

Data Lineage and Compliance

Data lineage is gaining momentum as the volume of data and complexity of systems environments and compliance requirements continue to grow. For those unfamiliar with the concept, data lineage is a way of looking at an organization's data throughout the entirety of their cycle—from original to final destination—by creating a visual representation of the flow of data throughout the organization.

Data architecture and data governance activities can be bolstered by data lineage. Specifically, these activities rely on metadata (data about the data). Data lineage can help to document the different processes, business rules, dependencies and other attributes that explain where data come from and how they were used to calculate results. These metadata can be created manually or by automated information processing means. Manual metadata can include any attribute that is important to IT or the business (**figure 1**), while metadata captured automatically can include information about when data were created; who created the data sets; how data were acquired, added or deleted from a data set, system or interface; and when they were last updated in any data repository or system in scope.¹

Figure 1—Examples of Metadata

Field Name	Field Type	Size (bits)
Object ID	UID	128
Object size	WORD	64
Title length	WORD	16
Author length	WORD	16
Copyright	CHAR	Varies
Description	CHAR	Varies
Author	CHAR	Varies
Audience	CHAR	Varies
Purpose	CHAR	Varies
Scope	CHAR	Varies

Source: E. Sweet. Reprinted with permission.

How Does Data Lineage Help Enterprises Meet Compliance?

Data lineage became more prevalent during the financial crisis (which started in 2007) when Global Systemically Important Banks (G-SIBs) were unable to report accurate risk exposure because they “lacked the ability to aggregate risk exposures and identify concentrations quickly and accurately at the bank's group level.”² This deficiency was considered significant and, as a result, the Basel Committee issued guidance to help banks improve their risk aggregation and reporting capabilities. A key component of the guidance was the use of data lineage as the means to discover the different sources of data and the data flow used to create risk exposure reports.

By January 2013, the Basel Committee issued Banking Supervision regulation number 239 (BCBS 239)³ to introduce 14 principles for effective risk data aggregation and risk reporting. Banks must be able to demonstrate the data flow used to create risk reports, and this is where data lineage becomes a powerful tool to:

- Discover data sources
- Document the data flow
- Create a visual representation of the data flow
- Establish a single source of truth that all stakeholders can trust and rely on for risk reporting

The 14 principles developed by the Basel Committee to provide guidance over risk aggregation and reporting are listed in **figure 2**.

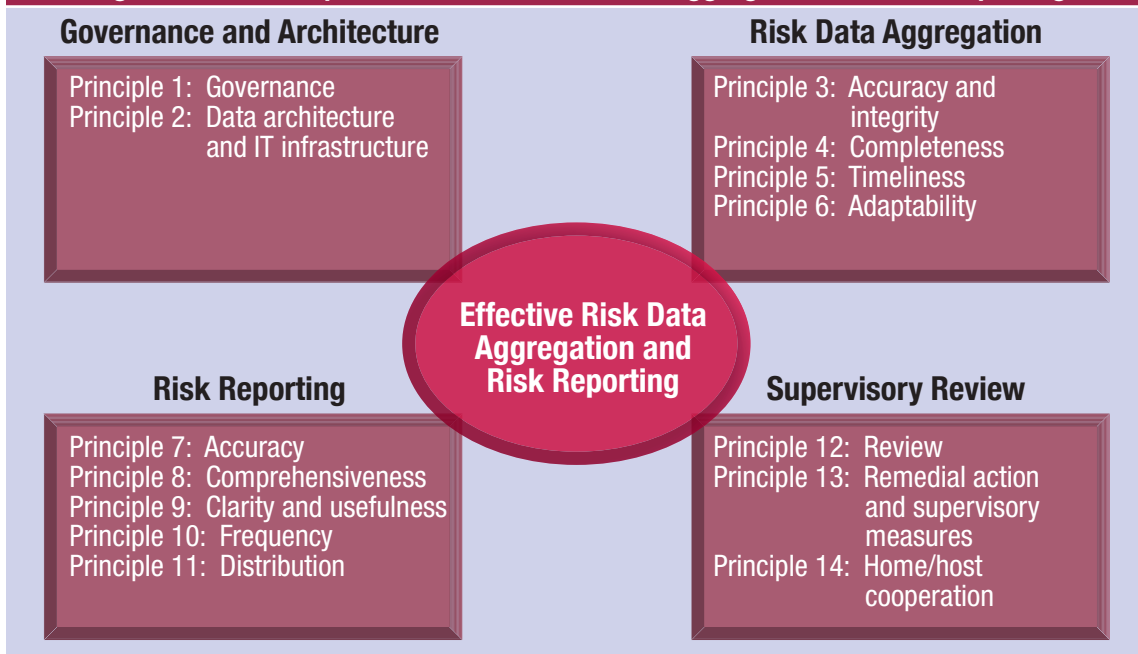
Other regulations such as the Fundamental Review of the Trading Book (FRTB)⁴ and the UK's Senior Managers Regime (SMR)⁵ require banks to fully comply with the principles and reporting requirements established by BCBS 239 to strengthen risk management and board and senior management responsibility and accountability. Banks that fail to comply with BCBS 239 may experience increased scrutiny, be required to increase available capital, and/or be limited in their risk-taking and growth opportunities.⁶

Data lineage prominence started in the banking sector, but in today's environment of statutory

Eva Sweet, CISA, CISM

Is a technical research manager at ISACA and IT professional with more than 15 years of experience in IT operations, security and audit. In her current role, she authors thought leadership and practical guidance publications that focus on risk and assurance.

Figure 2—14 Principles for Effective Risk Data Aggregation and Risk Reporting



Source: E. Sweet. Reprinted with permission.

regulations, its adoption is becoming an important step to meet compliance with regulations such as the US Health Insurance Portability and Accountability Act (HIPAA), the US Sarbanes-Oxley Act (SOX), and the EU General Data Protection Regulation (GDPR). Furthermore, enterprises that have large and complex IT environments and are concerned with sensitive data or looking to reap the benefits of big data can benefit from the ability to visualize data flows through the use of data lineage.

How Does Data Lineage Work?

To fully exploit the benefits of using data lineage to demonstrate reporting accuracy, enterprises need to lay the foundation by conducting a full discovery of data sources, processing systems and interconnections, and creating a metadata repository for all the data attributes needed to describe the data flow from origin to destination.

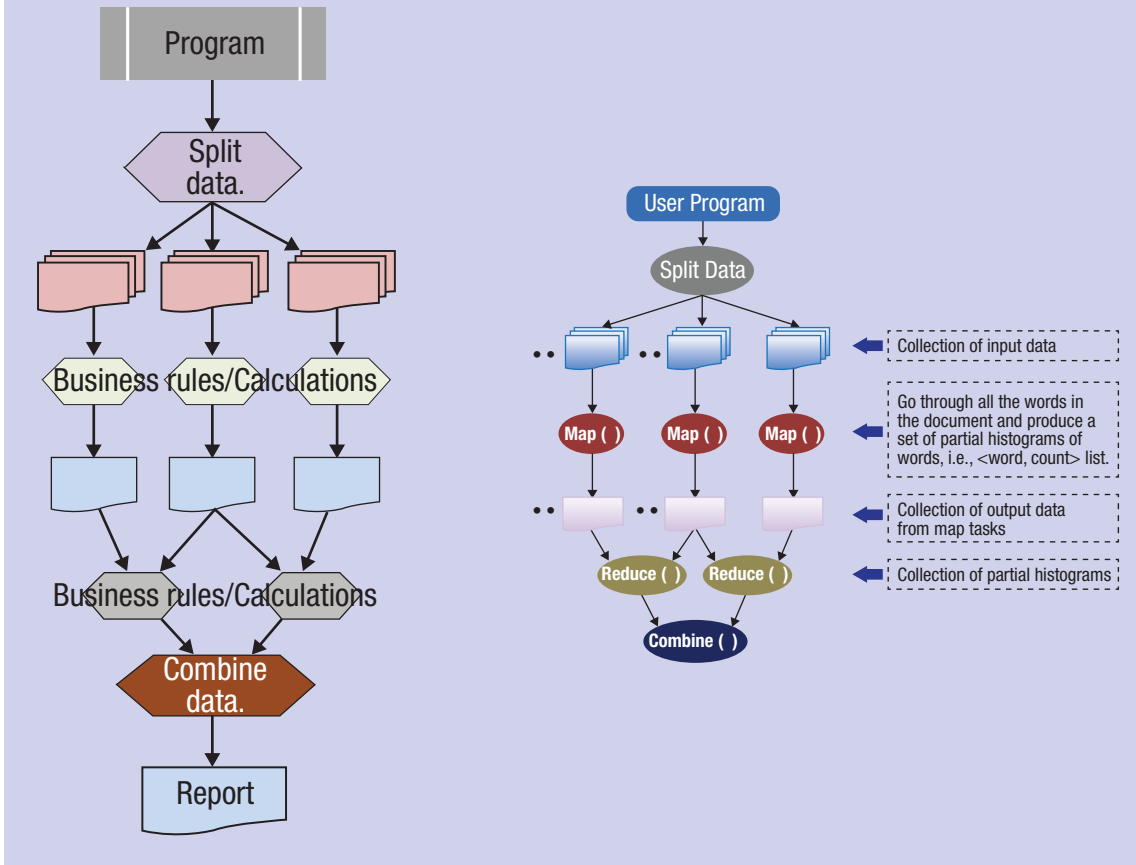
The metadata repository is a key element for data lineage quality, so it must be properly managed and

maintained. Some of the data attributes that should be stored in the metadata repository include:

- Users, systems and processes involved in the data flow, and their dependencies
- Business terms, rules, operational procedures and scripts used to process data
- Database names, table names and data types
- Reporting systems
- Time stamps for each time data are changed, created, deleted, added and processed
- Data location, owner, steward, format and retention policies

Data lineage uses the metadata repository to create a report that describes the entire flow of data, from all sources to the final report(s). The resulting data lineage report lists each field and value stored in the metadata repository related to a specific data flow. The data lineage report can be used to depict a visual map of the data flow that can help determine quickly where

Figure 3—Visual Representation of a Data Lineage Report



Source: E. Sweet. Reprinted with permission.

data originated, what processes and business rules were used in the calculations that will be reported, and what reports used the results. **Figure 3** shows the visual representation of a data lineage report.

Benefits of Data Lineage

The benefits for enterprises that adopt the methodology as part of their data governance program include:

- Accurate reports and metrics
- Identification of duplicate data or processes
- Establishment of a consistent glossary of business terms across all affiliates
- Increased efficiency in regulatory reporting
- Improved data analysis and decision making
- Realization of the benefits that data lineage offers
- Continuous improvement of data governance
- Identification of business rules discrepancies
- Potential to identify security breaches or exposure of sensitive data
- Identification of redundant processes, systems, data or business rules
- Compliance with regulations such as SOX and HIPAA
- Improved data and systems architecture
- Improved change management and new system implementation by understanding potential impacts on data flows

Conclusion

The amount of data that enterprises need to expand their business and gain competitive advantage will continue to grow just like the number of regulations needed to ensure that enterprises compete in an ethical and responsible manner. The combination of large volumes of data and more stringent statutory regulations can be sufficient justification to implement data governance practices, including data lineage. One example is Aspen Insurance, which adopted data governance as the first step to meet compliance with Solvency II, a new European regulation similar to BCBS 239, which mandates good data management practices to ensure proper solvency reporting. (BCBS 239 mandates good data management practices to ensure proper risk reporting.) Aspen Insurance is realizing the benefits of documenting data flows, including the creation and use of data, significant calculations, data quality rules and process controls by being able to meet Solvency II compliance.⁷

Compliance is a good business case for data lineage; however, the benefits expand beyond satisfying regulations and into enabling the business environment to use big data and analysis because good quality data will yield good quality reporting for decision making.

Editor's Note

This article describes one of the top compliance challenges discussed at the 2016 IT Audit Director Forum held at the North America CACS conference in May in New Orleans, Louisiana, USA. For more information on this forum, see the briefing white paper titled *Insights From the 2016 IT Audit Director Forums: Top IT Audit Leaders from Around the World Share Knowledge*, which can be accessed at www.isaca.org/it-audit-director-forum-insights.

Endnotes

- 1 Berson, A.; L. Dubov; *Master Data Management and Customer Data Integration for the Global Enterprise*, McGraw-Hill, USA, 2007, <http://searchitchannel.techtarget.com/feature/The-benefits-of-metadata-and-implementing-a-metadata-management-strategy>
- 2 Basel Committee on Banking Supervision, *Principles for Effective Risk Data Aggregation and Risk Reporting*, January 2013, www.bis.org/publ/bcbs239.pdf
- 3 *Ibid.*
- 4 Bank for International Settlements, *Fundamental Review of the Trading Book: A Revised Market Risk Framework*, October 2013, www.bis.org/publ/bcbs265.pdf
- 5 Financial Conduct Authority and Prudential Regulation Authority, *The Senior Managers Regime*, UK, July 2015, <https://www.fca.org.uk/news/fca-publishes-final-rules-to-make-those-in-the-banking-sector-more-accountable>
- 6 Ernst & Young, *BCBS 239 Risk Data Aggregation and Reporting: A Practical Path to Compliance and Delivering Business Value*, 2015, [www.ey.com/Publication/vwLUAssets/EY-bcbs-239-risk-data-aggregation-reporting-AU/\\$FILE/EY-bcbs-239-risk-data-aggregation-reporting-AU.pdf](http://www.ey.com/Publication/vwLUAssets/EY-bcbs-239-risk-data-aggregation-reporting-AU/$FILE/EY-bcbs-239-risk-data-aggregation-reporting-AU.pdf)
- 7 Colibra, *Customer stories*, www.colibra.com/stories/