Project One: Data Quality Plan

Creston Getz

Southern New Hampshire University

DAT 325: Data Validation: Quality and Cleaning

Ray Rabago

January 18, 2026

Purpose Statement

High-quality data is the cornerstone of modern businesses. Without it, organizations lose the competitive advantages and insights that data can provide. Data quality is critical because bad data leads to bad decisions. It is imperative that data from the new firm is carefully studied and verified before it is used. High-quality data generates more value than the resources required to validate the new firm's data. Good data improves operations by reducing the time spent fixing errors, which lowers overall workload. It also provides a competitive advantage through better forecasting of demand, pricing, and strategy (uCertify, 2026; Omol et al., 2024). High-quality data is a significant long-term business asset and supports the firm's ability to scale successfully during and after the integration of the new firm.

Organizational Goals

Obtaining high-quality data requires identifying business goals. Without understanding what an organization's goals are, achieving high-quality data is much harder. In our firm, improving decision-making, ensuring regulatory compliance, and increasing operational efficiency are all key goals. These are all areas that good data can support. High-quality data improves decision-making by replacing guesswork with facts we can trust. It allows us to identify trends, patterns, and predict outcomes (uCertify, 2026; Omol et al., 2024). For example, if a new firm has a product with many duplicate entries in its sales data, importing the data without first verifying it could lead to incorrectly reporting the number of sales for that item.

Our firm is also required to follow certain regulatory compliance standards. For instance, the Sarbanes-Oxley Act (SOX) mandates that public companies maintain strict financial data accuracy. Failure to comply, can result in the CRO or CFO facing criminal charges (Krantz & Jonker, n.d.). High-quality data improves operational efficiency in many ways. As mentioned

earlier, it reduces the time spent fixing errors, which lowers repeated work. Another operational benefit of good data is improved supply chain management and forecasting. If we can predict demand, we can be better prepared, which reduces stress and improves customer satisfaction. If the data from the new firm were already high quality, then the integration process will be very smooth.

Data Quality Characteristics and Procedures

High-quality data is defined by six characteristics: completeness, validity, uniqueness, consistency, timeliness, and accuracy. Obtaining perfect data quality is not possible; however, we can use our business goals to identify which characteristics are most important to us (Roundy, 2024). If we plan to implement AI tools in the firm in the future, high-quality data becomes even more critical, as bad data leads to bad models (Edwards, 2024). The two most important characteristics we need to address for the new firm's data are accuracy and completeness. The other four are also important, but not as critical as these two.

Accuracy refers to the degree to which information is correct. Are the entries filled with incorrect information? For example, if someone accidentally mixes up two of the new firm's sales objects during data entry, and the difference is significant, it could lead to incorrect insights. To help ensure accuracy, we can cross-check the data with other datasets, inspect the data sources, and set up automated validation processes (Ehrlinger & Wöß, 2022; Fortra, 2025).

Completeness refers to whether all required information is present and whether we have all relevant data from the new firm. We must ensure that all datasets include the necessary fields and that no critical data is missing. Without complete data, cross-checking becomes much harder, and we may miss important insights about the new firm. We need to define which data is critical and ensure it is fully complete. This can be achieved by working with the new firm's

management to confirm that we receive all the data needed for our key performance indicators (Omol et al., 2024).

Security and Personnel Responsibility

Security policies can impact data quality in many different ways. Data risk management helps ensure that data is safe and reliable for use. Data risk management helps identify, assess, and mitigate threats, both intentional and unintentional. When migrating data from the new firm, we need a plan to identify potential risks. Such as, if there are audit gaps or the firm did not follow the principal of least privilege for users of their data. As mentioned before, our firm must follow certain regulatory rules, such as SOX. Identifying the risks the new data poses in relation to complying with these rules is important. Leadership should be tracking items such as the new firm's data sources, metadata, and any security risks associated with how the data was previously handled (Fortra, 2025; Ladley, 2012).

For proper data governance, it is important that we appoint, at a minimum, a data owner. Data stewards will help the data owner fill critical gaps in the new firm's data and maintain high quality. Data custodians will be responsible for the technical implementation and correction of data issues. These roles can be adjusted to fit our needs; for example, we may have multiple data stewards if certain members of our firm have domain knowledge in specific parts of the new firm's data. A data owner plays the most important role because they are accountable for the protection, use, and quality of the data. Having a single person responsible for the data ensures there is a clear chain of command and that someone is actively monitoring it. The data owner for the new firm's data should be a member of senior management at the firm (Herzberg, 2021).

References

Edwards, C. (2024). *Data Quality May Be All You Need. Communications of the ACM, 67(7), 8–10*. https://doi.org/10.1145/3647631

Ehrlinger, L., & Wöß, W. (2022). *A Survey of Data Quality Measurement and Monitoring Tools. Frontiers in Big Data, 5*. https://doi.org/10.3389/fdata.2022.850611

Fortra. (2025, June 5). What Is Data Risk Assessment? Why & How to Do an Assessment. https://www.fortra.com/blog/what-data-risk-assessment-why-how-do-assessment

Herzberg, B. (2021, November 11). *The Datamasters: Data Owners vs. Data Stewards vs. Data Custodians*. Satori. https://blog.satoricyber.com/the-datamasters-data-owners-vs-data-stewards-vs-data-custodians/

Krantz, T., & Jonker, A. (n.d.). *SOX Compliance*. IBM. https://www.ibm.com/think/topics/sox-compliance

Ladley, J. (2012). *Data Governance : How to Design, Deploy and Sustain an Effective Data Governance Program*. Morgan Kaufmann.

Omol, E., Mburu, L., & Abuonji, P. (2024). *Unlocking Digital Transformation: The Pivotal Role of Data Analytics and Business Intelligence Strategies. International Journal of Knowledge Content Development & Technology, 14(3), 77–91*. https://doi.org/10.5865/IJKCT.2024.14.3.077

Roundy, J. (2024, September 9). *6 dimensions of data quality boost data performance. Search Data Management*. https://www.techtarget.com/searchdatamanagement/tip/6-dimensions-of-data-quality-boost-data-performance

uCertify. (2026). *DAT–325 - Data validation, quality, and cleaning* [Online course]. https://snhu.ucertify.com/app/?func=load_course&course=SNHU-DAT325.AJC1&class_code=09YKy