Joshua Mandell

RSAN 150

Assignment 2: Data Quality Whitepaper

Internet Training Solutions - Data Quality

To whom it may concern,

As the CFO of Internet Training Solutions, you are no doubt aware of my recent project to interview major stakeholders in the company about our current data quality issues. In addition to you and the accounting officer, which is to say the data producers in our firm, I interviewed the IT manager and one data user. Based on these interviews, I have some recommendations to make for our data quality program. I base these conclusions and recommendations on the current state of our firm and on the general principles of data quality to which we must adhere. Our interviews have demonstrated that there are needs in data quality that are not being met. As the IT manager explained to me, there are individuals within ITS who know the system for data inputting and are performing quality control on the data, but this is not sufficient. Data quality has to be able to be judged from the perspective of the end user, not just the person inputting the data – and must possess the following attributes: availability, usability, reliability, relevance, and presentation quality (Cai and Zhu). The current system has aspects which are currently working but which are more prone to eventual failure due to the lack of well-organized data governance.

Based on the information gathered during the interviews – and the fact that these interviews were needed in the first place – it appears that there is not sufficient direct communication between departments. The CFO stated several data *needs* – to monitor cash balances, to meet forecasts and give early warning signals if there are parts of the business underperforming. However, these goals cannot simply be stated – they have to be directed, and that direction has to come from somewhere within the organization. There is some disagreement about where this direction should come from. A general user suggested a top-down approach – that is, management has a commitment to a particular way of doing things, and therefore should be responsible for ensuring that particular way of doing things is implemented. However, ITS is resistant to the presence of a

dedicated position of Data Quality Manager, favoring a more decentralized approach. The decentralized approach may be more manageable when it comes to the company's available resources. However, over time the system has grown and this purely decentralized approach seems no longer tenable without new rules put in places, rules that would be best created and managed by some kind of governance entity. As the IT manager said, "you can't always put in 100% controls…" what the account relationship manager is doing is manually overseeing the quality of information – doing the job of data quality but without the necessary governance structure to *maintain* an ongoing data quality program.

One of the problems with the current situation is that the different parties interviewed all have different sets of priorities and different notions about how our data quality system should work, how quality should be ensured, and what should be done about making sure that our process works correctly. The general user desires a top-down approach while the IT manager wants everyone to be their own data quality manager. The CFO has a set of priorities but does not have a strong attachment to whether those priorities are met via a dedicated data quality manager or just someone doing the job of a data quality manager ("I don't think it makes any difference"). Unfortunately, in this situation, it likely *will* make a difference. Because different parties involved do not agree, there's a need for some kind of outside dedicated entity to ensure data quality. There is no properly defined Information Governance Framework that can bring these disparate wants together. It appears from the interviews that currently, regardless of how the system is supposed to work in theory, technical and business parts of the company are each essentially marching to their own drums. A gap between IT and business is a common problem of data quality in organizations, and it is bridged by an Information Governance Framework (Bhansali, pg. 224).

An Information Governance Framework consists of the elements of policies, people, and technologies. Enhancing each of these elements will improve the data quality situation at ITS. On the dimension of "policies" – a top-down approach when it comes to data governance will enable policies to be consistently applied. Currently, there are data entry staff who "know the system" as the IT manager has assured us. At first glance it may seem like this fulfills the "people" dimension of an Information Governance Framework, but the problem is that this does not create consistency in results. Each user has their own perspective, and the assumption that the current system will always work just because it currently does is built on the assumption that people will intuitively share their

perspectives with one another without any external means of ensuring consistency. It is very important to have people who know the systems, particularly when it comes to outside data capture such as that described here. It is good that there is an account relationship manager who ensures that information is input correctly – this ensures data quality in the sense of accuracy and standardization. But there is far more to data quality, and bringing in a data quality expert to the organization would probably help greatly to make sure data is not only accurate and standardized, but consistently useful across the organization – accessible, accountable, communicable, compliant, secure and private, and fit for use (Bhansali, pg. 229). Consider also the technological aspect of the Information Governance Framework. The CFO wants, in particular, monitors and early warning signals for parts of the business that may be underperforming. This can be accomplished in part by metrics – a technological aspect of an information governance framework. I think it's necessary to build up such a framework – and to have a data governance body with a data quality manager of some kind – because there are clear ways to improve all three of these major aspects.

When building a data quality framework and when determining the organizational structure that will serve the best data quality, it's important to keep in mind the basic principles of data quality and how these apply to our organization. These principles are as follows:

Data are a product, with customers, to whom they have both cost and value (Cai and Zhu). We need to consider what value we are delivering to our users and clients, and what our data is costing them (not just in terms of literal price, but also the results of their own efforts. Our clients want to market their courses to students, and if we give bad data, it is not just wasted costs and time that our clients will face – they will be making incorrect marketing decisions. As it stands, because of the way we acquire data (individual staff are getting data from the internet and each person involved is "their own data quality manager"), even if we assume that *currently* data is performing well on all possible metrics of quality, this could quickly change due to a change in personnel or problems with data sources. In light of this, it's important to know that a more dedicated position of data quality management – and a more organized data governance body – will go a long way to ensuring we are giving our clients the best data they can get.

As a product, data have quality, resulting from the process by which data are generated. As stated above, our current process might result in lower data quality without the help of a data governance body with a designated data quality overseer. Central management of data quality serves multiple functions: not only assessing and monitoring the quality of the data, but handling quality issues that individual data entry staff and IT staff may not be able to handle sufficiently without disrupting their existing work. Having a dedicated individual or team for data quality assurance means that there is a consistent voice pushing for better data quality, separate from the process of the actual collection and processing of data. Finally (although sequentially first in the process of setting up a data quality management framework) the data quality manager can define standards (Sebastian-Coleman). "Knowing the system" is not enough – overseeing a mechanical system to check the actual validity of values entered is not enough – standards must adequately set expectations for the fitness of use to our clients, not just the validity of data.

It is understandable that stakeholders including the IT manager and the CFO may not think it is necessary to have a singular office of data quality imposing a top-down standard for data quality. It might interfere in the way things are currently done, and might make it more difficult to get existing work done. However, these fears can be assuaged if we are deliberate about how we implement new data quality standards. Firstly, it is clear that there is already a de facto set of data quality practices – there is just no central body administering them. While I recommend creating such a body, that does not mean the existing standards have to go out the window. Existing standards for data validity will continue to work well, *and* it's important to keep relying upon the data entry staff's knowledge of the system of inputs and of their data sources (Sebastian-Coleman). This will provide for the general user's desire for a top-down management, and it will place more of the responsibility on management instead of on individual data entry or IT staff people to figure out standards. Although front end employees play a very important role in ensuring data quality, and could be considered "in charge" in that sense, they should not be saddled with the responsibility of creating standards that they do not have the authority to enforce – that responsibility falls on management.

It is clear that there is some resistance to this, likely because of the cost to the firm. Of course there will be the monetary cost of paying a data quality manager, and the organizational costs of the time spent getting the

new standards set up and implemented. It is understandable that there will be some resistance if the current way of doing things seem to be working fine for now. But part of the point of proper data governance is that – even if the current way of doing things works fine for now – it may not stay that way forever. There needs to be a consistent set of standards applied, ways to formally work toward the improvement of data quality, ways to manage identified issues with data quality. There will be plenty of upfront costs because of the factors just discussed, but in the long run, the benefits will be far greater. Formal processes for describing data quality are popular because they work – there would be no need for process improvement tools (Burns) if this were not the case. In the long run, upgrading our organization's data quality and data governance structures now will lead to better data quality in the future and will allow us to deliver a better product to our clients, while resolving any major problems that arise which may be too great for individual front-end users to solve without consistent central help and guidance.

— Clear explanation of some key issues
— I would have liked to see something more on data quality principles
— Strong analysis & specific recommendations
— This was the best part of the paper
— Not well organized & minor format issues (e.g. 2nd person voice)

Works Cited

Bhansali, Neera. *Data Governance*. *learning.oreilly.com*, https://learning.oreilly.com/library/view/data-governance/9781439879139/. Accessed 13 July 2022.

Burns, Robert. *Appendix E Quality Improvement and Data Quality*. p. 14.

Cai, Li, and Yangyong Zhu. "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era." *Data Science Journal*, vol. 14, no. 0, May 2015, p. 2. *DOI.org (Crossref)*, https://doi.org/10.5334/dsj-2015-002.

Sebastian-Coleman, Laura. *Meeting the Challenges of Data Quality Management*. Elsevier Science & Technology, 2022. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/brandeis-ebooks/detail.action?docID=6871491.