012). In short, “data is the lifeblood of IR,” and the effectiveness of an IR office and professionals is often measured by how well its practitioners understand and manage data (Ronco et al. 2012, p.678).

### **How Data is Used in Higher Education**

Regardless of the name used to identify the unit, IR offices or similar units, sit in a unique place within institutions. They are one of the few, if the only unit, that views the institution at the macro (institution-wide), as well as the micro (departments and individuals) level due to their extensive interactions with institutional data. Institutional data can be thought of as a matrix; when there is a change in one area, it impacts other units as it flows through the institutional data system. For example, when a professor receives and accepts an external research grant, this single act impacts multiple offices (data) across the institution. Typically, the research office oversees the grant application and operations, financial affairs manages the finances, and facilities provide and administer the research space (labs), while either academic affairs or the department are tasked to hire replacements (instructors) for the courses to be filled while the professor focuses time on the grant. All of these actions impact institutional data and create continuous updates from the initial submission of the grant through its completion. As such, IR staff have an institution-wide perspective (macro level), but are able to see the inner workings of an institution (micro level) due to their knowledge of and ability to see the data as it moves through the institution.

Coburn and Turner’s (2012) “The Practice of Data Use” is a useful paradigm to understand how data is used. “The Practice of Data Use” seeks to understand what happens when people engage with data in their normal work day (Coburn and Turner 2012, p.102). Spillane (2012) argues that to understand the use of data requires consideration of the situations it is used (context); e.g. how users notice and interpret new information in their daily practice. Understanding how data is used in practice and it’s the context surrounding it helps explain not only the outcomes of its use, but also the interrelationship between macro structures (social or organizational) and micro action (projects, reports, analysis) involving data. Coburn and Turner (2012) contend that there are three distinct types of

studies describing the practice of data use. All three help in conceptualizing what happens when IR professionals engage with data in their everyday roles and how it relates to organizational learning (in the case of IR, the institution learning about itself).

First, data is conceptualized as an interactive endeavor that involves understanding how data flows into streams of ongoing action and interactions as they occur (Coburn & Turner 2012). IR professionals are often integral members of the institutional management; particularly with how data is used on campus (ex: institutional assessments of students and staff). IR is also regularly tasked to provide either direct or indirect oversight of institutional data (in the role of data stewards) and often asked by senior leadership to provide leadership and commentary on data and its use throughout the institution. These regular, or routine, interactions with data create a pattern or recurrent way (i.e., role) for how IR professionals interact with data within an institution.

Second, the role of the environmental, organizational, and larger context influences how data is used (Coburn & Turner 2012). This is in-line with Terenzini's (1993, 2013) tiers two (issues intelligence) and three (contextual intelligence) of his "organizational intelligence," which has long been one of the standards in understanding the skills and competencies of IR professionals. Coburn and Turner (2012) argue that there are shared languages and common frames of reference by those using data which have developed over time and multiple interactions. Within IR offices, this is seen with how IR is often involved in establishing and managing data protocols, definitions, and reporting. Quite simply, the culture of working with data normalizes its use and meaning across an organization; within an institution IR is usually a leading unit in shaping how data is collected, used, and in general, perceived. It should be noted that institutional culture(s) and norms will shape the action with and meaning of data not only by IR, but also the entire institution.

Lastly, data use is viewed as a "situated phenomenon" that examines how data is used in real time and practice (Coburn & Turner 2012, p.103). Within IR, this is done through observations of practice as well as surveys and interviews of professionals in the field. The multiple surveys of AIR membership over the years (Lindquist 1999, Swing et al. 2016, Volkwein 1990, 2011) are clear accounts of professional practices (norms) as well as the tasks of IR professionals and their role with data. These surveys show that historically, IR has had multiple roles with data which have helped guide institutional leaders managing their institutions.

When IR professionals engage with data, their first step is often to determine the general purpose of its use: is it to describe something or an activity, or is it to infer something (McLaughlin et al. 2012)? The outcome desired from analysis will often dictate not only the methodology involved in the analysis, but also key facets (amount, scope, definition, date collected, etc.) of the data used. IR professionals need to use care in their choices of how they use data as there is a need for data to be used in a clear, thoughtful, and objective (impartial) way to allow IR to remain as neutral and apolitical as it can be.

Leaders within AIR, the Association for Institutional Research, identified five categories of IR work in their *Development of the Duties & Functions of Institutional Research* (2017) and directly link each category with data. The first category is to *identify information needs*. This is the iterative process of identifying relevant stakeholders and their decision support needs (AIR 2017). This is a proactive function where IR staffs anticipate the future informational needs of the institution through reviews of data and policies to assist stakeholders in defining or framing what information is needed. The second category is to *collect, analyze, interpret, and report data and information*. This involves an understanding of the data available and how it was collected to the processes of analyzing and reporting it. This category reflects the technical tasks undertaken by IR to provide data,

information, and analyses (AIR 2017). Some specific examples in this category include enrollment management analyses, assessments of student learning, and assessments of faculty teaching and research. The third category is to *plan and evaluate*. These are formative and summative evaluative processes conducted within an institution directly for planning and decision-making purposes (AIR 2017). Often budgetary, strategic planning, and governance tasks for compliance, accreditation and program review fall within this category. The fourth category is to *serve as stewards of data and information*. IR has a direct role of insuring an institute wide strategy on data governance and its analytics (AIR 2017). Tasks and roles in this category involve key performance indicators (KPIs), data warehouse, and business intelligence (BI) tool(s) development to overseeing the integrity of the data used. The fifth category is to *educate information producers, users, and consumers*. This encompasses the training and coaching related to the use of data, analyses, and information to inform decision-making (AIR 2017). This is often in a collaborative role as IR and information technology (IT) have overlapping leadership and support responsibilities in advancing data-driven decision within an institution (Dahlstrom 2016).

Over the last few decades, the volume, velocity, and variety of data handled by IR has greatly increased (Taylor et al. 2013, Visser & Barnes 2016). This in turn created increased expectations of what data can provide (Calderon & Mathies 2013). Two of the larger increases in recent years have been in the form of data metrics and analytics. Both involve combining multiple pieces of data together to provide an understanding of institutional performance (see also Chapter 6 on data visualizations for a more in-depth discussion). While they have many overlapping features and are often used interchangeably, data metrics and analytics though, are not the same (Patterson 2015).

Data metrics are more informational in nature and are standard measurements derived from past organizational operations. Examples include credits generated by faculty FTE, efficiency of resources used, and admission yield rates. Data analytics, though, are more strategic and future focused, which applies statistical techniques and models on past performance data to see patterns in the data or to predict a future outcome. Examples here include forecasting the number of students to admit in an admissions cycle, predict target enrollment numbers (based on historical acceptance rates, student persistence and graduation rates, and institutional changes such increasing faculty within a department), deriving the locations of focused admission recruitment (based on historical acceptance and persistence rates and the amount of previous resources expended in an area), and gaining insight into the student experience (ex: mapping student activities and practices online to when they use institutional facilities). Both data metrics and analytics look to provide data in singular mass form (as information) allowing institutions to take action and improve the student experience and institutional performance.

### **Examples of Data Misuse**

There have been multiple ways data has been misused in higher education (i.e., colleagues on campus, legislators, media) and, unfortunately, sometimes directly by IR professionals themselves. Often these misuses cast a large negative shadow over an institution (or unit) well after the incident is over (see University Oregon example below) (Gray 2016, New 2015, Read 2015). Below are four general categories or groupings of how data has been misused in recent years. While these categories or groupings are fairly comprehensive, it is not exhaustive and do not cover every instance of data misuse. It is important to note that in most cases, there is no distinction between intentional or unintentional misuse of data. Both cause essentially the same problems, though intentional could create additional issues (legal, financial, security) that would need to be addressed.

### ***Data that Provides Inaccurate Information/Misreporting***

Inaccurate reporting of data is becoming a frequent occurrence within higher education. When data (or information from the analysis of data) provides false information, it creates situations where resources (time, human, and financial) are needed to correct the inaccuracies and mitigate the consequences of its misuse. In many ways, inaccurate information or misreporting is similar to a type 1 (false positive) or type 2 (false negative) error in statistics. They make a data consumer believe something is true when in fact it is not.

One of the more recent and notable occurrences of inaccurate reporting of data comes from the U.S. Department of Education's College Scorecard. The College Scorecard is an online tool allowing students (and their families) to compare institutions on a series of data points. In January of 2017, the Department of Education admitted it had been publishing inaccurate information about repayment rates on its College Scorecard website (Mahaffie 2017). The Department of Education disclosed the loan repayment rates for many institutions was inaccurate due to a coding error leading to the undercounting of borrowers who failed to pay down any of their undergraduate student loan balance (Mahaffie 2017). There was roughly a 20 percentage point decline in the national rate of repayment after the adjustment (Fain 2017). Many advocacy groups and politicians (from both sides) have argued for more reliance on loan repayment rates than on loan default rates as metric of accountability (Fain 2017). This is not the only criticism the College Scorecard has faced over its data and calculations (particularly with the *salary after attending* data, as it is only based on students who received federal financial aid), and it has clear challenges to its long-term viability along with the lack of a stewardship plan (Whitehurst & Chingos 2015).

Providing inaccurate information or misreporting data has also been used as a form of propaganda. In this way, the inaccurate information is used to encourage a particular viewpoint to alter the attitude and perception associated with an institution, agenda, or issue. It is not objective and is often displayed by presenting facts selectively. Perhaps the most notorious misreporting of data in this manner is with the admissions data of incoming students. Institutions face a great deal of pressure to improve their academic profile, which in turn raises their prestige (and ranking). Some institutions have been publicly outed for inflating the admissions data of their first-year students in recent years as part of efforts to raise their academic profiles. Specific examples in the Emory University and Claremont McKenna College (two of the more well-known recent examples) cases include overstating the average test scores by excluding specific students (ex: bottom 10 percent) from its calculations, inflating the average class rank of incoming students, and deflating admission (acceptance) rates (de Vise 2012, Jaschik 2012, Strauss 2012).

### ***Flawed Data Governance***

As mentioned in Chapter 1, IR's involvement in institutional-wide data governance is important, as flawed data governance can lead to multiple misuses of institutional data. While Chapter 5 provides a more in-depth overview of data governance, but in brief, it encompasses the people responsible for data quality (data stewards) and the policies and processes associated with the collecting, managing, storing, and reporting of data (Koltay 2016, Young & McConkey 2012). The misuse of data through flawed data governance typically occurs when institutional data is allowed to be used in way it is not intended (e.g., as the next grouping of data misuse listed below details - *data used as it was not intended*) or when it is accessed by outside entities (data breaches).

When data is allowed to be used as it is not intended, it is often through a lack of data protocols, definitions, and structures, or a breakdown in their use (data mismanagement). Knowledge of the data, definitions, and structures are precisely the knowledge that IR professionals possess and thus should be central to the development of policies and procedures around institutional data. The consistency in application and accessibility of protocols, definitions, and use are key factors in limiting the misuse of data (Young & McConkey 2012). Koltay (2016) suggests adopting a clear data governance is advantageous to institutions, as it standardizes data, enables increased transparency of its use, and reduces costs (financial, time, and human resources). The regular monitoring and reviewing of an institution's data governance promotes a culture of improvement, which in turns increases the level of consistency of data use within an institution (Young & McConkey 2012). Having good data governances improves the overall quality of institutional data while providing assurances that institutional data can be trusted.

Data breaches are becoming more common in higher education, just as they are in industry and government. However, higher education institutions likely have a larger number of reported breaches, because of their open and transparent culture, than organizations in industry who typically report data breaches only when they are legally required (Grama 2014). Three high profile cases in 2016 illustrate different ways data breaches are occurring within higher education. First, the University of Central Florida (UCF) informed 63,000 current and former students and staff that hackers had comprised personal information (primarily names and Social Security numbers) (Binette 2016). This was undertaken by a direct intrusion (hack) into the UCF's computer network (Binette 2016). A second example at The University of California, Berkeley occurred when its financial system was breached by hackers and exposed social security numbers and bank accounts of 80,000 current and former students and employees (Sposito 2016). The data breach occurred when a sensitive piece of commercial software was being updated and connected the financial system to the Internet (Sposito 2016). Lastly, over 3,000 current and former employees of Tidewater Community College (TCC) in Virginia had their names, earnings, and social security numbers stolen in a phishing scam (McKinney 2016). The breach occurred when a TCC employee responded to a data request from a fake TCC email account and supplied the requested data (Mckinney 2016). While good data governance cannot eliminate all possibilities of data breaches, it can help institutions protect their data and mitigate security risks.

### ***Data Used As It Was Not Intended***

Data, when it is used as it was not intended, primarily happens when either data is not used as it was defined or when it does not measure what it intends (or claims) to measure. In many ways, it is often a definitional or a misapplication issue. Two of the more well-known examples involve measurements from external organizations to higher education institutions.

A first example is the use of standardized test scores in the admissions process. In the admissions of undergraduate students, the misapplication of standardized test scores has been known and documented for some time (Beatty et al. 1999, National Association for College Admission Counseling 2008, Wightman 2003). Over the years the use of standardized tests in undergraduate admissions been used as a mechanism to evaluate applicants efficiently but has evolved from a tool of inclusion to one of exclusion (Wightman 2003). Wightman argues (2003, p.83) that "despite extensive evidence to the contrary, test scores are being portrayed as an accurate, objective indicator of merit" which results in pitting the concepts of merit and academic standards against the benefits of diversity and educational opportunity. Graduate admissions are not immune to misuse of standardized test

results either as Posselt (2016) argues there is frequent misuse of GRE scores. Specifically, most graduate programs create GRE cutoffs or minimum requirements and use this as a primary strategy to quantify applicants' quality (National Association for College Admission Counseling 2008, Posselt 2016). Educational Testing Services (ETS), the company overseeing the GRE, however does not advise departments or institutions to use cutoffs.

Standardized test scores are intended to be one of many factors in admissions, not the primary or sole criteria. More appropriately, they are designed to be used as a direct evaluative measure of a student's achievement in course work or prior school performance (National Association for College Admission Counseling 2008). In undergraduate admissions, they are often misused because of the amount of applicants is so great that standardized tests are used to handle the volume and sort through the applicants efficiently (Beatty et al. 1999). Posselt (2016) contends that most graduate departments use the cutoffs and minimum GRE scores because they are pressed for time and need to be speed up the graduate admissions process. By creating cutoffs, departments are able to "throw-out" applicants, thus reducing their applicant pool and the time needed to decided which candidates are worthy for admission or not.

The over reliance on standardized test scores in the admission process is a misapplication, and raises questions of bias and the validity of the admissions process itself as there are concerns about the validity and bias of the tests themselves. However empirical research generally does not support that there is bias against test takers who are not white and male nor that the validity of the tests for the limited purpose for which they were designed (Wightman 2003). While the evidence supports the validity of standardized tests as predictive measures of student performance, their utility for admissions should not be simply accepted without question (Wightman 2003). The technical question of whether test scores are statistically related to an outcome of interest (e.g. student success) is not enough to determine how a standardized test should be used in the admission process (Wightman 2003). The standardized tests were designed to measure reasoning, critical thinking, and analytical writing skills of applicants and to be used as supplements to transcripts, recommendation letters, and other qualifications as deemed important by an institution (ETS 2017, Wightman 2003). The last part is perhaps the key point, "qualifications which are deemed important by an institution." Institutions make admissions decisions but when they do so, they need to recognize how standardized tests fit within their institutional culture and admission process and should not use standardized tests to simply increase efficiency. However, it should be noted that this is often easier said than done.

A second example of data used as it was not intended is with university rankings. University rankings are often viewed, primarily by external stakeholders, as a way to gauge institutional quality and benchmark institutions against one another. However, university rankings, both global and national, have received a significant amount of criticism for their methodology and data quality. In general, university rankings are primarily shaped by the availability of data, and many have subjective, as well objective, issues in their metrics (Hazelkorn 2015b). For most rankings the primary measures of quality are student and staff characteristics, bibliometric outcomes, and reputational surveys (Hazelkorn 2015a). Things like the added value from attending a particular institution (teaching and learning outcomes), the impact or benefit of research in society, or student experiences are usually not included in rankings (Hazelkorn 2015b). However these are the things that can perhaps tell us the most about the quality of the teaching and learning within an institution. Rankings metrics also create a very narrow perspective of what is quality in higher education and of the construction or organization of academic disciplines themselves (Marginson 2014, Robinson-Garcia & Calero-Medina 2014). Marginson (2014) argues that this creates situations where institutions are measured through either flawed data and methodologies or have been



subjected to a misapplication of measures, which were meant to express an outcome different than institutional quality (i.e., the criteria evaluating institutional quality are wrong).

But why have rankings increased in prominence in recent years when there is such strong criticism of their data and methodologies? In general, there is strong desire for institutions to obtain the highest ranking possible because it bestows academic prestige. This can, in turn, lead to an increase in student applications or more extramural funding for faculty research. Rankings are not new as they have been around for over 100 years (Hazelkorn, 2015b), but the social-political effectiveness or power associated with them is (Teichler, 2011). The status bestowed onto institutions by rankings in many ways becomes a circular game in which power makes itself (i.e., the higher ranked institutions are often the benchmark for excellence, so when compared to themselves they perform and rank well) (Marginson, 2014). Additionally, rankings popularity is derived from their simplicity (Hazelkorn, 2015b) as they are easy to understand, especially by external stakeholders. This is perhaps the key point as rankings are increasingly used by students in their selection of choice of enrollment (particularly international students) and governments and industry to compare and benchmark institutions against one another (Hazelkorn, 2015a). In the end though, rankings are merely measures of quantifiable pieces of institutions with many indicators being of wealth and not necessarily educational quality (Hazelkorn 2015b).

### ***Violation of privacy***

Institution officials have a duty to protect the confidentiality of student and employee data. When an individual's data is collected, accessed, or used not in-line with institutional policy, this constitutes a violation of privacy. Within the United States, the Family Education Records Privacy Act of 1974 (FERPA) governs the access and dissemination of student records. Often times FERPA, and similar laws in other countries, are used as the legal base to form institutional policies on data protection for community members (students and employees). While FERPA allows the U.S. Department of Education to enact a penalty of removing all or part of federal funding when there is a violation, the likelihood of this occurring is not very high. In most cases, the Department of Education will notify the institution of the violation(s) and then require specific changes to be made to bring the institution into compliance. Only if an institution refuses will the removal of federal funding actually occur. However, in today's digital world, privacy concerns are no longer limited to legal implications of laws like FERPA, as there are many ethical and policy concerns in a networked society which has increasingly new technologies.

Perhaps the most common way that data is misused in this category is when an employee views specific student records for whom they have questionable or no right to do so. An institution's policies governing student privacy often forms the base of what is legal, however employees are still often left with ethical choices of what they should or should not do in relation to accessing students' records. A recent case (2014) at the University of Oregon illustrates this issue well. The university was facing a lawsuit by a female student in relation to its handling of a legal case involving the student and members of the university basketball team (New 2015). The university's lawyers accessed the student's mental health (counseling) records in the process of defending itself (Gray 2016, New 2015). While ultimately considered legal at the time (the case subsequently caused significant changes in how student health records are protected in the state of Oregon), the release was in violation of the health center's policy and became public when two employees became "whistleblowers" as they notified state regulators at the state legal and psychologist Bar associations of the access of the student's health records against policy (Read 2015). The "whistleblowers" notified state regulators primarily over ethical concerns of providing the

student's health records without notification nor consent to members of the university's legal affairs office (Read 2015).

Another regular occurrence of violation of privacy occurs when students opt out of their personal information being able to be disclosed, but it is nevertheless included. The most typical occurrence here is when a student's demographic information is published in a directory even though the student requested that it not be published. Unfortunately, personal information being disclosed without consent is becoming more of a concern, as personal data (student and employee) are increasingly networked within data repositories allowing individuals who may be unaware of the privacy and legal considerations to interact with personal data. The confusing part, particularly for those not up to date on the laws, is there are many legal exceptions that allow personal information to be disclosed (U.S. Department of Education 2015). One of the more commonly used exceptions is for "school officials" to use student data for "legitimate educational interest." However institution leaders are required to define who is considered a "school official" and what constitutes a "legitimate educational interest" while at the same time notifying students of their definitions and processes (U.S. Department of Education 2015). While neither students nor parents can directly sue an institution for a FERPA violation, a complaint can be filed with the U.S. Department of Education, which can spur an official investigation into the conduct of the institution in question. Most institutions want to avoid a complaint being filed and, as such, have a tendency work with students (and families) to remedy the situation as best they can.

### **Parting thoughts and Recommendations to Minimize Misuse**

We are living in "post-truth" society where lies and the questioning of facts has become a common occurrence (Higgins 2016). The increasing misuse of data coupled with the current environment is threatening higher education's community values of scientific truth, openness, and transparency. This has left many institutions (and academics individually) wondering how they fit in and operate, as data and information (i.e. evidence) is no longer valued as much by the community it supposedly serves. Below are three suggestions of what institutions can do to reduce the likelihood of the misuse of data and (hopefully) improve its use.

First, there needs to be an acknowledgement for not only the need for data and technical expertise, but also context expertise. One constraint of effective data-driven decision making in higher education is the fact that many academics lack formal training and/or expertise with educational data (Horta et al. 2017). This creates a reliance on experts, such as IR professionals, to provide needed context and understanding of data; whether in its plural form (as a fact/statistic) or in its singular mass form (as information). The need for context expertise is critical as there are many questions around the legitimacy, intentionality, and even ideology of how data is used within institutions (Calderon 2015). Without context expertise, there is no questioning the motives, source, and relevance of how data is used which impacts the overall stability of an institution (Calderon 2015).

While the data analytic gap within senior management is improving, the increasing sophistication of analysis is outpacing their increased abilities (Ransbotham et al. 2015). The result is it often leaves a need for senior managers to become comfortable applying analytic results they do not fully understand or comprehend how they were developed (Ransbotham et al. 2015). Swing and Ross (2016) argue that perhaps the best way forward is for staff and subunits to not only have significant access to institutional data through a federated system, but also have formal training to develop data expertise. The notion of federated data is, on first blush, a seemingly possible solution to ease the data reporting often required of IR professionals. While IR professionals should and must collaboratively work with other

colleagues at one's institution, the use of data (through a federated system or not) without it being managed and supported by data and context experts leaves it vulnerable. Specifically, the use of data without the proper context increases the likelihood of the wrong conclusions to be drawn from data and analyses. As such, the distribution of data should only occur after there are comprehensive plans in place complete with data definitions and documentation. Without the integration of data and context experts, the misuse and misapplication of data is likely resulting in increased poor academic planning and decision support.

IR and IR professionals, as a profession and individually, need to make the case for context expertise within institutions. If they do not, IR offices will be reformed and likely absorbed into another institutional operation like IT services or strategic planning (Calderon & Mathies 2013). This is where another dimension is worth discussing; the ethics involved with data use. Value judgement(s) over the appropriate use of data while providing data, drawing inferences about data, or acting as a data custodian is essential. This is where IR can step up and provide leadership on this issue within institutions. The old adage "information is power" is very appropriate in this context. As Calderon (2015) argues that:

having access to data and controlling the way it flows contains also an element of power, as data in an aggregated or as a value-added form can be used to allocate scarce resources, impose demands for accountability, and drive calls for improvement or reform at all levels. (p. 300)

Without questioning the power and legitimacy of data and its use, there cannot be a truly ethical use of data. Put another way, without the proper context and questioning its use, the use of data is simply functional. This is not a good thing as data used just functionally limits its potential effectiveness as well as its relevance. Without the ethical use of data, then there can be no achieving a truly effective and legitimate evidence-based decision making within an institution.

Second, institutions need to create, strengthen, and monitor data governance and the access to their data. It is important that IR professionals be deeply involved in the institution's data governance and management process as these are professionals who can provide subject expertise (ex: assessment of student success) and engage with the institution's data over its entire lifecycle (from capture/creation to use through the final archiving and destroying). This is critically important when an institution is trying to make meaning out of an analysis of data, as the deeper an individual's knowledge about an issue, the less likely they are to misunderstand the high level (aggregate) data that is generally presented in institutional reports. Conversely, the more superficial understanding an individual has, the more likely they are to confuse observed associations with casual considerations. As Calderon (2015, p.302) states "... the fact that something is counted or measured does not make it right or relevant."

In regards to data access, the use or development of policies following the principle of least privilege is a good starting point: access to all data is restricted by default, unless it is specifically allowed. In terms of data governance, require institution officials to have clear data definitions and rules of use, definitions, and collection. These data definitions and terms of use should be regularly updated and widely accessible for entire campus community. This promotes the holistic usage of data and supports a community wide understanding, appreciation, and use of data in agreed upon manners. This also increases the visibility of data and its likelihood of being seen as a good source or a reference and ultimately being used in institutional (policy, programs, etc.) development.

Third, with the increasing use of data analytics and metrics, there is a need for clear institutional principles and guidelines for their use and development. In particular, there are significant concerns over the use of third-party analytic tools. Many institutions do not have

the resources to have in-house analytic capabilities (staff, hardware, etc.) and these third-party options offer a way to use analytics without incurring the associated long-term costs. The main issue with most third-party analytic tools is their core algorithms are proprietary and not shared with clients (Alamuddin et al. 2016). This creates situations where institutions receive an analysis from a third-party analytic tool, but have no way to replicate or gauge the integrity and flexibility of the algorithms (Aladmuddin et al. 2016). In short, it “raises questions about the ethics of making decisions ... based on a black box that administrators, instructors, and students do not understand” nor control (Aladmuddin et al. 2016, p.22). “The Leiden Manifesto” (Hicks et al. 2015) offers suggestions on some the best practices for metric-based research assessments. While originally focused solely on research metrics, the 10 principles are applicable in a broad sense for institutional data. Some of the principles include keeping data collection and analytical processes open, transparent, and simple, allowing those evaluated to verify data and analysis, and scrutinize indicators recurrently with regular updates (Hicks et al. 2015). Having clear institutional guidelines on the use and development of data analytics and metrics will reduce the misapplication of data and promote sound data-driven decision making.

Data should and will be used in higher education decision support. But it is difficult to assess the appropriateness of data, from the information it conveys to its actual use, without a clear understanding of the context and purpose it was collected and used. Technological advances, along with the increasing accountability pressures and decrease in public resources, is boosting the use of data and data-driven decision making within institutions. IR leaders have an important opportunity to contribute to this conversation through guiding their institutions to properly use data and help build preventative practices. IR professionals can be an influential voice in what data and performance measurements are valid, suitable, and relevant for their institutions (Calderon 2015). Institutional decision making should be based on quality processes which are informed by the highest quality of data. But like with any tool, data and its analytics needs to be used appropriately and with care. If it is not, it will continue to be misused, and institutional decision making will not be as informed as well as it could.

Page Break

## References

- Association for Institutional Research. (2017). Development of the duties & functions of institutional research. Resource document. Association for Institutional Research. <http://www.airweb.org/Resources/Documents/Duties-and-Functions-Summary.pdf>. Accessed 1 Apr 2017.
- Alamuddin, R., Brown, J., & Kurzweil, M. (2016). Student data in the digital era: An overview of current practices. Resource document. Ithaka S+R. [http://www.sr.ithaka.org/wp-content/uploads/2016/09/SR\\_Report\\_Student\\_Data\\_Digital\\_Era-090616.pdf](http://www.sr.ithaka.org/wp-content/uploads/2016/09/SR_Report_Student_Data_Digital_Era-090616.pdf). Accessed 27 Apr 2017.
- Beatty, A., Greenwood, M.R.C, & Lin, R. (1999). *Myths and tradeoffs: The role of tests in undergraduate admissions*. Washington, DC: The National Academies.
- Beynon-Davies, P. (2011). *Significance: Exploring the nature of information, systems, and technology*. London, UK: Palgrave Macmillan.
- Beynon-Davies, P. (2013). *Business Information Systems (2<sup>nd</sup> ed.)*. London, UK: Palgrave Macmillan.
- Binette, C. (2016). Intrusion into UCF network involves personal data. Resource document. University of Central Florida – colleges & campus news. <https://today.ucf.edu/intrusion-into-ucf-network-involves-personal-data/>. Accessed 30 Mar 2017.

- Calderon, A. (2015). In light of globalization, massification, and marketization: Some considerations on the uses of data in higher education. In K. Webber and A. Calderon (Eds.), *Institutional research and planning in higher education: Global contexts and themes* (pp.288-306). Routledge Press: New York.
- Calderon, A. & Mathies, C. (2013). Institutional research in the future: Challenges within higher education and the need for excellence in professional practice. In A. Calderon & K. L. Webber (Eds.), *Global issues in institutional research, New Directions for Institutional Research, No. 157* (pp. 77-90). San Francisco: Jossey- Bass.
- Coburn, C. & Turner, E. (2012). The practice of data use: An introduction. *American Journal of Education*, 118(2), 99-111.
- Dahlstrom, E. (2016). *Digital capabilities in higher education, 2015: Analytics*. Louisville, CO: Educause.
- de Vise, D. (2012, February 27). Claremont Mckenna SAT scandal: More at stake than rankings? *Washington Post*. [https://www.washingtonpost.com/blogs/college-inc/post/claremont-mckenna-sat-scandal-more-at-stake-than-rankings/2012/02/07/gIQAHIwVwQ\\_blog.html?utm\\_term=.343e4bc2231d](https://www.washingtonpost.com/blogs/college-inc/post/claremont-mckenna-sat-scandal-more-at-stake-than-rankings/2012/02/07/gIQAHIwVwQ_blog.html?utm_term=.343e4bc2231d). Accessed 30 Mar 2017.
- ETS. (2017). About the GRE general test. [https://www.ets.org/gre/revised\\_general/about/? WT.ac=grehome\\_greabout\\_b\\_150213](https://www.ets.org/gre/revised_general/about/?WT.ac=grehome_greabout_b_150213). Accessed 17 Mar 2017.
- Fain, P. (2017, January 16). College scorecard screwup. *Inside higher ed*. <https://www.insidehighered.com/news/2017/01/16/feds-data-error-inflated-loan-repayment-rates-college-scorecard>. Accessed 23 Mar 2017.
- Grama, J. (2014). Data breaches in higher education. Resource document. EDUCAUSE center for analysis and research. <https://net.educause.edu/ir/library/pdf/ECP1402.pdf>. Accessed 22 Mar 2017.
- Gray, C. (2016, February 16). Senate unanimous in bill protecting student medical records. *The Lund report*. <https://www.thelundreport.org/content/senate-unanimous-bill-protecting-student-medical-records>. Accessed 12 Feb 2017.
- Hazelkorn, E. (2015a). The obsession with rankings in tertiary education: Implications for public policy. Resource document. World Bank. [https://hepru.files.wordpress.com/2015/01/the-obsession-with-rankings-in-tertiary-education\\_wb\\_0115.pdf](https://hepru.files.wordpress.com/2015/01/the-obsession-with-rankings-in-tertiary-education_wb_0115.pdf). Accessed 20 Mar 2015.
- Hazelkorn, E. (2015b). *Rankings and the reshaping of higher education: The battle for world-class excellence (2<sup>nd</sup> ed.)*. London: Palgrave MacMillan.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden manifesto for research metrics. *Nature*, 250, 429-431.
- Higgins, K. (2016). Post-truth: A guide for the perplexed. *Nature*, 540, 9.
- Horta, M., Bouwma-Gearhart, J., & Park, H.J. (2017). Data driven decision making in the era of accountability: Fostering faculty data cultures for learning. *The Review of Higher Education*, 40(3), 391-426.
- Jaschik, S. (2012, April 18). Claremont McKenna admits extent of deception on admissions statistics. *Inside higher ed*. <https://www.insidehighered.com/news/2012/04/18/claremont-mckenna-admits-extent-deception-admissions-statistics>. Accessed 17 Mar 2017.
- Knight, W., Moore, M. & Coperthwaite, C. (1997). Institutional research: Knowledge, skills, and perspectives of effectiveness. In J.F. Volkwein (Ed.), *What is institutional research all about? A critical and comprehensive assessment of the profession* (pp.31-39). New Directions for Institutional Research, No. 104. San Francisco: Jossey-Bass.

- Koltay, T. (2016). Data governance, data literacy, and the management of data quality. *IFLA Journal*, 42(4), 303-312.
- Lindquist, S. (1999). A profile of institutional researchers from AIR national membership survey. In J.F. Volkwein (Ed.), *What is institutional research all about? A critical and comprehensive assessment of the profession* (pp.41-50). New Directions for Institutional Research, No. 104. San Francisco: Jossey-Bass.
- Mahaffie, L. (2017). Updated data for college scorecard and financial aid shopping sheet. Resource document. U.S. Department of Education.  
<https://ifap.ed.gov/eannouncements/011317UpdatedDataForCollegeScorecardFinaidShopSheet.html>. Accessed 20 Mar 2017.
- Marginson, S. (2014). University rankings and social sciences. *European Journal of Education*, 49(1), 45-59.
- McKinney, M. (2016, March 25). Data breach exposes information on more than 3000 TCC employees. *Virginian-Pilot*. [http://pilotonline.com/news/local/crime/data-breach-exposes-information-on-more-than-tcc-employees/article\\_6ab72a2f-52a0-533e-8060-a2d245c7f151.html](http://pilotonline.com/news/local/crime/data-breach-exposes-information-on-more-than-tcc-employees/article_6ab72a2f-52a0-533e-8060-a2d245c7f151.html). Accessed 23 Mar 2017.
- McLaughlin, G., Howard, R., & Jones-White, D. (2012). Analytic approaches to creating planning and decision support information. In R. Howard, G. McLaughlin, & W. Knight (Eds.), *The handbook of institutional research* (pp. 459-477). Jossey-Bass: San Francisco.
- National Association for College Admission Counseling (2008). *Report of the commission on the use of standardized tests in undergraduate admission*. Arlington, VA: National Association for College Admission Counseling.
- New. J. (2015, August 3). Staying confidential. *Inside higher ed*. [https://www.insidehighered.com/news/2015/08/03/privacy-loophole-remains-open-after-outrage-over-u-oregons-handling-therapy-records#disqus\\_thread](https://www.insidehighered.com/news/2015/08/03/privacy-loophole-remains-open-after-outrage-over-u-oregons-handling-therapy-records#disqus_thread). Accessed 19 Feb 2017.
- Oxford University Press. (2017). data, n. *OED Online*.  
<http://www.oed.com/view/Entry/296948?redirectedFrom=data#eid>. Accessed 12 Mar 2017.
- Patterson, L. (2015). Analytics and metrics: related but not the same. Resource document. MarketingProfs LLC.  
<http://www.marketingprofs.com/articles/2015/27462/analytics-and-metrics-related-but-not-the-same-an-explanation-and-a-checklist>. Accessed 11 Mar 2017.
- Posselt, J. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Cambridge, MA: Harvard University Press.
- Possey, J. & Pitter, G. (2012). Integrating the functions of institutional research, institutional effectiveness, and information management. *AIR Professional File*, 126, 1-32.
- Ransbotham, S., Kiron, D., and Prentice, P.K. (2015). Minding the analytics gap. Resource document. MIT Sloan management review. <http://sloanreview.mit.edu/article/minding-the-analytics-gap/>. Accessed 22 Mar 2017.
- Read, R. (2015, June 4). UO whistleblowers: giving student's confidential therapy records to campus lawyers felt wrong. *The Oregonian*. [http://www.oregonlive.com/education/index.ssf/2015/06/uo\\_whistleblowers\\_giving\\_stude.html](http://www.oregonlive.com/education/index.ssf/2015/06/uo_whistleblowers_giving_stude.html). Accessed 27 Feb 2017.
- Robinson-Garcia, N. & Calero-Medina, C. (2014). What do university rankings by fields rank? Exploring discrepancies between the organizational structure of universities and bibliometric classifications. *Scientometrics*, 98(3), 1955-1970.

- Ronco, S., Archer, S., & Ryan, P. (2012). Tools for measuring the effectiveness of institutional research. In R. Howard, G. McLaughlin, W. Knight (Eds.), *The handbook of institutional research* (pp. 673-687). Jossey-Bass: San Francisco.
- Spillane, J. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American journal of education*. 118(2), 113-141.
- Sposito, S. (2016, February 27). UC Berkeley system with data on students, workers, and alumni hacked. *San Francisco Chronicle*. <http://www.sfgate.com/news/article/UC-Berkeley-system-with-data-on-students-6857364.php>. Accessed 15 Mar 2017.
- Strauss, V. (2012, August 23). Elite university lies to college rankers for years. *Washington Post*. [https://www.washingtonpost.com/blogs/answer-sheet/post/does-it-matter-that-an-elite-university-lied-to-college-rankers-for-years/2012/08/22/e89b39d2-ec68-11e1-aca7-272630dfd152\\_blog.html?utm\\_term=.48cabcd2a554](https://www.washingtonpost.com/blogs/answer-sheet/post/does-it-matter-that-an-elite-university-lied-to-college-rankers-for-years/2012/08/22/e89b39d2-ec68-11e1-aca7-272630dfd152_blog.html?utm_term=.48cabcd2a554). Accessed 17 Mar 2017.
- Swing, R. & Ross, L.E. (2016). A new vision for institutional research. *Change*, 48(2), 6-13.
- Swing, R., Jones, D., & Ross, L. E. (2016). The AIR national survey of institutional research offices. Resource document. Tallahassee, FL: Association for Institutional Research. <https://www.airweb.org/Resources/ImprovingAndTransformingPostsecondaryEducation/Documents/National-Survey-of-IR-Offices-Report.pdf>. Accessed Apr 2 2017.
- Taylor, J., Hanlon, M., & Yorke, M. (2013). The evolution and practice of institutional research. In A. Calderon & K. L. Webber (Eds.), *Global issues in institutional research* (pp. 59-75). New Directions for Institutional Research, No. 157. San Francisco: Jossey- Bass.
- Teichler, U. (2011). Social consequences of university rankings: A meta-analysis of the ranking literature. In J.C. Shin, R. Toutkoushian, & U. Teichler (Eds.), *University rankings: Theoretical basis, methodology and impacts on global higher education* (pp. 55-69). Dordrecht: Springer.
- Terenzini, P. (1993). On the nature of institutional research and the knowledge and skills it requires. *Research in Higher Education*, 34, 1, 1-10.
- Terenzini, P. (2013). "On the nature of institutional research" revisited: "Plus ca Change ..."? *Research in Higher Education*, 54, 137-148.
- U.S Department of Education (2015). Resource document. FERPA. <https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html?src=rn>. Accessed 11 Feb 2017.
- Visser, H. & Barnes, G. (2016). Professional development for institutional research. In J. Botha and N. Muller (Eds.), *Institutional research in South African higher education* (pp. 75-96). Stellenbosch: Sun Press.
- Volkwein, J. F. (1990). The Diversity of Institutional Research Structures and Tasks. In J.B. Presley (Ed.), *Organizing effective institutional research offices* (pp.7-26). New Directions for Institutional Research, No.66. San Francisco: Jossey- Bass.
- Volkwein, J.F. (2011). Gaining ground: the role of institutional research in assessing student outcomes and demonstrating institutional effectiveness (occasional paper 11). Resource document. National institute for learning outcomes assessment. <http://www.learningoutcomeassessment.org/documents/Volkwein.pdf>. Accessed 3 Apr 2017.
- Whitehurst, G.J. & Chingos, M. (2015, October 15). Deconstructing the reconstructing the College Scorecard. *Evidence Speaks Reports*. <https://www.brookings.edu/research/deconstructing-and-reconstructing-the-college-scorecard/>. Accessed 14 Apr 2017.
- Wightman, L. (2003). Standardized testing and equal access: A tutorial. In M. Chang, D. Witt, J. Jones, & K. Hakuta (Eds.), *A compelling interest: Examining the evidence on*

*racial dynamics in colleges and universities* (pp 49-96). Stanford, CA: Stanford University press.

Young, A. & McConkey, K. (2012). Data governance and data quality: Is it on your agenda? *Journal of Institutional Research*, 17(1), 69-77.