

EDM Webinar

Data Governance Reality Check

Live Date: June 18, 2020

Featuring:

Danny Sandwell, Director of Product Marketing

Recording: bit.ly/3hFmuDx

Presentation: bit.ly/2CyTtcN

EDM Council 2020 Benchmark Report: bit.ly/2YbKer6

Dataversity 2020 State of Data Governance and Automation Report: bit.ly/3hMraY6

edmcouncil.org

erwin.com

WEBINAR Q&A:

Are RDF/Graph Databases also supported by erwin for lineage?

erwin mimics a graph approach when determining lineage in DI. At erwin, we support any JDBC compliant data source using our Standard connectors. Additionally, we could configure a Smart Connector for any Graph DB that is not JDBC compliant.

For an organization that is not very mature with data management practices and data governance - how mature should the organization be before even thinking of bringing automated tooling in?

This is the quintessential question. Many a DG initiative has failed, not due to poor technology or vision, but due to lack of executive and/or organizational support and understanding. Job 1 is to ensure you have buy-in, sponsorship and funding from executive management and the key parts of the organization that will be responsible for providing this capability. This is a journey of learning. One way we have seen customers overcome this challenge is by documenting their data landscape, architecture and technologies. Data modeling is one discipline that can help. Reverse engineering your data sources and aggregating them into an enterprise model will give you that kind of visibility and acumen around the problem. Also, cataloging those models as well as native metadata can form the foundation upon which you can communicate the use cases and value of DG/DI and any related automation.

Does this study cover also public administrations in EU, or only private ones?

This study covers public and private entities. It is primarily NA focused although many of the private entities have a global reach.

What do you look at when assessing metadata quality as a key starting point to a data quality program?

Metadata completeness is the first step. Metadata represents what “should” be in the data source in terms of structure, characteristics and control. Harvest the source metadata and assess where the gaps are. Data profiling and the statistics that result can help identify where instance data is not conforming. Enrich that metadata as a guide to address any vulnerabilities.

What’s the difference between Metadata Management and Data Catalog?

Great question. Metadata management is the process of harvesting, centralizing and organizing physical metadata for the purpose of documenting, de-siloing and relating metadata. This provides your data landscape and architecture. Data Cataloging is generally the next step. Metadata is a key foundational element for the Catalog, but it goes well beyond the physical metadata. The Data Catalog enables you to curate and enrich the physical metadata with business and semantic metadata to provide a holistic view into data source certification, socialization and activation. This usually includes business context, operational usage and alignment. The goal is to provide a one stop shop for DQ metrics, structure, business taxonomies and usage guidances that accelerate “fit for use” analysis and data access and preparation efforts.

Where does Data Governance live? From your experience, where should accountability for Data Governance sit to ensure it's continuously effective, sustainable, and evolves as business and data landscapes evolve?

Ideally Data Governance should be a joint initiative owned by the business (consumers) and the data management group (providers). Data Governance that is wholly driven by Data Management often becomes another enterprise silo and struggles to get the funding, visibility and support required to make it successful at the enterprise level. By ensuring that the business (non-IT) folks have skin in the game and a place at the table to define requirements, use cases, responsibilities etc., data governance has a much better chance of delivering the results and ROI required to keep it viable over time.

Is impact analysis only related to financial impact? Are there any other areas of impact that can attract executives buy-in?

Beyond financials (data valuation) and data management optimization, impact analysis also supports risk and compliance efforts, audit responses of all kinds, innovation and transformation management as well as overall business optimization.

What use cases did the generated code support?

- Reducing costs, efforts and increasing speed and agility in data movement and data integration efforts.
- Built in governance of data movement and integration platforms.
- Platform modernization initiatives. (Datavault, governed data lake, EDW)
- Platform and data migration initiatives. (i.e. Datastage to Talend, Oracle to Teradata)
- Cloud Adoption (Snowflake, Maria DB etc.)

What is the structure and mix of companies that participated in this benchmark survey?

- Surveyed North American organizations
- Respondents in data or data architecture roles, information/data governance professionals, business intelligence/analysts, data scientists and executive management
- Organizations in technology, government, consulting, insurance, finance, healthcare and manufacturing among others
- All sizes but majority 101 to 5,000 employees (40%) then 10,001 to 50,000 (17%)

What technology or platform did the generated code support? RDBMS or unstructured?

Ideally, the ability to auto-generate code should cross multiple technologies. RDBMS procedural code, Hadoop and other un-structured DBMS technologies, modern scripting languages and ELT tools, and of course, traditional ETL platforms. With erwin DI we support all of these platforms. Automated code generation (forward engineering) combined with our ability to auto-document existing code (reverse engineer) into technology agnostic metadata mappings gives you the ability to centralized the development and management of data movement processes, optimize existing platforms, migrate logic between platforms and automate the manipulation of data structures, data movement processes as well as the loading of data into these platforms.

Can the best practices on ontology and taxonomy be shared?

Best practices around ontology and taxonomy starts with deploying an ISO standard glossary with bulk import ability. ISO provides the standard and structure for glossary authoring and management. This allows you to leverage industry best practices (i.e. FIBO) to bulk populate proven terminology that can then be used as the foundation to massage to your specific terminology and usage. Additionally, importing naming standards and definitions from data models is another way to leverage existing work product and accelerate glossary development and deployment. erwin automates the ingestion of erwin data models, best practices data models and continues to develop relationships with industry best practices bodies such as the EDM Council to assure access to proven taxonomies. We provide delimited files of best practices taxonomies and ontologies on our support site for customers to download and ingest into our solution.