Critical Thinking Module 5: Data Governance in Software Engineering and AI

Alexander Ricciardi

Colorado State University Global

CSC502: Ethical Leadership in Software Development

Professor: Dr. Steven Evans

February 15, 2026

Critical Thinking Module 5: Data Governance in Software Engineering and AI

Data has always been at the center of software engineering, but with the emergence of generative Artificial Intelligence (AI) and shifting toward building, deploying, and maintaining data-intensive systems, these systems depend on, produce, collect, and manipulate large amounts of data. In real-world applications, data can be incomplete, inconsistent, hard to find, or locked behind organizations' departments, legal or regulatory constraints, and unclear ownership. To address these issues, software engineers build data pipelines, shared data platforms, metadata catalogs, validation checks, and access controls. However, these technical solutions alone are not sufficient, as they cannot establish decision rights. In other words, they fail to define who holds the authority and accountability for selecting, defining, and approving access to data, as well as setting quality standards, managing changes, and enforcing compliance. Data governance fills this gap by establishing how an organization makes decisions about data, who is responsible, and what rules apply. This paper argues that, in software engineering and AI, data governance is a framework that converts data from a raw product into a valuable, truthful, and trustworthy organizational asset by defining decision rights, accountability, quality standards, and enforceable rules for how data is created, shared, protected, and used across the data lifecycle.

What Data Governance Is

Before defining data governance, it helps to clarify how it relates to data management, which is the process of handling data, such as collecting, storing, securing, and using data (O'Keefe & Brien, 2023). Data governance, by contrast, is an organization's decision-making framework that defines how data will be managed and describes who holds authority and accountability for those decisions (Holdsworth & Kosinski, n.d.; Thomas, n.d.).

These frameworks, in the context of software engineering and AI, help organizations to answer questions like:

- Who defines what a customer means across the software systems?
- Who approves a new data source for production use?
- What quality thresholds must data meet before it can be used by the AI agent?
- Who or which software processes can access sensitive data, and under what conditions?
- How long should data be retained in our cloud database, and when must it be deleted?

As these questions suggest, implementing a data governance framework may require significant organizational cultural change, along with ensuring that its operation includes clear structures, processes, roles, responsibilities, and decision-making rules.

Data Governance Improves the Value and Quality of Data

Implementing a data governance framework enhances the value of an organization's data assets by ensuring they are trusted, usable, shareable, and compliant. These governance outcomes improve decision-making, reduce operational friction, and protect stakeholder needs (Thomas, n.d.). Additionally, data governance establishes standards for what "good data" means and how to measure it, improving the value of the data. Furthermore, data governance helps establish definitions and rules to improve data quality. Wang and Strong (1996) argue that data quality is not based solely on accuracy and correctness; it also depends on data dimensions such as completeness, timeliness, and relevance. Data governance helps organizations define and measure these dimensions. In software engineering, this translates into practical controls, like:

- Shared definitions for data fields; for example, what counts as or is an active user.
- Data validation rules and processes.
- Data quality monitoring processes.

Ultimately, by defining data quality and enforcing it, data governance allows developers and analysts to develop and deploy software applications with confidence.

Data Governance Promotes Data Stewardship and Reduces Complexity

To be useful, data needs to be not only valuable and of high quality, but also easily discoverable, interpretable, and reusable. In other words, it must be accessible to those who need it in a format they can use (Sheldon, 2014). Data governance typically establishes metadata standards, data traceability, definition documentation, and data stewardship. Data stewardship is especially important as it helps identify data ownership and empowers data stewards to manage the data lifecycle and enforce data quality standards and reusability. This is critical as in software systems, data flows through complex services and pipelines, and data governance provides engineers with a framework to manage this complexity effectively. In other words, it enables engineers to debug faster, rely on more accurate analytics, and integrate new team members into the project with significantly less friction. Furthermore, it ensures that a project's stakeholders know exactly whom to contact for technical questions, approvals, or fixes; therefore, reducing delays and improving communication.

Data Governance Strengthens Data Security, Privacy, and Compliance

The value of data must also be protected by preventing external security breaches, unauthorized access by employees (insider threats), and non-compliance with regulations or laws. Failing to implement these security and compliance controls properly can result in fines, reputational damage, and operational disruption. The National Institute of Standards and Technology (NIST) states that failing to manage privacy risk can cause brand damage, financial loss, and legal liability (NIST, 2020a).

Data governance can support these privacy and security controls by defining a set of rules, access controls, and retention policies that dictate how the data is handled throughout its lifecycle. Additionally, it can reduce risks associated with laws, contracts, and compliance by ensuring that data management practices are well defined. In the context of software engineering, this often translates to implementing controls that ensure data processing is transparent, auditable, and compliant with ever-changing legal requirements. Furthermore, because data is the bedrock of both software analytics and AI, as such, these tools and the outputs they generate are only as transparent, auditable, and compliant as the data they ingest or are trained on. NIST's AI Risk Management Framework emphasizes that governance is a cross-cutting function throughout the AI lifecycle and is a cornerstone requirement for implementing effective AI risk management practices and policies (NIST, 2023). This implies that data governance is a prerequisite for performing effective and trustworthy analytics and for the development and deployment of safe and responsible AI systems.

Practical Applications in Software Engineering and AI

In practice and in the context of software engineering and AI, data governance has the most impact when it is built into every aspect of software engineering and AI lifecycles. For example, engineering teams can leverage data governance to make decisions at critical stages of the development process by reviewing and approving new data sources before production use, standardizing shared definitions for data fields, and enforcing data-quality thresholds using tests and monitoring. Additionally, data governance mandates standardized documentation that makes datasets and AI models easier to audit, trust, and reuse. For example, requiring a short datasheet for each dataset that documents why it exists, how it was collected, what it contains, and what it should (and should not) be used for (Gebru et al., 2018).

Finally, data governance supports secure and compliant engineering practices. For example, the Secure Software Development Framework (SSDF) establishes a set of security development practices that can be integrated into the Software Development Life Cycle (SDLC) (Souppaya et al., 2022). When combined with a data governance framework, SSDF enforces how the technical security data controls are applied inside the SDLC, while data governance defines who has organizational authority and responsibility to apply these controls and ensures that they are consistently enforced within the SDLC. Ultimately, data governance, in practice, acts as the connective tissue between technical controls and decision-making, ensuring that data, AI, and software systems remain reliable, secure, compliant, and fit for the duration of their development life cycles.

Conclusion

Software applications are increasingly integrating AI into their workflows and features and, as a result, are being built, deployed, and maintained as data-intensive systems that depend on, produce, collect, and manipulate large amounts of data. However, data is often inaccurate, incomplete, inconsistent, duplicated, outdated, or poorly formatted, and it can also be difficult to access as it can be restricted behind legal, regulatory, or ownership constraints. To address these issues, software engineers implement technical solutions such as data pipelines, shared data platforms, metadata catalogs, validation checks, and access controls. Even so, these technical solutions alone are not sufficient; they fail to define decision rights and accountability. Data governance fills this gap by establishing who has the authority to define data, approve access, set quality standards, manage changes, and enforce compliance. Additionally, when implemented well, data governance improves data value by making it more trusted, usable, shareable, and compliant.

Furthermore, it also strengthens stewardship, reducing operational complexity and supporting privacy and security requirements. Moreover, in AI software systems, where outputs depend heavily on the quality and governance of the underlying data, these controls are not optional; they are imperative. Ultimately, they act as a bridge between technical controls and organizational decision-making, ensuring that data, AI, and software systems are managed consistently, responsibly, and effectively across their full life cycles.

References

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). *Datasheets for datasets* (arXiv:1803.09010). arXiv. https://doi.org/10.48550/arXiv.1803.09010

Holdsworth, J., & Kosinski, M. (n.d.). *What is data governance?* IBM. https://www.ibm.com/think/topics/data-governance

NIST. (2020, January 16). *NIST privacy framework: A tool for improving privacy through enterprise risk management (Version 1.0)* (NIST CSWP 01162020) [PDF]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.01162020

NIST. (2023, January). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1) [PDF]. National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

O'Keefe, K., & Brien, D. O. (2023). Chapter 5: Ethics and data management (including AI). *Data ethics: Practical strategies for implementing ethical information management and governance*. Kogan Page.

Sheldon, R. (2014, November 19). *Data as a Service: The Next "As a Service" Wave?* https://www.red-gate.com/simple-talk/cloud/platform-as-a-service/data-as-a-service-the-next-as-a-service-wave/

Souppaya, M., Scarfone, K., & Dodson, D. (2022). *Secure software development framework (SSDF) version 1.1: Recommendations for mitigating the risk of software vulnerabilities* (NIST Special Publication 800-218) [PDF]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-218

Thomas, G. (n.d.). *The DGI data governance framework* [PDF]. The Data Governance Institute. https://neweditions.net/sites/default/files/sites/default/files/ACLDataCouncil/Data%20Governance%20Institute%202014%20Data%20Governance%20Framework.pdf

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099