An Evolutionary Strategy for Leveraging Data Risk-Based Software Development for Data Integrity

Organizational decision-making is largely dependent on the availability of adequate supporting data, data elements or records. These are perhaps the most important elements that can aid enterprises in making critical decisions promptly and efficiently. For example, healthcare organizations rely on electronic health and medical records to inform important decisions. However, in many organizations, when it comes to product code and software development and design models, data elements do not appear to carry equal importance compared to functionality. But there has been a gradual rise in recognition of the importance of data risk-based systems development across industries. Data risk-based systems development involves performing data element selection, analysis, cleansing and integration to highlight the critical data points and attributes to determine how to significantly limit risk. It recognizes that less data risk is linked with the objective of systems development with least possible risk.

Routine traditional software development methodology does not serve as a one-size-fits-all approach for every industry. For example, organizations specializing in life sciences and banking have certain key data elements that play critical roles in enterprise processes and are highly scrutinized via regulations and laws. Therefore, giving equal importance to identified critical data elements as functional specifications during the design and development phases results in effective systems development and helps avoid the cost of subsequent additional development activities to address data quality and regulatory findings.

Taking into consideration findings from data element risk analysis during the software development life cycle (SDLC) adds a higher level of confidentiality, integrity and protection controls to organizational data and business-critical processes. This can be implemented by proactively incorporating the results of data element risk analysis into user requirements, functional requirements, customization of coding standards, data modeling, system architecture and database design. Organizations operating in industries such as pharmaceuticals, finance and clinical research are routinely exposed to situations in which they must invent innovative products and systems while remaining compliant with regulations. Updating the traditional software development methodology to include data risk-based development enables organizations to avoid any penalties that may be imposed due to a lack of controls and noncompliance with regulations and laws, which makes data element risk-based development a highly lucrative approach for information systems development in many industries in the public and private sectors.[1, 2]

Traditional vs. Data Risk-Based Software Development

Traditionally, code or feature development in software development is either functional or performance-focused, which limits the entire system to functional and nonfunctional requirements gathered and illustrated through use cases. Functional requirements are focused on the functionality of the system, and performance requirements are focused on requirements such as load speed, maximum number of transactions, and scalability. Such a traditional approach can expose product development projects to failure as it does not consider the various classes and types of data to be included in the system. Conversely, the combination of data risk-based development and traditional

SASIDHAR DUGGINENI | CISA, CISM, ITIL FOUNDATION

Is a compliance manager at PPD, part of Thermo Fisher Scientific. Duggineni's professional background is in healthcare technology compliance, data integrity research, quality assurance, information systems auditing, information security, good practice (GxP) compliance and server administration. He has worked with specialized teams to prepare for International Organization for Standardization (ISO) standard ISO 27001 certification. He is also a lead auditor for his organization's supplier audit program and internal audit program. Duggineni participates in various operational committees within his organization related to information security, internal quality compliance and supplier compliance, and he has mentored many of his team members in information security auditing and data integrity.

It is arguable that the value of an information system almost always depends on the size and type of data that are stored, managed and used to conduct business processes.

systems development methodology not only gives adequate attention to critical data elements and classes that are managed and stored, but also aligns the data elements with the functional and performance attributes to be delivered.

It is arguable that the value of an information system almost always depends on the size and type of data that are stored, managed and used to conduct business processes. Departing from the traditional belief that effective information systems are those that fulfill functional and performance requirements, modern and evolved evidence supports the belief that such systems gain significance based on the data they process as inputs to generate meaningful outputs that can be utilized to the enterprise's benefit. Thus, a hybrid systems development model that incorporates data element risk analysis and adequate checks and balances (e.g., audit trails, security logs for critical data inputs, and forensic capabilities to ensure regulatory and legal compliance along with functional compliance) is essential. Also, by appropriately classifying and identifying the types of data elements to be captured by the system, enterprises bound by government regulations and laws can better respond to investigations, conduct forensics and manage incidents.[3,4]

The Birth of DQaaS

In response to the rising demand for inclusion of data element risk-based systems development across industries, a new form of service has emerged: Data Quality as a Service (DQaaS). DQaaS involves hiring third-party independent consultants who master the art of data profiling via exploratory data analysis and mining activities tailored to the industry compliance needs in question. The main activities involved in DQaaS include profiling, validation, cleansing and data stewardship-related services. These services can immensely help in determining an organization's position and need for development of new information systems. This is especially true for industries that need to include a data element-based development approach in their systems development methodology of choice, since the focus of this approach is to comprehensively classify and analyze the most relevant and valuable data elements in a system.[5]

Regulatory and Legislative Expectations

Meeting regulatory expectations for data integrity is crucial for many local, national and global organizations. These expectations are enforced in the form of regulations, rules and laws, such as:

- The Code of Federal Regulations (CFR) 21 enforced by the US Food and Drug Administration (FDA), which ensures data integrity in the pharmaceutical industry. CFR Part 11 went into effect in 1997 to extend data integrity regulations into the modern era with electronic records and electronic signature standards.[6]
- Guidelines on data integrity compliance introduced by the World Health Organization (WHO), the European Medical Agency, and the UK Medicines and Healthcare Products Regulatory Agency.[7,8,9]
- The US Sarbanes-Oxley Act (SOX), a US federal law that sets strict standards for safeguarding the integrity of financial data.[10]
- The US Health Insurance Portability and Accountability Act (HIPAA), which provides federal protection of patients' health data against

During the stages of requirement gathering for a product, emphasis should be placed on effective translation of data integrity requirements into formal product requirement documents for simple communication of data integrity expectations

misuse or exposure and requires technical and administrative controls to ensure compliance.[11]

Noncompliance with these laws and regulations can lead to serious consequences such as heavy fines, loss of reputation, products banned from markets, financial losses and lawsuits.[12]

Implementing Data Risk-Based Development

To help reduce data integrity risk and maintain compliance, data risk-based development can be implemented by incorporating a variety of controls during various phases of systems development. There is no one-size-fits-all approach to accomplish this goal. Beginning with coding standards, developers can be proactive and incorporate data integrity and data quality as code throughout the build processes, such as compilers and binaries, and in deployment technologies and processes involved in centralized source code management. Examples include:

- Ensuring that the code does not involve integrity violations such as modules or libraries from untrusted sources.
- Performing code reviews with targeted data integrity benchmarks.
- Developing critical system functionalities with required attributes such as audit trails, traceability and authentication.
- Enforcing the integrity of values in database columns and rows.
- Implementing logs and input data validations.
- Educating the workforce about the importance of the inclusion of data quality and integrity requirements in work processes.
- Cleansing and clearly defining all data elements that are critical for data integrity, quality and accountability, and further incorporating those specifications into coding standards to eliminate the gaps in data forensic and data audit trail system capabilities.
- Identifying appropriate data elements on a comprehensive scale to fit the functional and regulatory needs of the system, which helps significantly increase efficiency and coverage for proactively integrating those identified critical data elements into code building processes. This also enhances the compatibilities with external systems integrations for data transfers. This strategy leads to a significant decrease in reactive implementation of those specifications because of a discovered data integrity violation or regulatory noncompliance after the rollout of a specific build version of a system.

During the stages of requirement gathering for a product, emphasis should be placed on effective translation of data integrity requirements into formal product requirement documents for simple communication of data integrity expectations to product designers and developers from the people who use the product to conduct daily business operations.

Industry Case Study

Clinical trials are a good example of the value of implementing data risk-based development. The data gathered, stored and processed by information systems for clinical research organizations (CROs) play an important role in determining clinical efficacy and making critical decisions that impact patient safety. In addition, these data often are reported to regulatory agencies. However, critical requirements for data elements cannot be entirely accommodated by traditional software development methodologies wherein data classification and data elements are not given adequate attention during the application development stage. The validation of whether a clinical trial's success is directly dependent on the availability of supportive and attributable data, which is why the inclusion of data element risk-based development in any software development methodology of choice is critical. Systems developed with this hybrid methodology can readily perform data capturing, analysis and reporting, which are critical processes in applications such as clinical trials, finance operations and hospital procedures. Industries also benefit from data risk-based systems development inclusion because it allows them to retrieve information with data integrity, security, data trails and forensics built into the systems. This enables

LOOKING FOR MORE?

- Read *Defending Data Smartly. www.isaca.org/defending-data-smartly*
- Learn more about, discuss and collaborate on risk management in ISACA's Online Forums. *https://engage.isaca.org/onlineforums*

organizations to investigate any malpractice, privacy breaches or integrity compromises.

Figure 1 tabulates the differences between hybrid SDLC methodology with data element risk analysis and nonhybrid traditional SDLC methodology in terms of compliance requirements that both methodologies readily support.

Figure 2 is a comparative analysis of compliance defects encountered in a typical application life cycle in the span of 10 years using hybrid SDLC methodology with data element risk analysis vs. nonhybrid traditional SDLC methodology.[13]

Data Audit

When discussing data risk-based systems development, the concept of data audit cannot be ignored. Data audit is aimed at evaluating the quality and integrity of data that are available or accessible to business processes. It involves the adoption of a variety of tools and techniques to ensure that the current data sets are not only fit for purpose, but also of high quality and integrity. Here, data quality implies the authenticity and accuracy of data sets regarding organizational aims and objectives. For example, a data audit of a clinical trial enterprise would include checking the data for accuracy, integrity and reliability relevant to patient cases and the intended patient-specific outcomes. As another example, a financial sector organization may perform data audits to check the integrity and correctness of

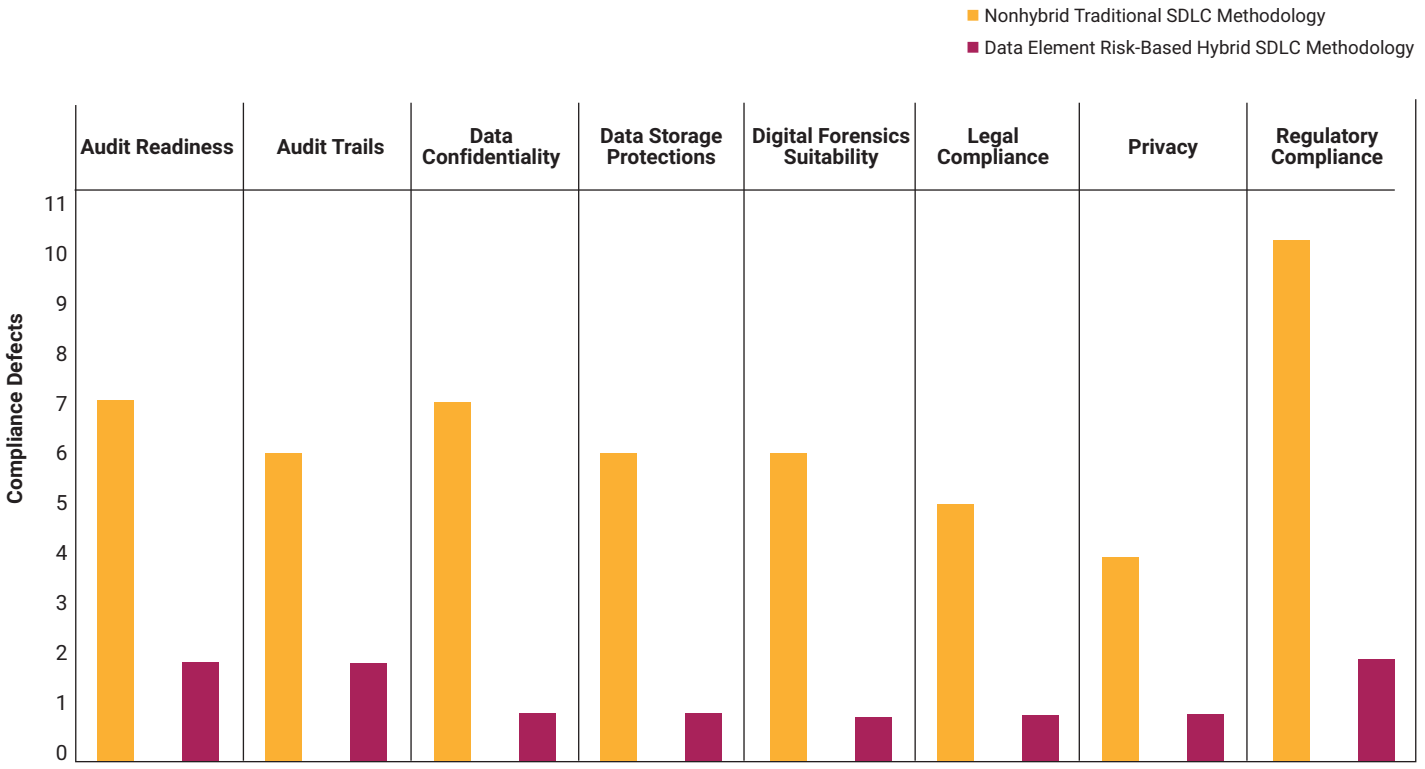
FIGURE 1

Compliance Compatibility

Compliance Requirement	Hybrid SDLC Methodology With Data Element Risk Analysis	Nonhybrid Traditional SDLC Methodology
Privacy	Fully supports	Partially supports
Audit trails	Fully supports	Partially supports
Data storage protections	Fully supports	Partially supports
Audit readiness	Fully supports	Does not support
Regulatory compliance	Fully supports	Partially supports
Legal compliance	Fully supports	Does not support
Digital forensics suitability	Fully supports	Partially supports
Data confidentiality	Fully supports	Partially supports

FIGURE 2

Compliance Defects in the Application Life Cycle



its financial statements and performance throughout the calendar year. In this sense, systems developed with the inclusion of data element risk in their development methodology are better suited for both regulatory and functional compliance audits because they comprise the key data elements and attributes that are directly relevant to the regulatory and functional compliance needs of the industry. In other words, these systems are ready for a divide-and-conquer approach to performing data audits, wherein they specifically identify and analyze the data elements of greater significance and compliance relevance. Such an approach also saves time and effort pertaining to data risk audits.[12]

Conclusion

The role of data in organizational decision-making is critical across all industries. Data provide goods and services and meaningful insights that enable systems owners and regulatory authority to make accurate and mission-critical decisions and confirmations. Although the traditional approach to software development tends to neglect data elements due to a higher focus on the code aspect of system development, combining the traditional software methodology of choice with data element risk-based system development puts data in a meaningful position in the systems development life cycle. Data element risk-based systems development involves performing data element selection, analysis, cleansing and integration to ensure that the system at hand fulfills its intended purpose completely. Data element risk-based systems development can help enterprises across all industries remain proactive in terms of compliance and control while also fulfilling the relevant functional and performance requirements of a system.

Endnotes

1 Kogan, A.; B. W. Mayhew; M. A. Vasarhelyi; "Audit Data Analytics Research—An Application of Design Science Methodology," *Accounting Horizons*, vol. 33, iss. 3, 2019, p.69–73

2 MD Group, "Importance of Data Capture, Analysis, and Reporting for Patient-Centric Clinical Trials," 1 October 2020, *https://mdgroup.com/blog/the-importance-of-data-capture-analysis-and-reporting-for-patient-retention-in-clinical-trials/*

3 US Department of Defense, Military Standard on Software Development and Documentation (MIL-STD-498), USA, 5 December 1994, *http://everyspec.com/MIL-STD/MIL-STD-0300-0499/MIL-STD-498_25500/*

4 Grow, "Why Is Data Important for Your Business?" *https://www.grow.com/blog/data-important-business*

5 Rajan S.; S. Narayanan; "Data Quality as a Service: Practical Guide to Implementation," Data Quality Pro, 2022, *https://www.dataqualitypro.com/blog/data-quality-as-a-service-practical-guide*

6 Code of Federal Regulations, *Part 11 - Electronic Records*; *Electronic Signatures*, USA, 20 March 1997, *https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11*

7 European Medicines Agency, "Data Integrity: Key to Public Health Protection," 8 November 2016, *https://www.ema.europa.eu/en/news/data-integrity-key-public-health-protection*

8 World Health Organization, *Guideline on Data Integrity*, Switzerland, June 2020, *https://www.who.int/docs/default-source/medicines/norms-and-standards/current-projects/qas19-819-rev1-guideline-on-data-integrity.pdf*

9 Medicines and Healthcare Products Regulatory Agency, "MHRA GxP Data Integrity Definitions and Guidance for Industry," United Kingdom, 21 July 2016, *https://www.gov.uk/government/news/mhra-gxp-data-integrity-definitions-and-guidance-for-industry*

10 HR 3763 Sarbanes-Oxley Act of 2002, USA, 2002, *https://www.congress.gov/bill/107th-congress/house-bill/3763*

11 Edemekong, P. F.; P. Annamaraju; M. J. Haydel; "Health Insurance Portability and Accountability Act," National Library of Medicine, USA, 3 February 2022, *https://www.ncbi.nlm.nih.gov/books/NBK500019/*

12 D'Halluin, C.; "The Importance of Data Integrity in the Finance Industry," *Payments Journal*, 6 December 2021, *https://www.paymentsjournal.com/the-importance-of-data-integrity-in-the-finance-industry/*

13 Deloitte, *Under the Spotlight: Data Integrity in Life Sciences*, United Kingdom, 2017, *www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-data-integrity-report.pdf*