

Data Governance

Loic Niragire

TIM-8130 v4: Data Curation

Dr. Ulrich Vouama

February 20, 2022

Article 1: Trusted Decision-Making: Data Governance for Creating Trust in Data Science Decision Outcomes

In this article, the authors sought to understand the role of data governance as a boundary condition for creating trust in data science decision outcomes by analyzing two data science case studies in asset management. Whereby boundary conditions refer to socio-technical constraints that need to be satisfied to trust data science outcomes. They concluded that an organization's data governance capability plays a vital role in accepting data science decision outcomes. Furthermore, they emphasized establishing data governance capability as a prerequisite for elaborate data science development.

The first case study analyzed in this paper was a data science project utilizing predictive maintenance to improve the efficiency of road maintenance. The second case study identified fraudulent electricity use within medium and low-tension electrical grids without infringing privacy regulations. The authors analyzed a total of ten datasets in the second case study and more than forty datasets on the first case study. The following sections dive into data governance practices described in both case studies.

Data Governance

In both case studies, the authors used multiple data sources and adopted a multi-method research design by relying on individual interviews, relevant market research, policy documents, and websites. They compared aspects of data governance found in internal documentation against those exposed in the interviews and used a combination of within-case and cross-case analysis to analyze their results.

The organization implemented policies for data management, data exchange with third parties, data processing methods, and individual training. Moreover, policies for assessing and

publishing the monetary cost of data assets were introduced to raise awareness of the importance of data quality management. Also, the organization appointed a data owner to oversee the translation of their policies into a data strategy in which opportunities and risks can be managed.

The organization implemented a data quality framework to improve its data quality control, which follows an eight-step process. Furthermore, they developed an automatic auditing tool in combination with a manual auditing tool to monitor the quality of the data. Finally, the organization adopted a centralized data quality measuring to ensure a standardized working method. For security and compliance, privacy officers are assigned while the CIO has the final responsibility to ensure data security.

Both case studies utilize data pipelines that feed various data sources into a data lake. In the second case study, data originating from smart grid terminals was complicated to understand as it lacked metadata, and subject matter experts with in-depth knowledge about the data were hard to find.

Article 2: Data Work in Context: Value, Risks, and Governance

The authors in this article elaborate on the processes involved in turning data into action and associated data governance to mitigate risks. They argue that advancements in pervasive infrastructure of networks, smartphones, sensors, and other devices have given rise to new data-driven interactions between people and organizations. Consequently, a more connected society has emerged whereby organizations and people increasingly rely on data to inform their decisions and routinely share data among entities. Therefore, turning data into action creates various risks at the personal and organizational level, requiring a constant trade-off between value and risks.

The authors discuss data governance in healthcare and business settings following a value-added model, which begins with data and proceeds from information and knowledge to action. This model distinguishes data from the processes that can be applied, such as organizing, analyzing, judgmental, and decision. These value-adding processes are called data work. The healthcare setting case illustrates an example where the starting point for analysis can be problem-oriented, such as resolving clinical problems. Meanwhile, the business case gives an example where the starting point for analysis is to discover patterns in data rather than resolve specific problems.

Data Governance

The first data governance practice discussed in the healthcare setting is metadata management, namely organizing collections of several kinds of data about patients, processes, and financial transactions. This collection involves a range of data sources, including remote sensors, meters, fingerprints, genetics, handwriting, x-rays, and more. Consequently, metadata collection and management in this setting highlights data security management concerns due

to personally identifying information. The second data governance practice discussed is data quality management through data ownership. Two arguments are made in this regard, the public should be allowed free access to information, and creators of original content should be well compensated. The third data governance practice discussed is document management, namely regulating data exchange involving key documents such as terms of service and privacy policy.

Article 3: Data Governance 101: IR's Critical Role in Data Governance

The authors in this article discuss the challenges of effectively managing various information data streams to maximize student and institutional outcomes throughout colleges and universities. Additionally, they offer helpful tips and important considerations for building and implementing an effective data governance program.

Data Governance

Data governance aspects discussed include data modeling to connect tables and disparate data systems, documentation of variables and their meaning in different contexts, and use of data warehouse. Moreover, data quality issues are discussed since institutional data are increasingly disconnected and undocumented. Additionally, senior leadership's primary concern should be data security and access control by establishing guiding principles for data stewards to use as a basis for their decisions. The authors also discuss data management through web-based tools that simplify the collection of documentation and security protocols. In addition, web-based tools allow subject matter experts to contribute their expertise in defining data resources.

Data Governance Comparison

Table 1 Data governance comparison

	Article 1	Article 2	Article 3
<i>Metadata collection</i>	- Interviews	- Sensory data	- Web-based tools
<i>Data quality management</i>	- Implementation of a data quality framework - Automatic auditing tool - Centralized data quality measure	- Establishment of data ownership - Value-added processes	- Involvement of subject matter experts
<i>Data architecture management</i>	- Data lakes - Data pipelines	- OpenTable - Xignite	- Data warehouse

Data Governance Synthesis

Proper and complete data governance within an organization begins with acknowledging data as an institutional asset and not a departmental resource (Association for Institutional Research, 2022). It is critical to assert that the process remains a neutral facilitator with the sole purpose of ensuring data integrity and availability. Data sharing expectations are driven by senior leadership, who also provide continuous guidance and support needed to embrace the new expectations. Moreover, senior leadership must establish an acceptable risk threshold for the required data resources investment (Henderson & Earley, 2017). Finally, all stakeholders must be provided an opportunity to participate in the creation of data and must be empowered to address data quality, integrity, and availability. Most importantly, data security and access policies must be outlined from the start (Zeid, 2014).

References

- Association for Institutional Research. (2022, February). *Analytics can save higher education. Really*. Retrieved from EDUCAUSE: <https://changewithanalytics.com>
- Henderson, D., & Earley, S. (2017). *DAMA guide to the data management body of knowledge (2nd ed)*. NJ: Basking Ridge.
- Zeid, A. (2014). *Business transformation: A roadmap for maximizing organizational insights*. NJ: Wiley Publishing.