Data governance and data quality: Is it on your agenda?

ANNE YOUNG AND KEVIN MCCONKEY
The University of Newcastle, Australia

Abstract

Data governance is a relatively new and evolving discipline. It encompasses the people who are responsible for data quality (the stewards); the policies and processes associated with collecting, managing, storing and reporting data; and the information technology systems and support that provide efficient infrastructure. Higher education institutions are paying more attention to data governance as we move into a funding environment that focuses on performance measures, targets and accountability. This article describes the establishment of a university-wide Data Governance Advisory Group (DGAG) at the University of Newcastle and highlights some of the short-term achievements as well as the longer-term goals of the group. The DGAG is responsible for providing advice to senior management on data governance policies, standards and strategic approaches; data quality initiatives; data privacy, compliance and security; data architecture and integration requirements; and data warehouse and business intelligence priorities. Membership of the DGAG includes key stakeholders from planning, quality and reporting; student and academic services; human resources (HR); finance; information technology (IT); research services; library; facilities management; external relations and corporate services. The monthly meetings have provided a forum for discussion of data quality and governance across the whole life cycle from collection to reporting to decision-making for a range of teaching and learning, research, administration and services data collections. Outcomes include the establishment of a register of data collections, a mapping of how data are used in official reports and benchmarking/ranking projects; the development of a common data dictionary, the sharing of good practices and promoting a collaborative culture.

What types of data are collected in your university? Who is responsible for the quality of that data? How and where are the data stored and accessed? How are the data used and by whom? Who makes decisions about the data collections? What do you do when errors are detected? Are the processes for ensuring the security of data documented? Is there a data dictionary (or glossary) to explain the meaning of common terms and where is it stored? If the answer to all, or even some, of these questions is not clear, then your organisation is in need of better data governance. This article will provide an overview of what data governance encompasses, describe the establishment of a university-wide Data Governance Advisory Group (DGAG) at the University of Newcastle and highlight some of the short-term achievements as well as the longer-term goals of the group.

What is Data Governance?

According to the Data Governance Institute (Thomas, n.d., p. 3) data governance is described as 'the exercise of decision-making and authority for data-related matters'. The main functions of data governance are to help establish policies that facilitate the appropriate use of data and to develop strategies for controlling and monitoring that use. More specifically, data governance has also been described as 'a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods' (Data Governance Institute, n.d.a).

What Should Be The Focus of a Data Governance Group?

Areas of focus for data governance can differ between institutions. According to the Data Governance Institute (Thomas, n.d.), areas of focus may include:

- Policy, standards and strategy—a concentration on the high-level data and metadata available within the organisation and how it is being used. Tasks may include reviewing and aligning current policies; creating new policies; clarifying roles, responsibilities and decision-making delegations; and assessing the functioning of business rules.
- Data quality—a review of the quality, completeness, integrity or useability of data. Often such programs evolve from the identification of system-wide issues that require a consultative approach to establish priorities for assessing and monitoring data quality. These initiatives help identify stakeholders and data custodian/stewards, and establish how decision-making, accountabilities and cross-unit communication regarding data quality operate in the organisation.
- Privacy, compliance and security—within the university environment the importance of data privacy, confidentiality, access, permissions, and regulatory and contractual requirements are paramount. Governance processes are required to assess, monitor and control risk, as well as to ensure regulatory and compliance requirements are met.
- Architecture and integration—a data governance group may be formed when a major new data collection system is installed or an existing system upgraded. These projects are challenging and require strong cross-unit cooperation. Typical tasks include building a common data dictionary, agreeing on architectural policies within an integrated framework, updating data and metadata and communicating these changes across the organisation.
- Data warehouses and business intelligence—these projects may start in a focused way with the introduction of a data warehouse, data mart or reporting tool but often grow to encompass other areas of data governance as issues arise. In assessing the business rules required to build a warehouse, the quality and integrity of data need to be assessed and the existing structure of decision-making and accountabilities assessed. The introduction of a new functionality can be seen as a great opportunity to review data definitions and rules for data usage and to establish who the data custodians are and where the gaps are in current processes.
- Management alignment—the move to establish a formal data governance committee often comes about when management actions about data stall because of uncertainty about the consequences of certain decisions. What downstream systems might be affected? Which reports use these data and will they still run if the planned changes are made? In such situations, managers need to come together to discuss and assess interdependencies and make strategic decisions. A bonus is that communication is strengthened and the need for

enhanced data governance is evident to the main stakeholders, rather than being dictated from senior management as an obscure requirement.

Why Do We Need Data Governance?

Organisations may differ in the approach they take to data governance, depending on which of the areas of focus is more important to them at the time, and there may be multiple areas of focus! Organisations need to be able to make decisions about how to collect and manage data, maximise the value of the data for decision-making, in an efficient way within a complex legal, regulatory and compliance environment.

Regardless of whether data are used widely across the organisation or in 'silo' areas, it needs to be managed in a responsible way by staff who are designated as accountable for that data. One of the principles of data governance, however, is that individuals and areas do not 'own' data—it is an enterprise asset that may flow through many processes, be stored/moved/transformed by information technology (IT) systems and end up in an array of reports. Hence, a 'federated accountability' approach to data stewardship is recommended, whereby segments of the data lineage (from data creation/acquisition through to reporting) are documented and accountability for each segment is assigned. This approach is challenging, time consuming and complex but it is an approach that works (Data Governance Institute, n.d.b).

Who Should Be Involved in a Data Governance Group?

The concerns to be addressed in an organisation will often dictate who should be involved, and to what extent, in the data governance group. It may include the 'data stewards' who are responsible for the data as well as the 'data stakeholders', who are interested in how data are collected, coded, processed, manipulated, stored, made available, reported or archived. One benefit of meeting as a data governance group is that decisions can be made according to an agreed-upon process, with input from all interested parties. For the group to achieve its goals, it must receive appropriate levels of leadership support.

Where the data governance function sits, within the organisational chart, is less important than having the right people involved and committed to the purpose of the group. The organisation needs to be ready for a more formal approach to data governance and this often evolves as issues arise and networks of colleagues develop across the different areas of the organisation. These informal networks often include data stewards (those responsible and accountable for the data) as well as staff from areas that store, extract, merge, use and report the data. Providing a forum where data-related issues can be discussed can lead to a wider acknowledgment of the importance of data quality and of improving data processes, involving the relevant stakeholders and promoting an institution-wide perspective to decision-making. In order to maintain momentum, participants need to see 'what's in it for them', that they are not wasting their time and that the group is making a difference! Publicising the group's accomplishments also helps to promote the work of the group and the value of its contribution to continuous improvements within the organisation.

How Do You Establish a Framework for Data Governance?

In order for a data governance framework to be understood, accepted and embraced in an organisation, it must be tailored to the culture and structure of that organisation and evolve in a consultative way. The Data Governance Institute has a website with resources to support

data governance projects including whitepapers, case studies, best practices and non-technical briefings on data-related issues, a framework to assist with thinking and communicating the concept, along with a 'how to' guide and who-what-when-where-why-how information about data governance (Data Governance Institute, n.d.a). Models of data governance include

- Top-down—where decisions are made by the senior executive and are communicated down.
- Bottom-up—where data-related decisions are made by individuals or groups at the local level and are communicated up. These tend to be process issues rather than policy issues.
- Centre-out—where decisions are made by one or more senior staff in a central unit that is most familiar with the issue and who make a recommendation about what is best for the organisation.
- Silo-in—where representatives from multiple groups collectively agree on a course of action that weighs up the needs of each group and the needs of the organisation as a whole, where the group can propose recommendations or has been given delegated authority to make decisions.

In any of the models described above, establishing multiple communication channels and educating stakeholders about why decisions have been made is important to achieve buy-in and compliance.

Why Should Data Governance Be Important to Institutional Researchers?

According to Delaney (2009, p. 29), 'institutional researchers can best serve higher education in the twenty-first century by enhancing their current roles and adopting new roles to exert greater influence on decision making'. New roles include increasing their knowledge of an institution's internal culture and external environment and serving as 'knowledge industry analysts, knowledge generators, and knowledge brokers' (p. 39). The mandate to ensure the integrity of the institution's data and information to contribute to that knowledge creation is clear.

Although the need for data governance in higher education institutions has been considered from an IT perspective (Cheong & Chang, 2007; Yanosky, 2009), it is rare to hear it discussed in institutional research forums. An exception is the recent presentation by Yonai and Anderson at the 2011 AIR Forum in Canada (Yonai & Anderson, 2011) where they described the collaborative approach taken at Syracuse University by the Office of Institutional Research and Assessment and the Office of IT and Services to establish a data governance program. Some of their issues that prompted the program included perceived problems with access to data, an inefficient process for requesting information, lack of data quality audits, duplication of data, no comprehensive list of data quality initiatives, insufficient training and education about data and the use of shadow systems. Challenges to their project included a history of lack of trust and support between units, differences in terminology between units and competing goals rather than complementary ones. The solution was to undertake a needs assessment process and prioritise outcomes that would benefit most users. Their initial findings identified a need to better align business processes, including coordinating system changes with other departments; eliminating shadow data collection systems; reviewing privacy, security and compliance; improving data access and acquisition; improving access to historical data and better communicating how to obtain data.

Progress towards improving data governance can have a positive impact on the work of institutional researchers through the full cycle of activities, from design and collection of

valid and reliable data through to analysis and reporting of results. The typical goals of a data governance group (Data Governance Institute, n.d.c) are to:

- enable better decision-making
- reduce operational friction
- protect the needs of data stakeholders
- train management and staff to adopt common approaches to data issues
- build standard, repeatable processes
- reduce costs and increase effectiveness through coordination of efforts
- ensure transparency of processes.

Specific activities may include addressing infrastructure issues or validating some of the core business data definitions, business rules and processes. All these goals are consistent with the goals of institutional research teams.

Why Are Consistent Definitions Important?

Although some data collections may exist in 'silo' areas of the organisation, high-level data may be used by groups of analysts across the institution and interpreted by a range of stakeholders. Without consistent and accessible definitions, a data element may mean one thing when it is collected, but an analyst may assume a different meaning. Sometimes these differences are only detected at the reporting stage when results provided to senior management are questioned and the issue then needs to be resolved. Worse still are situations where data are used incorrectly but the error is not easily detected and incorrect data are reported and used in decision-making. Examples in higher education where definitions may differ include a student's address (term, home, postal), number of enrolments (headcount, program enrolments, point in time, year to date). Data definitions should describe a single concept, be mapped to a particular data element, be precise in every situation in which the user would employ the term and be easily understood by the reader—not a simple task!

What can go wrong when establishing a Data Governance group?

Some authors have listed the 'worst practices' in data governance (Sherman, 2011). One pitfall can be putting too much focus on the data alone, rather than how to ensure the integrity of the data, processes, documentation and dissemination so as to maximise its use to inform decision-making. Another pitfall is trying to tackle too many issues at the same time or starting the ball rolling with a complex organisation-wide project. A staged approach with incremental gains will be more sustainable and the lessons learned can be incorporated into the next phases, rather than risking having a bad experience at the outset.

Failure to gain support to implement recommendations is a real risk in data governance projects. Assigning clear accountabilities, with agreed deliverables and timelines, is one way to control that risk and by providing resources where needed to support the activities. Increasing the workload of busy staff, without reprioritising existing tasks or providing more resources, is an unrealistic approach to implementing change. Another myth is that IT systems alone will fix data-related problems. The real asset in an organisation is the people and their interactions and their collegial approach to finding the best solution for the organisation as a whole (Harris, 2011).

Concentrating on the high-level transactional systems within the organisation is important, but ignoring the other data collections that exist is unwise. Key information may

often be collected and held in smaller systems scattered throughout the organisation. A task for the data governance group early on is to document and better understand these systems and consider how integrating that information may provide benefits to the organisation.

How Does Data Governance Promote a Culture of Continuous Improvement?

Regular monitoring and review of activities are key components of continuous improvement. The typical quality enhancement cycle begins with development of a strategic plan that defines an organisation's vision, purpose and goals. This, in turn, determines the nature of actions and information required, followed by implementation of a monitoring system, analysis and reporting of results, and recommendations for improvements that align with the overarching strategy. The same steps should be followed in a data governance project. After identifying opportunities for improvement, it is important to ensure that findings are acted upon and those actions are evaluated. Effective data governance requires a 'significant and sustained change management effort' (Harris, 2011, p. 8). Consistent with good strategic planning, the data governance group and other stakeholders should know what success looks like and how it will be measured. Key questions for the data governance group are:

- What will the area/s of focus be?
- Why is this a priority?
- What outcomes are we seeking?
- How will we achieve them?
- How will we know if we have been successful?

Case Study: Data Governance at the University of Newcastle

The Data Governance Advisory Group (DGAG) at the University of Newcastle was established in 2010 in response to a growing interest in having a cross-institutional forum for the exchange of ideas on data-related issues. The implementation of a data warehouse drawing data from several systems across the university, and the subsequent reporting on that data, highlighted areas for redress. Some examples include changes in names or coding of data elements that were not communicated to all users, changes in input systems whereby some data items were no longer collected, spikes in missing data for some elements and system upgrades that had unintended downstream impacts on the warehouse. Preliminary discussions took place during meetings of the IT Governance Group (ITGC) and a working party was formed to draft the terms of reference of a DGAG.

Consistent with contemporary literature, as described in this article, the Data Governance Advisory Group was deemed to be responsible for providing advice to senior management on data governance policies, standards and strategic approaches; data quality initiatives; data privacy, compliance and security; data architecture and integration requirements; and data warehouse and business intelligence priorities. Some of the key questions then to consider were: who should be involved, what type of commitment would be required and how should the meetings be run?

It was agreed that the Data Governance Advisory Group would have representatives (Directors and IT system support staff) from Planning, Quality and Reporting; Student and Academic Services; HR; Finance; IT; Research Services; Library; Facilities Management; External Relations and Corporate Services—with the group being chaired by the Director, Planning, Quality and Reporting. Monthly meetings were scheduled for the first 12 months

with the intention of reviewing the membership and timing of meetings as the group's work evolved and establishing working groups as necessary. The meeting agendas were structured around the terms of reference, with a call for additional agenda items prior to each meeting. The first few months for the group were primarily a period of induction to the concept of data governance and discussing the model that might work best at the University of Newcastle, and building a shared awareness of the data-related issues to be addressed.

One of the early outcomes of the group was the establishment of a register of data collections, using a template to collect and collate the information. The template gathered information from the data stewards of each collection about:

- the main fields of data collected
- the source of the data
- where the data are stored
- what types of documentation are available to describe the data
- reports produced from the data (by whom and where the reports are stored)
- whether the data are sent outside the university (if so to where, and when)
- whether the data are used in benchmarking projects or activities.

Completed templates were received from all many areas and included a diverse set of data collections ranging from official collections (Higher Education Student Data, the HERDC publications) to finance and HR collections, library repositories, student and staff survey data, complaints data, the contact relationship management system (for alumni and donors etc.) and more.

The templates have provided an opportunity to begin the process of understanding the data held within the university and how it is used internally and externally. We have summarised how data elements are used in official reports and benchmarking/ranking projects so that the data stewards are more aware of the downstream uses of those data and the consequence of poor data quality for performance measures that depend on those data items. Some examples include the importance of validating students address data (currently used to define low socioeconomic status) and academic staff nationality (currently used as a measure of international diversity in the DEEWR performance portfolio as well as several international ranking systems and benchmarking projects).

Discussions have commenced about developing a common data dictionary and the meetings have provided a forum for the sharing of good practices and promoting a collaborative culture. Representatives from various areas have presented an overview of their data collections and data-related issues at the meetings, such as promoting understanding and compliance with our data security policies, and have benefited from the collegial discussions about how the group can work together to make improvements.

Several data quality projects have begun, such as assigning more detailed coding for majors and specialisations in some generalist degrees. Detailed field of education codes are now available for students who complete, meaning that although the time series trends for some fields of education may change, we can explain those changes and demonstrate the value of the more specific coding.

As the role of the data governance group has been more widely disseminated, other areas that are responsible for data collection systems, such as marketing and recruitment, have approached the group to become involved. Although the focus of the early meetings has

been on understanding our larger data collections and addressing known data-related quality issues (and we had a long list!), the group hopes to integrate more data into the warehouse to help monitor performance against the Strategic Plan, particularly in areas where activities and outputs are less easily measured, such as community engagement, partnerships and collaborations.

Although the group began by focusing on addressing data quality issues, we are now moving to a more preventative approach to minimise the occurrence of data-related problems by reducing ambiguity, establishing clearer accountabilities, developing and documenting processes, disseminating and communicating across units and providing more support to groups using the data.

The Data Governance Advisory Group has accepted the challenges as described by the Data Governance Institute (Data Governance Institute, n.d.d), to:

- contribute to standardised data definitions
- prioritise the need for policies and help draft those policies
- identify and reconcile gaps in processes and documentation
- assign accountabilities in areas where that has been unclear
- review and update accountabilities to reflect current structure and responsibilities
- report progress on data quality improvement initiatives
- report on compliance with policies
- track decision-making for data-related processes
- establish data quality rules, particularly with respect to missing data
- better understand those data items that are reported externally for regulatory or benchmarking purposes
- discuss and improve how data are being reported internally
- actively monitor data quality
- discuss collection, storage and access to sensitive data
- assess risks and controls around data security
- discuss system integration requirements
- bring cross-unit attention to integration challenges
- establish rules for data usage and data definitions
- identify stakeholders, clarify accountabilities and confirm decision rights.

Conclusion

Establishing the Data Governance Advisory Group at the University of Newcastle has provided a forum for cross-unit, collegial conversations around data-related issues covering the broad spectrum from data collection to reporting to decision-making for a range of teaching and learning, research, administration and services data collections. As we move into a funding environment that focuses on performance measures, targets and accountability, a sound data governance framework will be crucial for the institutional researcher to contribute to and influence decision-making.

References

Cheong, L.K., & Chang, V. (2007, December). *The need for data governance: A case study.* Paper 100 presented at the meeting of Australasian Conference on Information Systems, Toowoomba, Australia. Retrieved from http://aisel.aisnet.org/acis2007/100/

Data Governance Institute. (n.d.a). *Data governance: The basic information.* Retrieved from http://www.datagovernance.com/adg_data_governance_basics.html

Data Governance Institute. (n.d.b). *Assigning data ownership.* Retrieved from http://www.datagovernance.com/gbg_assigning_data_ownership.html

Data Governance Institute. (n.d.c). *Goals and principles for data governance.* Retrieved from http://www.datagovernance.com/adg_data_governance_goals.html

Data Governance Institute. (n.d.d) .*Data governance with a focus on policy, standards, strategy.* Retrieved from http://www.datagovernance.com/fc_policy_standards_strategy.html

Delaney, A.M. (2009). Institutional researchers' expanding roles: Policy, planning, program evaluation, assessment, and new research methodologies. *New Directions for Institutional Research, Fall*(143), 29–41. doi: 10.1002/ir.303

Harris, J. (2011). The collaborative culture of data governance: Data quality success means taking responsibility and playing for the same team, enterprise wide. *Information Management, 21*(1), 8.

Sherman, R. (2011). *A must to avoid: Worst practices in enterprise data governance.* Retrieved from http://searchdatamanagement.techtarget.com/feature/A-must-to-avoid-Worst-practices-in-enterprise-data-governance

Thomas, G. (n.d.). *The DGI data governance framework*. Retrieved from http://www.datagovernance.com/dgi_framework.pdf

Yanosky, R. (2009). Key findings: Institutional data management in higher education. *Educause.* Retrieved from http://net.educause.edu/ir/library/pdf/EKF/EKF0908.pdf

Yonai, B., & Anderson, E. (2011, May). *Data governance at large research university: A collaboration between the Office of Institutional Research and Information Technology.* Paper presented at the meeting of the Association of Institutional Research, Toronto.