INSTITUTIONAL DATA QUALITY AND THE DATA INTEGRITY TEAM

Katherine A. McGuire

About the Author
Katherine A. McGuire is director of institutional research at Oxford College of Emory University.

Abstract
Data quality has become a pressing issue for many campuses in recent years, as colleges struggle to extract timely, accurate, and consistent information from ever-proliferating institutional data sources in order to meet strategic decision-making and accountability demands. In this mixed methods study, a survey and semi-structured interviews were used to examine data integrity teams, which are groups that try to improve the accuracy and usefulness of data in computing systems at institutions of higher education. A survey sent to a random sample of institutional researchers revealed that many campuses did not have data integrity teams. Where campuses had them, those teams frequently did not engage in activities like data auditing, creating or maintaining data standards documentation, or training staff on data standards issues. Interview participants from campuses with an established team reported that the greatest benefits were greater communication, collaboration, and awareness of data quality issues. Both survey respondents and interviewees reported that more data governance resources, including dedicated staff time, were needed to improve data quality. The implications of these findings for strategic data quality and best practices for institutions are discussed.

Keywords: Data quality, data governance

BACKGROUND

Computerized database systems have created a revolution in the capacity of organizations to store and rapidly retrieve information about their processes and people. The routine operations of colleges and universities have been profoundly affected by these broad-based changes in information management. All administrative and academic departments on a campus require access to information contained in institutional databases for their daily activities, whether it be directory information, student enrollment and academic records information, financial aid data, accounting and billing data, faculty and staff personnel data, donor records, grants management data, or facilities and scheduling information. In recent years, demand for information for accountability, institutional decision-making, and planning has placed increased scrutiny on data quality and data processes at postsecondary institutions.

Since early in the development of the field of institutional research, practitioners have expressed concern about the accuracy of data contained in student information systems. In a 1989 Association for Institutional Research (AIR) Professional File paper entitled "Data Integrity: Why Aren't the Data Accurate?," Gose described a number of major types of data errors, and noted that the human element was essential in maintaining data systems free from various types of "data corruption." By "human element," he presumably meant that improving communication between departments and individuals about data problems and data standards is crucial to improving data quality.

McGilvray points out that a persistent problem with data quality is that data management is one area where the trend toward greater integration and collaboration in organizations has lagged behind: "Our applications and business needs for information are integrated, but our behavior has not

changed to work effectively in this world. For example, your company may need information to support end-to-end processes and enterprise decision-making, but the information is being created by an individual contributor from the business who has no visibility to other needs for the same information" (2006, p. 2).

Thus, data entry responsibilities frequently fall to the lowest-ranking and newest member of a department, someone who does not understand the needs of end users and in whom just enough training is invested to get the job done at a basic transactional level. Such employees tend to be rewarded for speed rather than accuracy, and often the department where data entry occurs is not directly impacted by data errors.

Colleges have adopted various strategies for improving campus data, all of which could be described by the umbrella term "data governance." Such strategies might include codifying data standards, creating standard operating procedures for data processes, developing master data sets for reporting, and assigning to specific personnel oversight of data in campus functional areas. All these strategies require that critical stakeholders regularly communicate and collaborate to identify problems, set standards and policy, oversee and review data and data processes, and help manage change that impacts data integrity. Some college campuses have instituted data integrity teams to serve this function. Data integrity teams are groups of stakeholders from diverse functional areas on campus that meet regularly to try to collaboratively

address data problems as they arise, as well as to proactively implement improved data management policies and procedures.

Young and McConkey (2012) and McLaughlin, Howard, Cunningham, and Payne (2004) have described many of the activities that are appropriate for data integrity teams in higher education. Teams should first identify data stakeholders and their needs. They should institute consistent data definitions across the institution, such as by creating a data dictionary, and they should establish data use rules. They should draft data policies, communicate the importance of those policies, and monitor and report both the status of data quality efforts and compliance with standards. They should assign data stewards or custodians so that there is no ambiguity about who is responsible for data in a given area, and they should update such assignments when necessary. They should seek to understand external accountability and internal research and planning data needs, and should incorporate these needs into data standards decisions. Teams should be aware of data quality issues surrounding documentation, process gaps, and missing data. They should address issues of access, security, and integration of multiple data systems. Finally, data integrity teams should track how data decisions are made, as well as how conflicts between departments or members are resolved.

The present study examined the staffing, scope of activities, institutional environments, and effectiveness of data

integrity teams on college campuses by means of a concurrent mixed methods research design, including an online survey and semi-structured interviews of postsecondary data users. Some of the research questions the study addressed were these:

1. What percentage of postsecondary institutions have formal data integrity teams? Can any institutional characteristics or organizational conditions be identified that seem to promote the development of data integrity teams?
2. Who typically serves on data integrity teams? Which institutional departments play leadership roles in data governance?
3. How well are data integrity teams supported by executive leadership and what authority do teams have to make and enforce data policy?
4. What are the typical tasks undertaken by a data integrity team? How effective do team members believe their teams are at solving various types of data quality problems?
5. What do team members perceive as the barriers to institutional data quality? How do they think these might be overcome? Are there any types of data problems that are insurmountable or unavoidable?

In the first phase of the study, randomly selected members of the higher education professional association, AIR, were invited to participate in a 20-minute online survey that asked

questions about the demographic characteristics of their institution and whether it had a data integrity team. If the institution had such a team, questions followed as to who served on the team, core team activities, and team accomplishments and challenges. A second qualitative phase of the study interviewed individual data integrity team members at postsecondary institutions about their teams' activities and challenges. This study differs from previous data integrity research done by higher education information technology (IT) groups like EDUCAUSE (see Yanosky 2009) by focusing on the perceptions of professional institutional researchers rather than on IT leadership or staff, as well as in having a qualitative component.

METHODOLOGY

The quantitative phase of the study consisted of an online survey created and maintained in the online web survey tool SurveyMonkey (www.surveymonkey.com), and administrated by AIR. The survey contained item tracking so that AIR members whose institutions did not have data integrity teams or who were not members of their schools' data integrity teams answered a different set of questions than respondents who were on campuses and/or served on data integrity teams. A sample of 519 randomly selected members of AIR were sent an e-mail from AIR explaining the purpose of the survey and inviting them to participate by clicking on a hyperlink in the e-mail message. Descriptive data analysis was performed using Statistical Package for the Social Sciences (SPSS).

Table 1. FTE Enrollment of Respondents' Institutions

FTE Enrollment	Frequency	Percent
Fewer than 1,000	18	11%
1,000–2,999	36	21%
3,000–9,999	55	32%
10,000–19,999	36	21%
20,000 or more	27	16%
Total	172	

Note: FTE = full-time equivalent.

The qualitative phase of the study consisted of structured individual interviews. Each interview subject was a data integrity team member from a different postsecondary institution. Participants were recruited through the e-mail lists of two institutional research groups: the Georgia Association for Institutional Research, Planning, Assessment, and Quality (GAIRPAQ) and the Higher Education Data Sharing Consortium (HEDS). Additional potential subjects were located by a Google search of terms such as "university data integrity team," "college data governance," etc., and e-mail contact was made with relevant staff at institutions for whom data integrity team information was found online. Subjects were interviewed by phone using the online tool Skype, and interviews were recorded to MP3 files using the Skype recording tool Evaer. All semi-structured interviews were transcribed manually from the MP3 files, and the resultant data were coded and analyzed in QDA Miner Lite. Both thematic and content analyses were performed where appropriate.

RESULTS

Survey Results

A total of 205 AIR member respondents submitted the survey, for a 39% response rate. Of these, 197 responded to at least one item on data integrity and were included in the final analysis of survey results.

The majority (87%) of respondents were employed at postsecondary institutions. Of the 172 respondents employed on postsecondary campuses, by far the largest group was at institutions with both undergraduate and postgraduate programs (66%). Smaller percentages of respondents were from institutions with two-year (22%), four-year only (9%), and graduate-only (3%) programs. There were slightly more respondents from public (55%) than from private institutions; only four respondents (2%) were from private proprietary schools. The diversity of institutional student enrollment sizes represented in the sample can be seen in Table 1. Exactly half of the respondents were

at multicampus systems, illustrating the potential complexity of data management at the institutions in the study.

Fewer than half (44%) of the 172 respondents from postsecondary institutions reported that their school had a data integrity team, and only 38 respondents (22%) reported leading or serving on a data integrity team. Table 2 shows the institutional characteristics of institutions that had data integrity teams.

Executive Advocacy of Data Quality Efforts

Respondents indicated they believed that campus executive leaders were overall supportive of efforts to improve data quality (see Table 3). With the exception of the chief business officer, whose rating decreased slightly when disaggregated, this confidence in leaders' support of data integrity was even more pronounced for respondents who were members of their institutions' data integrity teams.

Respondents' Ratings of Institutional Data Quality

Sixty-six percent of all institutional respondents said that they "Agreed" or "Strongly agreed" with the statement, "The overall quality of data in my institution's administrative computing system is high." There was virtually no difference in the percentage of respondents who rated institutional data quality highly who were on data integrity teams from those who were not. Respondents who reported that their campus did not have a data integrity team were asked why they thought it did not (see Table 4).

Table 2. Characteristics of Institutions with Data Integrity Teams

Institutional Characteristics	Number of Institutional Respondents with Data Integrity Team*	Percent of Institutional Respondents with Data Integrity Team*
Institutional Type		
Two year	18	47%
Four year only	5	46%
Four year plus graduate and/or professional	48	52%
Graduate and/or professional only	4	80%
Institutional Control		
Private for-profit	1	25%
Private not-for-profit	36	58%
Public	38	49%
Institutional FTE		
Fewer than 1,000	5	36%
1,000–2,999	19	59%
3,000–9,999	27	53%
10,000–19,999	15	54%
20,000 or more	9	50%

Note: FTE = full-time equivalent.
* "I don't know" and "No response" omitted from numerator and denominator.

Table 3. Support of Campus Leaders for Data Integrity Efforts

The following campus leaders support efforts to address data integrity at my institution (Strongly agree or agree)	All Institutional Respondents (n=169)	Data Integrity Team Members Only (n=32)
President/Chief executive officer	56%	76%
Provost/Chief academic officer	69%	90%
Chief business officer/Chief financial officer	68%	61%
Chief student affairs officer	56%	68%
Chief Information officer	71%	84%

Table 4. Reasons Respondents' Institutions Do Not Have Data Integrity Teams

To the best of your knowledge, what are the reasons that your institution does not have a data integrity team? (check all that apply) (n=70)	Percent
Data quality is not a problem at my institution.	14%
Data quality issues are too contentious/political.	20%
Decision-makers are not aware of data quality issues.	27%
Decision-makers are not interested in data quality issues.	20%
Decision-makers do not have time to devote to data quality issues.	40%
Decision-makers do not have resources to devote to data quality issues.	43%

Composition and Leadership of Data Integrity Teams

Over 80% of survey respondents who were on data integrity teams worked in institutional research or assessment offices, as might be expected given the population sampled. As shown in Table 5, by far the most common functional area of team leaders was institutional research and related departments,

followed by IT. Various other leader functional areas were mentioned in the open-ended comments for this survey item, including associate vice president and bursar, as well as cochairing arrangements.

Additional team members mentioned in the open-ended comments sections were online or e-learning coordinators,

athletics, career services, the veterans' affairs office, and student life.

Data Integrity Team Characteristics

Over 80% of the respondents who served on their institution's data integrity team had been on the team for more than three years, and only about 15% had served for less than a year. The most common regular meeting schedules were monthly (24%) or quarterly (18%); a combined 32% said they met either irregularly or on an as-needed basis rather than keeping a regular schedule.

About 30% of the respondents said their data integrity team reported to the institutional research, institutional effectiveness, or assessment functional area. Another 16% reported to IT, 13% reported to academic affairs, and about 10% reported to the president or chief executive officer. A few other teams reported to executive cabinets or other entities. Several respondents said that their team either did not report to anyone or that they were not sure who their team reported to. Respondents indicated that the team reported to the individual or entity that oversaw it by face-to-face meetings or presentations (42%), memos or reports (13%), or both methods (40%). Most teams reported that they had only a limited range of data policy-making authority and that they referred data policy violators to another entity or person (see Table 6).

Team Activities and Effectiveness

Data integrity team members reported their team doing a variety of common data quality–related activities, as

Table 5. Team Leader's Department and Representation on Team

	Team Leader's Department (n=32)	Represented on Team (n=41)
Institutional research/Institutional effectiveness/Assessment	47%	100%
IT/Computing	24%	71%
Other (please specify)	16%	13%
Academic affairs/Faculty	3%	58%
Admissions/Enrollment management	3%	71%
Development/Advancement	3%	34%
Registrar	3%	79%
Business/Accounting	0%	66%
Financial aid	0%	74%
Human resources	0%	45%

Table 6. Team Authority to Make and Enforce Data-related Policy

Which best describes the team's authority to make data-related policy on your campus? (n=33)	Percent
We have a broad range of policy-making authority.	23%
We have a limited range of policy-making authority.	45%
We can make recommendations only.	29%

Which best describes the team's authority to enforce data-related policy on your campus? (n=32)	Percent
We have policy enforcement authority (e.g., can limit data systems access).	13%
We refer individuals who violate data policies to other entities (e.g., their supervisors).	53%
We have no authority to enforce policy.	27%

Note: "Other" responses not included.

summarized in Table 7. The activities that were most often cited as a focus of the team were identifying data gaps and inconsistencies, identifying data stewards, and considering institutional strategic reporting needs. The two items that respondents cited least often as being a focus of the team concerned data auditing and policy assessment.

Team members also reported on institutional and departmental environments and outcomes for data quality, as shown in Table 8. Although respondents indicated that advocacy and awareness of data quality issues existed on their campuses, only slightly over half agreed that having a data integrity team had improved institutional data quality. Many of the typical activities associated with data integrity teams, such as creating data documentation, training staff, documenting data steward responsibilities, and monitoring data quality, were occurring at a third or fewer of the institutions. Only a quarter of the respondents agreed that data users knew the procedure for reporting data problems.

Views of Non-Team Members on Data Integrity Practices

As noted previously, many of the AIR member respondents either did not serve on their campus data integrity team, were employed on a campus that did not have a data integrity team, or were not employed on a college campus. Respondents who reported that they were not currently on data integrity teams answered opinion questions about data quality issues on college campuses. Of these respondents, 85% agreed with the

Table 7. Frequency of Data Integrity Team Activities and Team Effectiveness

How often does the data integrity team focus on the following issues, and how effective is the team in each area?	Frequency of Team Activities (Percent "Sometimes" or "Often") (n=32)	Team Effectiveness (Percent "Effective" or "Highly effective") (n=31)
Identify data gaps and inconsistencies.	97%	66%
Identify data stewards (people responsible for maintaining data quality and reporting data issues).	97%	68%
Consider internal strategic data reporting needs.	93%	54%
Create new data policies.	90%	55%
Review current data policies.	87%	71%
Align data policies between departments.	87%	54%
Seek input from data stakeholders.	86%	57%
Address compliance or regulatory issues.	86%	61%
Establish needs, roles, and responsibilities of data stewards.	86%	58%
Determine who has or needs access to data.	79%	61%
Assess effectiveness of data policies.	79%	48%
Monitor data quality.	79%	57%

statement, "Every college or university should have a data integrity team." The majority of respondents (55%) believed that data integrity teams should report to the office of institutional research or institutional effectiveness; only 11% stated that the team should report to an IT function.

Respondents were also asked what they thought the activities of a data integrity team should be (see Table 9). The activities that respondents not on a data integrity team were likely to think most important differed somewhat from the activities that data integrity team members reported as teams' most frequently addressed issues, with data auditing and policy assessment assuming greater importance to the non-team-member respondents.

About a third of the respondents not currently on data integrity teams had served on one in the past; of these respondents, 65% rated their previous data integrity team to be highly or moderately effective.

Open-Ended Survey Comments
Around two dozen respondents gave additional reasons or commentary about why their institution did not have a data integrity team. About a third of the comments indicated that data quality issues were handled in an informal, ad hoc manner in response to specific problems or projects with whatever departments were impacted

Table 8. Institutional Environments and Activities for Data Quality Reported by Data Integrity Team Members

Indicate your level of agreement with the following statements about your institution:	Percent "Agree" or "Strongly Agree" (n=37)
My supervisor is aware of the importance of data quality.	90%
Data integrity team members serve as advocates for good data in their departments.	77%
Data quality is a strategic priority.	65%
Data stewards/managers exist in each functional unit that has data access and responsibilities.	58%
Having a data integrity team on my campus has improved data quality.	55%
Data quality is continuously monitored.	48%
Significant resources are devoted to data quality improvement efforts.	42%
The institution has a usable and complete data dictionary.	33%
All data users have easy access to data field documentation.	32%
Staff who work with data receive training about data standards.	32%
Data steward/manager responsibilities are clearly documented.	30%
There are regularly scheduled comprehensive data quality audits.	26%
Individuals who use data know how to report a problem or issue with data quality.	26%

by the particular issue. Similarly, several other respondents indicated that data quality issues were handled in a decentralized fashion within departments. Three participants said that they had previously had a data integrity team that had stopped meeting, and several others said that their institution was in the process of forming a data integrity team. Two respondents expressed the belief that data integrity teams were not useful because data quality issues were too complex to be solved by a single team.

Most of the respondents who served on data integrity teams commented on how data integrity could be improved at their own institution. Typical comments cited the need for more buy-in by both senior leadership and staff. More centralization of data quality efforts and user accountability for data quality were also mentioned by several respondents. Training for data users was one of the most frequently mentioned needs, as was creating or updating a data dictionary. The need for additional staff was a concern, and several respondents said that they believed their institution needed dedicated staff to oversee data integrity issues.

About 40% of the respondents not currently serving on a data integrity team answered the open-ended question, "How can data integrity be improved at institutions?" Twenty-five percent of the comments mentioned the need for greater executive buy-in and accountability, and nearly 20% of comments mentioned the need for some kind of accountability for data entry or data reporting staff. As Table 10 shows, team members and non-team members mentioned similar data quality solutions.

There were also several comments from both team members and non-team members about the need to understand the origins of information and filter out bad data before such data got into centralized data systems,

Table 9. Top Five Activities that Respondents Not on a Data Integrity Team Indicated Should Be Part of the Charge of a Data Integrity Team

What activities should be part of the charge of a data integrity team? Select all that apply. (n=139)	Percent of Respondents
Identify data gaps and inconsistencies.	94%
Review current data policies.	93%
Assess effectiveness of data policies.	88%
Monitor data quality.	87%
Seek input from data stakeholders.	87%

by technical validation or automation where appropriate: "Garbage in = garbage out. One of the most difficult challenges is controlling quality and consistency from point of entry."

SEMI-STRUCTURED INTERVIEW RESULTS

Demographics of Participants and Their Institutions

Interviewees were data integrity team members from seven institutions in the continental United States. Six participants were institutional research or institutional effectiveness administrators at the director level or higher; the other was an IT manager who specialized in data governance. Several different Carnegie types were represented among the institutions in the interview sample, including four baccalaureate colleges, one master's college, and two research universities. Regionally the South, Mid-Atlantic, Northeast, Pacific Northwest, and Midwest were represented. Six of the interviewees came from private

not-for-profit institutions, and one was from a public institution. The total enrollments of the institutions ranged from just over 2,000 students to nearly 26,000 students.

Cross-case Analysis

As seen in Table 11, participants' institutions are compared side by side on a number of variables relevant to data integrity. These data were derived from the interview transcripts; in a small number of cases participants were not sure how to answer a question or became sidetracked to another issue when they were asked about it due to the loosely structured and organic nature of the interviews, so that the information could not be clearly ascertained from the transcripts.

Team Structure, Membership, and Leadership

There was a wide degree of variability in the structure of the data integrity teams represented in the sample. Some data integrity teams were effectively user groups for the main student

information system on the campus, while others were outgrowths of the institution's business intelligence units. Sometimes there was only one team on a campus involved with data integrity, but at some institutions there were several teams with different specific functions. In some cases, this diversification of the data integrity function had to do with a working group of middle managers needing to rely on a higher-level executive committee to make policy; in other cases, it had to do with the size and complexity of the institution and the data issues encountered.

For some of the data integrity teams, particularly those that functioned as user groups for a specific data system (e.g., Datatel or Banner), membership was voluntary for those who had an interest in solving problems with institutional data. At other institutions, data integrity team membership was part of the job description for manager positions that involved working with data. Additionally, attendance might be expected at all meetings for some core members, while other staff attended only when there was a specific issue or problem being discussed that required their input.

Despite this variability in team structure between campuses, there was a relatively high degree of uniformity in the functional roles that were represented on campus teams. Typically, a single representative from each relevant department participated on the team. As might be gleaned from the demographic description of the study participants, institutional research and IT offices

Table 10. Topical Summary of Open-Ended Comments on How Institutions Can Improve Data Integrity

How can data integrity be improved at your institution/at institutions?	Data Integrity Team Member (Percent of Comments; n = 25)	Not on a Data Integrity Team (Percent of Comments; n = 52)
Increased accountability.	16%	19%
More/better training.	16%	12%
Greater executive buy-in.	12%	25%
Greater staff buy-in.	12%	10%
Centralization of data integrity efforts.	12%	6%
Dedicated staff.	12%	6%
Create/improve data dictionary.	12%	2%
Better communication or collaboration.	8%	14%
More staff overall.	8%	4%
Different unit in control of data integrity.	8%	Not mentioned
Automation of data entry or data validation.	4%	4%
More local unit autonomy in data quality decisions.	4%	Not mentioned
More time devoted to data quality.	4%	Not mentioned

were represented on such teams, and were frequently leaders or occasionally cochairs of the team. Additionally, staff from the registrar's office, financial aid, human resources, academic affairs, student affairs, and admissions office were members of nearly all the teams. Staff members from business and accounting, as well as development and alumni affairs, were represented at some but not all the institutions included in the interview sample. The differences in team membership and structure were often reported to be due to the existence of multiple different data systems on campuses, such as separate athletics, admissions, communications, or advancement databases, for example. Participants indicated that this multiplicity of data systems added an additional layer of complexity to data quality. Sometimes the data integrity team included users of a number of databases, and sometimes it included only users of the main student information system, which could be problematic when one database was used to populate another.

Activities and Processes of Teams

Different teams had different regular meeting schedules and agendas. Most typically, the main data integrity team met once a month. The frequency of team meetings seemed to vary with the structure of the data integrity function: the two teams with business intelligence or an analytics function were those meeting weekly. Typically, a meeting agenda was created at least in part from a call for topics, issues, or problems from team members.

Table 11. Cross-case Analysis of Interview Participant Data

Participant	Name of Team	Team Leader	Entity to Which Team Reports	Data Dictionary	Data Warehouse	Executive Sponsor
Participant A	Data management group	IT staff person	Administrative computing advisory group	Yes	Yes	None mentioned
Participant B	Data quality/data governance	Business intelligence manager	No formal reporting structure	Yes	Yes	None mentioned
Participant C	Data governance	None mentioned	Provost	Yes	Yes	VP for IE, provost
Participant D	Data standards group	IR	Steering committee composed of data stewards	Yes	Yes	Provost
Participant E	Data standards committee	Cochaired by IR and an academic dean	Executive-level cabinet	No	No	Academic affairs associate dean
Participant F	Data committee	CIO	Large ad hoc group of VPs	No	No	None
Participant G	Users group	Cochaired by IR and IT	Voluntary group, no formal reporting structure	No	No	CIO

Note: CIO = chief information officer; IE = institutional effectiveness; IR = institutional research; IT = information technology; VP = vice president.

Additionally, team meetings also usually spent time on updates of ongoing data quality projects. A few teams had regular reports from specific offices or groups, such as IT staff that were working on projects that might affect data and impact data users:

Participant: So we meet monthly. And we have split the meeting into several different things that happen. One thing that happens is that our project manager for our PeopleSoft implementation always gives an update because this is the only place where people who are not at very high levels can find out what's going on with our implementation. . . . For example, we're thinking of purchasing some BI [business intelligence] tools. The people who are going to have to work with these BI tools are the people at the data

standards committee meeting, not the cabinet. And so those are the people who need to know that this might be happening. . . . This is the only place where that . . . where they get that kind of update. So we always devote part of our meeting to that.

Problem solving and change management were activities of the data integrity groups in the study that were frequently mentioned. Typically, problems or projects were submitted to the committees as an agenda item:

Participant: So once a month we put out a call for topics. We really just ask people, so OK, what's rubbing the wrong way? What's an issue now? And people bring these things up.

Additionally, changes in externally mandated compliance reporting or changes to institutional programs requiring adjustments to data collection and reporting strategies were often brought up in the data integrity teams. Examples of external policy changes that were mentioned were the change to the current federal Integrated Postsecondary Education Data System (IPEDS) race and ethnicity and human resources reporting standards. Technology changes, such as data system conversions or upgrades, might also typically be discussed in the data integrity group.

Frequently mentioned was the need for the formation of subgroups or subcommittees of team members with a particular interest or expertise in a

specific data problem. Sometimes this was an issue of change management. These subcommittees would occasionally draw on personnel who were not regular members of the data integrity team if their expertise or input was needed. The typical protocol seemed to be for these subcommittee members to work on a problem outside the data integrity team, and then report back to and seek feedback from the team at its regular meetings until a data issue was resolved.

Data dictionaries were sometimes an activity of the data integrity team. Four of the participating institutions had data dictionaries and three did not. Both of the research institutions had data dictionaries and, perhaps not coincidentally, also had business intelligence models for reporting and analytics. Almost all the schools that had data dictionaries also had data warehouses, so it is probable that there is a relationship between the two outcomes. One of the research university participants belonged to a school that used the Data Cookbook, a commercially available data dictionary tool. This institution's participant described the tool as playing a positive role in developing consistent and accessible data standards and processes across campus, but also admitted that implementation and maintenance of the technology had been labor intensive.

Authority

Authority of the team to make and enforce data policy was handled in a number of different ways at the campuses in the study. Some teams had a clear charge from

executive leadership while others were exclusively voluntary in nature. Teams seldom seemed to have broad authority to make data policy decisions. As indicated in the cross-case analysis, the usual arrangement was for a group of midlevel data managers to make data decisions at the field or project level, but to defer to an executive body on campus-wide policy decisions. Also noted in the cross-case analysis was that only about half of the participants reported having an executive advocate. Those that did spoke highly of the value of having an executive-level sponsor for data quality, particularly at the point of getting data integrity teams started:

Participant: And we have an advocate with my vice president, thank God, who used to be the CIO [chief information officer] here. . . . She's just that type of person that can just . . . that runs everything. But she's been a huge advocate for us. . . .

Interviewer: So she knows what the issues are.

Participant: Yeah. And you have to have an advocate, I would say. At least one.

There was some ambivalence from interviewees in response to questions about how much support data integrity teams and their efforts got from executive leadership. On the one hand, participants seemed to believe that leadership generally was supportive of the team itself. Where teams referred policy or strategic data decisions to an executive steering committee, participants reported that

the steering committee respected their expertise and was willing to endorse their recommendations on most data policy matters.

> *Interviewer: So, do you feel like you get pretty good buy-in from executive leadership, then? You had said that, you know, recommendations go up to the cabinet level. I mean, are they pretty likely to approve things that the group, the data standards group, has recommended?*

> *Participant: Yeah, I think as long as it's well-reasoned. I have to say, they're great about, what is you . . . I mean, what are you trying to do, why are you trying to do it, what's the benefit for the institution, what are the liabilities for the institution? And if you can present that, and they're all well-reasoned, they're like, "OK." . . . We have like 16 people on the data standards committee from across the institution. Everybody in that group buys into some things, and they've communicated back with their areas about it. We've probably picked up most of the rocks and seen what's underneath them. So when we go to the cabinet and try to make a recommendation, we've really, you know, we've really looked under a lot of rocks.*

On the other hand, a number of the participants expressed the opinion that most leaders on their campus didn't have a very profound appreciation for the strategic importance of data quality or understand the kinds of data

problems that existed on their campus. Additionally, some participants voiced frustration that data quality issues did not get the time, attention, or priority they needed:

> *Participant: I asked our interim provost—our provost is away briefly—so, I said to him, "Is it that people don't care? Because we had this one meeting, where everybody agreed we needed to meet, and we haven't met again. What's going on?" And he said, "I don't think it's that people don't care. It's that it doesn't seem "urgent." Something else usually . . . you know, that "urgent versus important" grid. It's very important, but not being seen as urgent."*

Resource Issues

One of the greatest resource issues for teams was that of staffing and the related issue of staff time. With one exception, in which a data governance manager at a large research university oversaw the data quality processes at that institution, almost all participants mentioned team leaders as well as members who had other primary job responsibilities. Whereas there are clear benefits to having data integrity team members with deep understanding of the data needs of one or more specific functional areas, this arrangement can also mean that every person on the data integrity team has other, more-pressing responsibilities, making it difficult for team members to find time to dedicate to data integrity team projects. Several participants mentioned attendance problems at meetings. Workload was also given

as a reason for not having data dictionaries or data warehouses. Of all the participants, only the two research universities had dedicated data governance staff or plans to add any.

Tools for communication between the team and data users were cited as a resource issue. Some participants mentioned that they placed data integrity group minutes or documentation like portable document formats (PDFs) of data dictionaries on an intranet site or used a tool like Moodle. Sometimes users accessed them but reportedly they often did not. Other interviewees said data policy decisions were sent to stakeholders by campus e-mail once, at the time they were implemented, which seemed to be problematic in terms of providing ongoing and readily accessible documentation to users. Only one of the institutions had implemented a "live" interactive metadata management tool. A perhaps related finding was that most participants reported that their team did not have a budget.

Benefits and Challenges

Almost every participant spoke about information-sharing and communication as key benefits of the data integrity team. The data integrity team was cited as a place where stakeholders were identified, impact of data decisions was explored, and users learned how data were created and used in other functional areas. Frequently the data integrity team was where users first became aware of compliance issues, technology changes, or program changes that might impact data collection or reporting needs.

Participant: So do we want to add that field? So we bring it to the table: Who-all does this affect? We think it affects me, institutional research, and the registrar's office. But who-all cares? It turns out financial aid. So it turns out, oh, this affects you, or maybe just confirming our instincts.

Interviewer: So it's a place for finding out who stakeholders are in decisions?

Participant: Yes. Yes, how does this affect, you know, other offices? That's a huge topic of the conversation, and that's been a huge benefit to this face-to-face meeting of folks.

Most of the participants also spoke of increased awareness of data integrity among data users as a benefit of the team, and several indicated that they thought that the team had raised the profile of data quality as a strategic issue on their campus.

In spite of this information-gathering function of the team, communication was also often cited as a challenge to working on data quality issues. Because members had different areas of domain expertise, they could not always easily explain to team members from another unit why a data element was problematic for them or how they knew a specific data point was incorrect. Members frequently used different technical vocabularies or conceptualized data or problems in varying ways. Even defining what constituted a data quality issue could be difficult:

Participant: Sometimes if you ask them, it's "No, we don't have a data quality problem," and then you go back and actually look and, "Well, yeah, actually, you do." "Oh, that's a data quality problem?" And then you talk about that. So it's getting people to kind of understand what their roles are and identify what it is they need to do.

A challenge that institutions seemed to struggle with was maintaining accessible documentation of not just data field standards, but also of procedures. One of the participants told how his school had recently "consolidated all of the handbooks—the students, the employees, the staff handbooks—into one college handbook, and that has reference to just about all the policies and guideline sets." However, this was not typical, as other institutions reported not having adequate documentation of policies, particularly those concerned with identifying and reporting data problems:

Interviewer: So in terms of the kinds of procedures you might have in place, you said you had a manual that has field-level kind of procedures. Are there also procedures for how you would report a problem? Like if you find a field that seems to have some discrepant or inaccurate data in it, and you think that maybe there's some sort of systematic issue, is there a written or formal procedure for how to initiate that?

Participant: Not really. What ends up happening is, either if it's an

immediate problem they go to the IT helpdesk; if they think it's more of a systematic problem, it goes to the data standards group, which meets quarterly. It goes to them to reach a conclusion or a compromise on what should be done.

Training also seemed to be a challenge. None of the respondents reported that their institutions required any form of consistent training on data standards for all new personnel. In general, the standard seemed to be that departments within the college or university were in charge of training their own personnel, because of the difficulty in providing data systems training general enough to meet the differing, technically specific needs of users in diverse functional areas.

The participants mentioned several data areas as particularly problematic for users and teams. Parent names and contact information came up a number of times as an example of data that are of high importance to advancement offices but that are difficult to keep updated and challenging to use. Faculty data frequently were mentioned as a challenge, in part because two offices—human resources and academic affairs—are typically involved in creating and using these data, but also because those offices have different operational and reporting needs. Tracking student hiatuses (leaves) was cited as challenging. Also mentioned as problematic was integrating data from different campuses, or data from online and other special programs. In most of these cases it was clear that

the complexity of the persons and activities represented by these data (online students in the military, faculty on sabbaticals, students whose parents were divorced and/or estranged) and not just technology limitations contributed to the difficulty of creating consistent and usable data.

Other Issues

A general observation was that the larger schools with a business intelligence and analytics orientation seemed to have more-advanced data quality processes. These institutions were more likely to have data dictionaries and data warehouses. Data governance tends to be a core component of a business intelligence and business analytics strategy. One of these respondents was careful to note, however, the integral role that a traditional institutional research orientation played in data quality.

> *Participant: The data needs to be in a way that people are confident in it, and you know how it's defined. . . . And I don't think anybody thinks about that like IR [institutional research] does. You don't have a research function in a typical corporate environment. You have a marketing or planning team or something like that, but not to the level that IR thinks about data governance. So it's been good for them to have us consulting on that. . . . Business intelligence, it won't work without good data. It won't. And you can't have good data coming in out of transactional systems that are not designed for reporting without some very formal sort of guidelines.*

Data system customization was also mentioned by some respondents as a factor in contributing to poor quality data. Although becoming less common as commercial enterprise resource systems replace legacy systems, users frequently have had the option to customize their data system and its fields to institutional needs. Frequently these customizations were poorly designed or documented, or documentation for the change has been lost over the years. In some cases, no current user knew the reasons for or specifics of the customization, which might no longer be necessary. Such customization can make finding and fixing data problems more difficult.

Finally, creative user methods of working around poor-quality data were mentioned as a barrier to improving data processes. Such strategies could include data silos like "shadow" spreadsheets kept by individual users, hasty "cleaning" of bad data to meet contingency needs, and insufficient documentation:

> *Participant: I think it's more that it's not being seen. The ways in which the system is broken are not immediately apparent, and the impact is not apparent. Because people have done an amazing job around here of work-around fixes.*

INTEGRATION OF QUANTITATIVE AND QUALITATIVE RESULTS

The most striking finding from the survey is the fact that only about half of the respondents said that

their institutions had data integrity teams. This could explain why subject recruitment for the qualitative study was somewhat challenging. Both of the studies identified the same group of "usual suspects" among functional areas of team members, with institutional research and IT being the most common areas represented; staff from those areas frequently serve as team leaders. In addition, the studies identified broad representation by other campus departments. In both study phases, development or advancement was the most likely major function not to be included on the team, probably due to the development-specific data systems used at many schools as well as the unique types of data that advancement offices collect and use. Both methodologies found that IT and academic affairs were the most likely executive advocates for data integrity efforts.

Most data integrity team members in both the survey and the interviews reported that their teams were improving data quality on campus. Very few survey respondents from institutions that did not have data integrity teams believed that not having such a team indicated a lack of data quality problems on the campus; rather, it seemed to be related to a lack of resources, including time. This finding accords well with what interviewees said both about the difficulty of getting buy-in to data quality improvement efforts on campus and why their data improvement efforts were not as comprehensive as they would like them to be, and might explain why many institutions

did not have a data dictionary or data warehouse even though they believed that having these resources would be beneficial. Data integrity work is by and large work that team members do in addition to their regular assignments, and respondents often reported difficulty maintaining momentum, particularly when organizational changes or crises demanded team members' attention. Several survey respondents from institutions without a data integrity team remarked in the open-ended comments that they had previously had a team but it could not be sustained. One of the interview participants reported having advised a department that she worked with that it needed to hire someone to attend to data governance issues, and several of the survey respondents stated in their open-ended comments that they believed dedicated staff were needed to oversee data integrity.

Both qualitative and quantitative study participants reported that their teams were participating in many of the same activities: identifying data issues, problems, and stakeholders; determining which offices did have or should have responsibility for which data; and evaluating current data policies and potential compliance or programmatic changes in data needs. Most of the participants in the qualitative study reported that their campus had identified data stewards, although their responsibilities were not always well-documented or official. Data dictionaries, a best practice recommendation in the data standards and data governance literature, were not found at most institutions in the survey sample, and were found in only

half of the institutions in the interview sample. It would be reasonable to suppose that this absence is due to a resource issue. Most survey respondents reported that their teams were not performing data auditing and monitoring activities. Although mentioned by one or two of the interviewees, on the whole they did not discuss auditing when describing core team activities.

A subgroup of the survey respondents whose institutions did not have data integrity teams reported in the open-ended comments that their campuses preferred to deal with data issues in an ad hoc or decentralized fashion. Since many of the data integrity team interviewees cited communication and "getting everyone together at the table" as a benefit of the data integrity teams, this opportunity can be lost when data problems are dealt with in an ad hoc way. It is worth noting that interview respondents saw the value of having smaller groups working on specific problems that mainly impacted their respective units, as long as they reported back to the team. In the same vein, another interesting though divergent finding is that relatively few respondents in the open-ended items called for increased centralization of data integrity efforts, even though bringing diverse functions together was an often-mentioned strength of the team for interviewees.

Both parts of the study found that most teams had authority only at the data field level, and needed to defer to higher-level individuals or groups to make campus-wide policy decisions. Some interview participants believed

the lack of policy-making authority of data integrity teams was a mechanism for keeping leaders in the loop about strategic data issues that might impact the institution as a whole. Although most respondents in both parts of the study believed that their campus leadership and their own supervisor supported data integrity efforts in a general way, they also believed that data quality issues were not very well understood by leaders. One of the ways that leaders support initiatives is by dedicating adequate resources to them, so it says something about executive buy-in that lack of resources was typically given as a reason that data quality efforts did not receive adequate attention.

Finally, both phases of the study identified similar benefits and challenges for data integrity teams. Better communication, awareness of data quality issues, and ability to collaboratively plan for organizational change impacting data systems were among the benefits mentioned by interviewees. Improving communication was also recommended by both survey respondents who were data integrity team members and those who were not as a way to improve data quality on campuses. Training was mentioned as a challenge by both survey and interview participants, as was maintaining readily accessible documentation about policies and procedures.

DISCUSSION

This study has shown that cross-functional data integrity teams on college campuses are identified with

several positive outcomes by team members. Such teams provide a forum for communication about data gaps and problems, foster greater awareness about data systems quality issues, and can facilitate the creation of consistent campus-wide data standards as well as data user policies. However, the study also found that many campuses have not created or do not see the need for such teams. Additionally, teams often lack resources such as time or staffing to implement recommended best practices such as data dictionaries and data auditing.

McLaughlin et al. (2004) have put forth a number of data process models for postsecondary settings that could be applied to these results. For example, the evolution of information management is described as consisting of three stages: (1) decentralized data operations, (2) centralized data administration, and (3) distributed data management. The majority of respondents in this study reported struggling against decentralized data operations, where only internal reliability and immediate operational needs are considered. Data integrity teams were slowly moving campuses toward centralized data administration, focusing on how data will be used for reporting as well as on operational needs, and evaluating data in terms of internal validity as well as reliability. McLaughlin et al. argue that the increasing desire for integrated data by decision-makers necessitates that institutions must move toward distributed models, meaning models that account for data that are spread out over many different software systems. The challenge

of good data increases as data are expected to serve ever-higher-level needs in the organization. One of the interview respondents articulated the importance of distributed systems:

> *Participant: An IR [institutional research] team to be effective really cannot manage it all by themselves. You have to have a distributed model, you know. Or you're going to die. Or you're not going to be successful. So that's what we're working on, is just getting it out into other people's hands. In a centralized data governance process, but distributed down the way that everyone feels confident pulling data, understands how it works.*

Finally, McLaughlin et al. (2004) have posited that there are three ways organizations can respond to data architecture failures: (1) masking or hiding problems, (2) coping and trying to circumvent data shortcomings, or, when these tactics inevitably fail to meet the need for enterprise analytics, (3) correcting deficiencies in the design of data systems and processes. The interview respondent quoted in the results section beautifully illustrated the strategy of coping with her description of "work-around fixes," as well as the role of these kinds of patches in concealing systemic data quality problems. Other interviewees and respondents to open-ended survey items described the creation of departmental or individual data silos as coping strategies. In the case of one research university in the interview sample that underwent reorganization,

the critical need for strategic data was a driver in correcting existing data problems.

It is important to note that this study did not purport to directly measure campus data quality in any way, but only to measure participants' perceptions of data quality. However, for the purpose of this research such indirect measurement was deemed to be adequate because the term "quality data" is defined as data that are adequate to end users' needs. Since the respondents were business data end users, their subjective opinions about data quality were presumably based on professional experience and specialized knowledge or expertise.

A potential limitation of the survey is the small number of survey respondents who were serving on data integrity teams. There might also be a selection bias toward respondents who are very satisfied or very dissatisfied with their data integrity team and data quality on their campuses. Additionally, although AIR draws its membership from many different fields within higher education, AIR's member population might be weighted toward larger and/or more-affluent institutions with the budget resources to pay AIR's conference and membership fees, or toward larger institutional research offices whose staff are more easily able to get away from the office for professional development activities.

At least one group of researchers has identified a lack of connection in the data quality research literature between technological solutions and applied business information

systems contexts (Sadiq, Yeganeh, & Indulska, 2011). In other words, teams need to be aware of and consider using technological solutions to the problems of data quality, whether this means implementing automated data validation and auditing systems, or electronic metadata management tools. If technological tools can help address resources limitations, the development of open-source data quality tools would be a promising applied research area.

To overcome the reluctance of campus leadership to invest in data quality efforts, better methodologies are needed to determine costs to higher education of poor data quality. Better research about the costs of poor data quality might be a necessary tool in moving data integrity front and center with institutional leaders who can set the data governance charge on their campuses. Another possible motivation for paying more attention to data quality could be the recent national press given to several high-profile cases of college and university data problems. The net effect might be to make stakeholders wonder if they are "minding the store" with respect to data quality on their own campuses. Ultimately, the case for data quality for colleges and universities is the business case of more-efficient and more-effective pursuit of educational mission in a time of resource constraints and high expectations.

REFERENCES

Gose, F. J. (1989). Data integrity: Why aren't the data accurate? *AIR Professional File, 33.*

McGilvray, D. (2006). Data governance: A necessity in an integrated information world. *Information management: How your business works.* http://www.information-management.com/issues/20061201/1069951-1.html

McLaughlin, G. R., Howard, R. D., Cunningham, L. B., & Payne, E. (2004). *People, processes and managing data* (2nd ed.). Tallahassee, FL: Association for Institutional Research.

Sadiq, S., Yeganeh, N. K. & Indulska, M. (2011). 20 years of data quality research: Themes, trends, and synergies. In H. T. Sheng & Y. Zhang (Eds.), *Proceedings of the Twenty-Second Australasian Database Conference* (Vol. 115, pp. 153–162). Darlinghurst, Australia: Australian Computer Society.

Yanosky, R. (2009). Institutional data management in higher education. White Paper. EDUCAUSE Center for Applied Research. http://net.educause.edu/ir/library/pdf/ers0908/rs/ers0908w.pdf

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