Data Governance in Theory vs. Data Governance in Practice

Live Date: January 11, 2022

Featuring:
Rachel Haines, Principal Solutions Architect, Quest Software

Recording: View webinar

Presentation: View slide deck

EDM Council Homepage: edmcouncil.org **Quest Homepage:** quest.com

WEBINAR Q&A:

Thank you to Rachel Haines for providing the below answers to all questions posed during the live webinar. For more information or additional questions, contact rachel.haines@quest.com.

I don't see architecture any place on the slide. Are we considering this covered under Technology?

I think of "architecture" as a technical discipline which, in a people-process-technology framework, could be part of either the technology landscape or process models.

Is there a practice + "hands-on" + "real-thing" part to this presentation? So far it has been like someone reading a book...

Thanks for your feedback. In such a short presentation it's hard to be all things to all people. That said, if you would like, I'd be happy to schedule a "hands-on" "real-thing" session with you, your data, and your processes so that I can provide my feedback to you about your Data Governance practice. As for the book, I'd be happy to send you a copy of my book when it becomes available.

Much of what is being discussed as Data Governance is classified as Data Management by some frameworks (DCAM, for example). These frameworks consider DG to be oversight of the DM activities. My experience has been that this broader interpretation of DG is fairly common outside of Financial Services. Could you say a bit more about your distinction between DG and DM and any observations about the terminology used in different industry sectors?

I think of the term Data Management as an umbrella, or holistic term which encompasses a number of distinct disciplines. These distinct disciplines can include Data Quality, Master Data Management, Security, Enterprise Architecture, DevOps, etc. (basically, any activity which is concerned with the creation, management, and usage of data assets).

With this in mind, Data Governance is more about the policies, processes, and procedures which "govern" the execution of these disciplines.

For me, Data Governance is the glue that exists between the Data Management discipline silos, and as such, seeks to provide a level of consistency and transparency across the enterprise.

What do you think is the importance of a Business Glossary is vs a Data Dictionary?

I think both are key part of the data management architecture. The difference between a Business Glossary and a Data Dictionary is, for me, the context of the user requesting information. For example, a Business Glossary may be thought of as a collection of business-facing logical concepts. Individual terms in the Business Glossary may or may not be represented by 1 or more physical assets in 1 or more data repositories, but terms in the Business Glossary may represent purely logical concepts (or domains) such as "customer" or order-to-cash. I tend to think of a Data Dictionary as more of a technical catalog of columns in a database. Whereas the Business Glossary may detail metadata like term owner, classification, and business rules, examples of the metadata associated with attributes in the Data Dictionary could be data type, length, nullability, PK/FK relationships, etc. In a mature data management environment, a user (business or technical) would be able to enter either the Business Glossary or Data Dictionary and be able to search for the information they need, then be able to, for example, drill down, drill up, or drill sideways in order to be able to understand, for example, the physical data associated with a Business term, or the Business terms associated with a particular physical attribute.

it is important to always have a vision, charter and a strategy for DG. In the context of M&A what kind of "framework" has been found to be successful to successfully initate and implement a data governance program?

Specifically for M&A, it is vital that the data, processes, and functionalities be rationalized and harmonized into either the parent organization or the newly created organization resulting from the merger.

With this in mind I would suggest that one of the primary goals of the Data Governance framework would be to facilitate this rationalization and harmonization. This probably means a greater emphasis on:

a) having a solid set of documentation around the data, processes, and roles of the acquiring company

b) rapid ingestion and documentation of the acquired data, processes, and people

c) analysis of gaps between the datasets, functions, and roles of the 2 organizations

d) prioritization of the gaps

e) rationalization and harmonization to close the gaps

I feel that the Data Governance framework should be able to provide tools, templates, and patterns which facilitate and provide guidance for these steps.

Isn’t it important to identify change champions? If so what kind fo change management strategy is applicable for DG programs? Are there more than one if so what are they?

Yes, there will be people throughout the organization that are more “bought-in” or visionary or influential when it comes to moving the culture of the organization forward to a more data-centric approach, and it is important to work to keep these people in your Data Governance corner.

But beyond allying with certain individuals, a key element of the Change Management strategy is that the Data Governance team communicate widely and regularly. Each Data Governance Team should develop a detailed Change Management communications plan that identifies what you want to communicate, to whom you want to communicate, what communication channels are available, who will initiate communications, and when those communications will be delivered. And don't forget to include channels for a 2-way conversation with your stakeholders.

Data Governance should be both transparent as well as responsive to be successful.

Do you differentiate b/w DG and Enterprise DG? Are they the same and do they encompass the same kind of elements/components?

Often organizations will build one or more “federated” instances of data governance before moving to an Enterprise Data Governance program. By building federated data governance first, the organization has the opportunity to grow a set of experienced and knowledgeable individuals who can contribute to an Enterprise DG strategy. Federated DG also provides the opportunity to build a set of tested policies, processes, and procedures, which can then be scaled up to an Enterprise level.

Of course as we build out multiple federated instances, we should take care to maintain alignment, consistency, and re-use across those instances. One guiding principle for federated data governance should be to think globally but act locally and thus, to build collateral which can easily be scaled to the Enterprise.

For me, Enterprise Data Governance is like “Data Governance University” whereas local or federated Data Governance is more a place to learn the practical skills required to operationally govern data.

DG Operating Model: where is a Data Governance Office???

On slide 13 the entire right side of the diagram is concerned with the Data Governance Office and the capabilities that the office should possess.

Can existing data governance framework be applied in the governance of data ingested and produced by an AI tool?

Regardless of where or how data is created, part of a robust data governance / data management framework should be to be able to fully understand and visualize the lineage of the data. In the case of AI-generated data, the inputs to, and outputs of the AI model should be documented in the same way the inputs and outputs for any application / data repository should be documented.

How important are DCAM Assessments in the data governance journey?

Successful completion of the Data Governance journey requires, like all journeys, an evaluation of the where you want to go, what resources you have to get there, and what is the best roadmap to arrive at the destination within the constraints of time and resources.

Completing an assessment (like DCAM) before setting out on the journey will be a valuable tool for:
a) understanding your current state
b) establishing future state goals and metrics for measuring attainment of those goals
c) identifying gaps between the current state and desired future state
d) building and executing a realistic roadmap for undertaking the Data Governance journey

In the data governance cycle, don't we have a risk to include data in catalog before analysing and improve if there are qualities' issues?

For me, part of the reason we create catalogs is to document everything we do or can know about the underlying data. This includes understand what the definition of “clean” data is. If there are constrains on the maximum of minimum values for an element, or is there is a specific format mask for the data, we need to documented these business rules in the data catalog. These documented business rules then become inputs for the Data Quality function to test the real data against the expected level of data quality, and where defects are discovered, those defects can be triaged, prioritized, analyzed, and remediated.

Further, having the rules for clean data documented in the data Catalog allows us to continuously monitor the quality of the data over time, and identify when and where corrective actions should be taken.

How do you think DG has changed in the past 3 years?

I see Data Governance becoming more automated. And that's a good thing. As our datasets become larger and move at an increased velocity, it is important that we leverage automation. Without automation we, as Data Governance practitioners, run the risk of never being able to get caught up and clear our technical debt. 2 specific ways automation has made a big impact are in the areas of ingestion and classification of data. Being able to auto-document huge databases, data schemas, and data models has become "table-stakes" for any modern data governance system. Similarly the use of AI to identify and classify PII, PHI, and PCI is helping to accelerate our compliance to regulations while at the same time lowering the risk associated with data breach or data misuse.

What are some typical domain models (i.e. subject areas) for practical ownership for data in the catalog? I.e. Are there industry standard models?

There are industry standard models for many industries. A good source for some of these models is ADRM.com. I have specifically used their Gas & Electric Utilities model to help accelerate a Data Governance Implementation program at a utility in Southern California. I found their model to be quite comprehensive and inclusive of many areas of the business (like generation, distribution, and asset management).

Data Governance typically has a data-centric starting point, applying governance across Data Domains that cut across organisational structures. Then many organisations have moved to SAFe / Agile delivery structures that adds an additional virtual structure / complexity. What's your experience on "what works" in terms of integrating data-centric data governance in this context?

If users come to the metadata from a data-driven context, great. If the organization is more product or delivery focused, then Data Governance should be flexible enough to be able to answer the types of questions which may result for that context.

The key is that Data Governance should always strive to be "useful" to its stakeholders. To this end, we should provide Business-facing glossaries and technology-facing dictionaries as well as being able to present data/metadata in tables, visualizations, reports, and dashboards.

Are there any excel templates that you could point to for documenting processes and data dictionary?

I don't know that I could point you to a specific Excel template for documenting processes, but I would suggest that you investigate a SIPOC methodology for this task. SIPOC stands for Suppliers, Inputs, Processes, Outputs, and Customers. It is a way to examine a task and understand what the task is doing, what data is input to the task, what are the deliverables of the task, and who is involved. Here's a link to an article you may find useful: https://searchcio.techtarget.com/definition/SIPOC-diagram-suppliers-inputs-process-outputs-customers

What are some key roles for the organization's Data Governance Office - this is beside Chief Data Officer, also all other functions such as CRO, CFO, CIO may exisit outside the Data Governance Office?

Within the Data Governance Office I tend to have a leader (who I call the Chief Data Evangelist), one or more Data Governance Analysts (who can help stakeholders get familiar with the concepts and tools of DG and can help complete metadata documentation), a technical writer (who will assist with the drafting of policies, standards, and processes), and perhaps a trainer / course developer (who can help build out instructor led and self-service training collateral).

Of course the size and makeup of the Data Governance Office will vary depending on the amount of data being governed and the specific goals of the organization.

You mention DGO is the facilitator + educator - what advice can you provide when business stakeholders and even the DG sponsor believe the DGO should be performing the fieldwork e.g.: filling in data dictionaries, documenting business processes, etc. work often performed by business analysts?

The only group that knows the Business meaning of the data is… the Business. Yes, the Data Governance Office can help the Business understand how to capture and store the metadata and business processes, but the source of truth will always be the Business. If the DGO is tasked with doing all the work, then there may be scaling issues around having enough DGO staff to complete the work. I feel that the proper use of the DGO is to facilitate and educate so that the Business can develop the skills to be able to effectively manage their own data. And as a bonus, when the Business becomes involved, they become invested in the result... which leads to a better, more complete correct and consistent set of metadata.

How do you think Data Governance has changed in the past 3 years?

I see Data Governance becoming more automated. And that's a good thing. As our datasets become larger and move at an increased velocity, it is important that we leverage automation. Without automation we, as Data Governance practitioners, run the risk of never being able to get caught up and clear our technical debt. 2 specific ways automation has made a big impact are in the areas of ingestion and classification of data. Being able to auto-document huge databases, data schemas, and data models has become "table-stakes" for any modern data governance system. Similarly the use of AI to

identify and classify PII, PHI, and PCI is helping to accelerate our compliance to regulations while at the same time lowering the risk associated with data breach or data misuse.

What pitfalls to look for creating data issues register?

One pitfall I've experienced is where a leader tries to jump to the front of the line and get their issue fixed first. To counter this I would suggest developing a multi-dimensional issue scoring mechanism which is applied evenly across all issues and which outputs a score that can be compared to all other issues. Examples of dimensions which might be considered in the score could include expected ROI for the fix, complexity of the issue, resource availability, business impact, regulatory impact, security impact, availability of a work-around, risk score, opportunity cost, etc.

How are you measuring ROI for data governance?

Over time DG should enable faster development cycles because data meaning is more readily ascertained. Documenting the meaning of each term / attribute should also lower the risk associated with misuse of data. But these are long-term metrics and will take a while to collect and analyze. More immediately we can use the number of views or searches performed against the metadata as a way of understanding if the documented metadata, policies, and standards are being found to be "useful". More views = more usefulness. We can also poll the sentiment around the provided collateral (data glossary, data dictionary, documented standards and policies) as a way of understanding if the stakeholders are finding Data Governance to be useful or a roadblock.

Is it better to have tool first or process/policy first as you work on Minimum Viable Data Governance Program?

I personally like to come at it from the perspective of "what do we need to do this job" rather than "how to we use the functions of the tool to do the job". To some extent I would say it's more important to understand why we are doing things like gathering metadata and documenting policies rather than how to use a specific tool. I think having a foundational understanding of the mechanisms of DG will help us select the tool that better meets our needs.

Data Ownership is a challenging aspect to manage. People are reluctant to take ownership. What in your experience has helped increase data ownership in a firm?

I think we need to make it clear that "ownership" is more about the rights and obligations of data creation and use than it is about punishment for data errors. It seems to me that too often people are reluctant to take ownership because they feel that ownership means they will be held responsible for data errors (which may not be their fault).

Can you share some use cases where new technologies e.g. AI, Block Chain support Data Governance?

AI is being used to identify and classify data for PII, PHI, PCI, and privacy concerns. I've not heard of a use case for blockchain in the management of metadata.

Whose role is to do data classification? IT/Data Governance Office or Busiess stewards?

I believe that the Business Stewards have the most contextual knowledge about what the data means and the impact of changes to the data. They should, therefore, be involved in the development of a classification scheme which would be applied consistently to all data. For example, we may want to ask a series of questions about the data to determine its classification: IS the data PII/PHI/ PCI ? If so, classify as “Critical”. Does the data have a regulatory constraint ? If so, classify as “Critical”. Is the data key to the completion of a Business process? If so, classify as “Critical”. and so on and so forth.

Integration and execution of the final set of classification questions into the data/metadata documentation process will be the responsibility of all data stakeholders (IT / Business / DGO).

Do you think it is important to identify Critical Data Elements or can you implement data management based on data flows only?

You can take it from a data only perspective. You want to manage all of the data in this flow. It may not all be critical, but you prioritize the data you need to solve a specific problem. There are many ways to eat this elephant. Critical data tends to rise to the top because of compliance, audit or privacy issues. There are other ways to group data but as long as we're not trying to do everything all at once. So yes, doing it by data flow would be legit.

What is the relationship between Enterprise Data Governance and Enterprise Architecture, if any?

I think architects and data governance people have to work together. As we look at our architectural landscape, we know there will be places where data is more real-time or when it is coming from third-party sources. We want to be able to understand the content, and look to architects to and say "hey this is more how we should be crafting this data because of its use and classification, or its criticality." So those disciplines are our siblings within the organization, but one of the biggest success criteria for a data governance program is to communicate with everybody. It's important to have the right elevator pitch, that way when you're with the CIO talking about data governance, you can say this is what we're trying to do. Have a pitch for C-suite, middle management and the technology saying this is what we're putting together and this is what we want you to participate in. Communication with architects and the people making decisions within your company is important.

How important are data subdomains in the implementation of data governance?

You have to start somewhere. We're looking to do all of the customer data at a high level. We need to make sure when we communicate with our customers, we're precise and correct. So you can do it at a high level or you could step down to a sub domain and think about just addresses for a customer. We need to make sure that invoice gets to the right place. Not the mailing address, but the financial address or the shipping address versus the marketing address. So if you have a problem at a data sub-domain level, let's attack the problems where they are, whether high level or low level.

Set of questions: Would you have a concrete example where the practice failed due to a failure in implementing the Theory? Is the "theory" always the same regardless the industry, business? What would you advise to properly "balance" theory vs practice for a successful DG program? What about buy-in and top management involvement/sponsorship for its success?

There is much ink spilled about how data governance should be. Authors have put out books; you can go to conferences and they're going to say the business needs to run as a collaboration. There are big policies and processes that need to be put in place. Sometimes if you try to build a data governance from top down, the organizations may not have things to do. It's important to work with the data, but we don't need to have all the policies in place and we don't need to know everything before we get started. I find that if you try to build a data governance organization from the top down, that's a hard lift. Bottom-up? Same sorts of things. People are working with data every day, so they have people doing bottom-up governance. They recognize that data is important to their business function, and the data needs to be fixed correctly. They can only get so far if there's not consistency across that bottom layer. So let's do

both. Let's do top-down, let's do bottom-up. It's important to have buy-in from leadership and want people that will resolve these issues at the lower levels. That's where the balance comes. Making sure that you're not doing it for the sake of doing it. You're doing it to meet some enterprise goals, but also doing it at a tactical level and you're learning, growing and maturing that data governance organization.

Is there a risk if your data management programme is lead predominantly by technology rather than the business?

I've been in on the technology side as a developer involved with database design. If we're looking from an IT perspective, we're looking to manage the widgets, but what do the widgets mean? The only people that can tell you actually what the widgets mean are the business. The business defines what that data is and it has a use. It's collected for a business function or to answer a business question. The risk letting IT run the show is that maybe they missed some of the nuances in the data. Maybe it's not 100% classified correctly. It's like when you get into the database and you see the status indicator is the indicator of status. Well okay, that's the definition but it's not a useful one. So the business needs to really be involved because their neck is on the line for the use of the data.