HOW DATA GOVERNANCE AND A SEMANTIC LAYER SUPPORTS DATA MESH

By George Firican

George is a passionate advocate for the importance of data, a frequent conference speaker and a YouTuber, being ranked among Top 5 Global Thought Leaders and Influencers on Big Data, Digital Disruption and Top 15 on Innovation.

ATSCALE

Data are Assets - Valuable and Impactful

"Data is valuable!" "Data is an asset!" We hear this a lot.

Clive Humby, a mathematician and the co-creator of the Tesco Clubcard – the world's first supermarket loyalty card, coined the phrase, "data is the new oil", in 2006. We also hear the expressions "data is valuable" and "data is an asset" a lot.

The message that Mr. Humby and others want to convey is that a company's data is a tangible asset. In fact, data is arguably one of the most important assets that any organization has. Why? We need data for two key reasons: to help us answer key business questions and to provide feedback about our performance, customers, markets and competitors. We hear so much today about the benefits of using machine learning - but what makes machine learning so valuable is the data that it uses. Enterprises learn from data. Data is the catalyst for improving enterprise relevance and distinctiveness. Data helps accelerate productivity, efficiency, and competitive advantage.

Companies that learn from their data improve their performance. Research from Tableau confirms that 83% of CEOs want their companies to be data driven, but only 46% are achieving that goal. For those that do achieve it, the rewards are plentiful. According to a survey of more than 1,000 senior executives conducted by PwC, highly data-driven organizations are three times more likely to report significant improvements in decision-making compared to those that rely less on data.

When Performance Matters - Insights and Analytics Deliver

The benefits of using data to generate actionable insights and analytics are real and extensive across every industry and functional capability.

For example, in banking the value could be driven by improving fraud detection and real-time analysis of market data. For insurance agencies, insights and analytics help predict and mitigate risk. The logistics industry relies on insights and analytics to optimize inventories and coordinate shipments to meet demand while minimizing costs. The healthcare industry uses analytics to improve the diagnosis of illness and medical conditions in patients while also improving predictions and planning for infectious disease threats or outbreaks.

Perhaps no other industry has transformed itself by applying data, insights and analytics like the sports industry. Sports has moved from a "gut feel" culture to an analytical / statistical-based culture where most strategic and tactical decisions - selecting players, improving plays, improving fitness and training - are guided by data, insights and analytics. Why? Because in sports, performance is measurable and, ultimately, winning is what matters. The winners reaping the greatest rewards.

Delivering Actionable Insights

Data is most valuable when it's used to create actionable insights that lead to improvements in business outcomes. These benefits can be seen across all industries. For example, for banking the value could be driven from improving fraud detection and real-time analysis of market data. For insurance agencies, we see how big data aids in mitigating risks or reducing the calculation time of the value at risk. The supply chain industry relies on data to ensure the inventories as well as the shipments are optimized to reduce costs while meeting customers' demands. The healthcare industry can use data to better help diagnose illnesses and medical conditions in patients, while also enabling better predictions and planning for infectious disease threats or outbreaks. The sports industry is a favorite for anyone who watched Moneyball, as it showcases how data can be used to improve performance strategically and tactically. Even governments rely on data for crime prevention, better emergency response, and smart city initiatives that provide more personalized and efficient services for constituents.

Recent research reveals that 60% of enterprise organizations use data and analytics to drive strategy and change, improve processes, and realize cost-efficiency (MicroStrategy, 2020).

The resulting investment in big data technology reveals the scope of this transformation. According to research firm International Data Corporation (IDC), worldwide spending on big data and business analytics (BDA) solutions in 2021 was forecast to reach $215.7 billion, an increase of 10.1% over 2020. Furthermore, IDC forecasts that BDA spending will gain strength over the next five years as the global economy recovers from the COVID-19 pandemic. The compound annual growth rate (CAGR) for global BDA spending over the 2021-2025 forecast period will be 12.8%, much larger than every other category of IT spending.

Modern Data Governance - The Catalysts

We can't deliver actionable, effective, consistent insights and analytics without governing the inputs. To that end, enterprises need to understand and manage how the data sources, processes for integrating data, and the data reports are created and deployed. Increasingly, enterprises need data to understand the decisions and actions that are being taken, and the impact they have on customers and compliance.

This is why we need to deploy effective data governance.

Traditionally, data governance has been viewed as bureaucratic, controlling and restrictive activities that create constraints and slow down progress. In many cases, the processes were slow because the systems used to manage governance, i.e. managing access for users and usage, were done manually with approvals done in batch mode vs being done continuously.

With organizations migrating to cloud-based data platforms, they are demanding that data governance do the same, and move from being restrictive to enabling greater speed, scale and productivity. This is what modern data governance helps deliver. Recent research from Deloitte shows that modernizing data is among the #1 or #2 reasons for moving to the cloud, along with security and cost savings.

TOP DRIVERS FOR CLOUD MIGRATION

Percentage of respondents who ranked each category as No.1 or No.2

■ No.1   ▢ No.2

Security and data protection

| 37% | 21% |

Data modernization

| 22% | 22% |

Cost and performance of IT operations

| 15% | 17% |

Modern Data Governance - Enabling Speed, Scale and Cost Effectiveness

We can't deliver actionable, effective, consistent insights and analytics without governing the inputs. To that end, enterprises need to understand and manage how the data sources, processes for integrating data, and the data reports are created and deployed. Increasingly, enterprises need data to understand the decisions and actions that are being taken, and the impact they have on customers and compliance.

This is why we need to deploy effective data governance.

Traditionally, data governance has been viewed as bureaucratic, controlling and restrictive activities that create constraints and slow down progress. In many cases, the processes were slow because the systems used to manage governance, i.e. managing access for users and usage, were done manually with approvals done in batch mode vs being done continuously.

With organizations migrating to cloud-based data platforms, they are demanding that data governance do the same, and move from being restrictive to enabling greater speed, scale and productivity. This is what modern data governance helps deliver. Recent research from Deloitte shows that modernizing data is among the #1 or #2 reasons for moving to the cloud, along with security and cost savings.

What's driving the evolution of modern data governance? Fundamentally, we use data to answer business questions, and to answer those questions businesses need actionable insights and analytics, delivered with speed, scale, governance and cost-effectiveness.

To make progress and have an impact with data, companies will need to do the following:

- Deliver actionable insights faster via automation, self-service and a hub-and-spoke delivery model
- Achieve scale in terms of data sources, users and usage
- Manage costs via cloud-based infrastructure
- Reduced redundancy, reuse, collaboration and compute optimization

SPEED	SCALE	COST-EFFECTIVENESS	GOVERNANCE
Faster time to insights with fewer resources	More data sources, users and uses, including self-serve	Improved productivity and infrastructure utilization / optimization	Governed access, activities, usage and compliance

Actionable Insights and Analytics - Relevant, Actionable, Impactful

Why is Data Governance Important?

Before we define modern data governance, here's a list of the core questions that need to be addressed by data governance:

- What foundation do we need to have for collecting clean data, documented metadata, and categorized and classified data? We need some policies in place.
- What do we need for creating repeatable steps to clean our data, to make it consistent, to provide access to it, to secure it, and to define it? We need to establish and follow processes.
- How do we ensure consistency in our cleanliness, definitions, and categorization? We need to establish and comply with certain standards.
- Who's going to create all of these policies, processes, standards, rules and definitions? Who will approve them, who will maintain them, and who will enforce them? We need to define and assign certain roles and responsibilities.

Defining Data Governance

The answers to the core questions above help define why companies need to implement data governance.

So then, what is data governance?

Data governance is a collection of processes, roles, policies, and standards that ensure the effective and efficient use of data in enabling an organization to achieve its goals. It establishes the policies, processes, standards, roles and responsibilities that ensure the quality and security of the data used across a business or organization. Data governance also defines who can take what action, upon what data, in what situations, using what methods while following set standards and definitions.

Defining Modern Data Governance

Let's define modern data governance. Modern data governance is defined as the use of cloud-based technology and tools to govern the use of data effectively and continuously focusing on the following five key capabilities:

1. Modern Cloud-based Data Platforms
2. Data Democratization
3. Data as a Product
4. Federated, Hub-and-Spoke Delivery
5. Data Observability / Accountability

MODERN DATA GOVERNANCE – KEY CAPABILITY ELEMENTS

MODERN DATA PLATFORMS	DATA DEMOCRATIZATION	DATA AS A PRODUCT	HUB-AND SPOKE DELIVERY MODEL
Cloud-based data platforms and tools	Rapid data access and self-service enablement	Create and manage data as a product	Decentralize insights creation - centralize data management
Data Observability / Accountability - data usage, decisions, actions and compliance			

In this whitepaper we will cover in more detail the key capabilities enabling modern data governance, including a brief review of the core elements of all data governance programs. Then, we will cover the importance of using a semantic layer to deliver improved data governance for data products.

Modern Data Governance - Enabling Speed, Scale and Cost Effectiveness

What's driving the evolution of modern data governance? Fundamentally, we use data to answer business questions, and what businesses need are actionable insights and analytics to answer those business questions - delivered with speed, scale, governance and cost-effectiveness.

To make progress and have an impact with data, companies will need to do the following:

- Deliver actionable insights faster via automation, self-service and a hub-and-spoke delivery model,
- Achieve scale in terms of data sources, users and usage and
- Manage costs via cloud-based infrastructure,
- Reduced redundancy, reuse, collaboration and compute optimization will continue to power progress and impact.

MODERN DATA LANDSCAPE – KEY EVOLUTION DRIVERS

SPEED	SCALE	COST-EFFECTIVENESS	GOVERNANCE
Faster time to insights with fewer resources	More data sources, users and uses, including self-serve	Improved productivity and infrastructure utilization / optimization	Governed access, activities, usage and compliance
Actionable Insights and Analytics - Relevant, Actionable, Impactful			

Next, let's explore the current challenges, the importance of data governance, and what needs to change to modernize data governance for everyone's benefit.

The Challenge - Delivering Actionable Insights at Scale

Delivering actionable data is a process that requires many steps. Data that is actionable is often created from multiple data sources, each of which needs to be assessed, cleansed, prepared and made ready for use. The steps to create actionable insights often require as many as seven steps, which are as follows:

1. Access
2. Profile
3. Prepare
4. Integrate
5. Extract / Aggregate
6. Analyze
7. Publish / Present

Further, the data that's sourced and transformed into actionable insights must also address data governance requirements for the following:

- Quality - Consistent, accurate data delivered on time with actionable recency
- Security - Secure storage, access, transformation and usage
- Privacy - Compliance with privacy standards
- Legal Compliance - Compliance to local laws regarding data privacy and usage
- Data Usage Compliance - Compliance with standards for how data is used to prevent identification

Core Elements of Traditional Data Governance - The Basic Building Blocks

The core elements and deliverables of data governance include:

- Roles and responsibilities
- Data policies,
- Data standards
- Data processes
- Defined metadata

Roles & Responsibilities

Data governance ensures that the right people are assigned the right data responsibilities. The main responsibilities are as follows:

Data governance lead - Responsible for all aspects of defining and operating the data governance policies and supporting the multiple data domains. They are ultimately responsible for implementing the data governance program vision, promoting the role of governance, and enforcing policy, all while following data governance best practices.
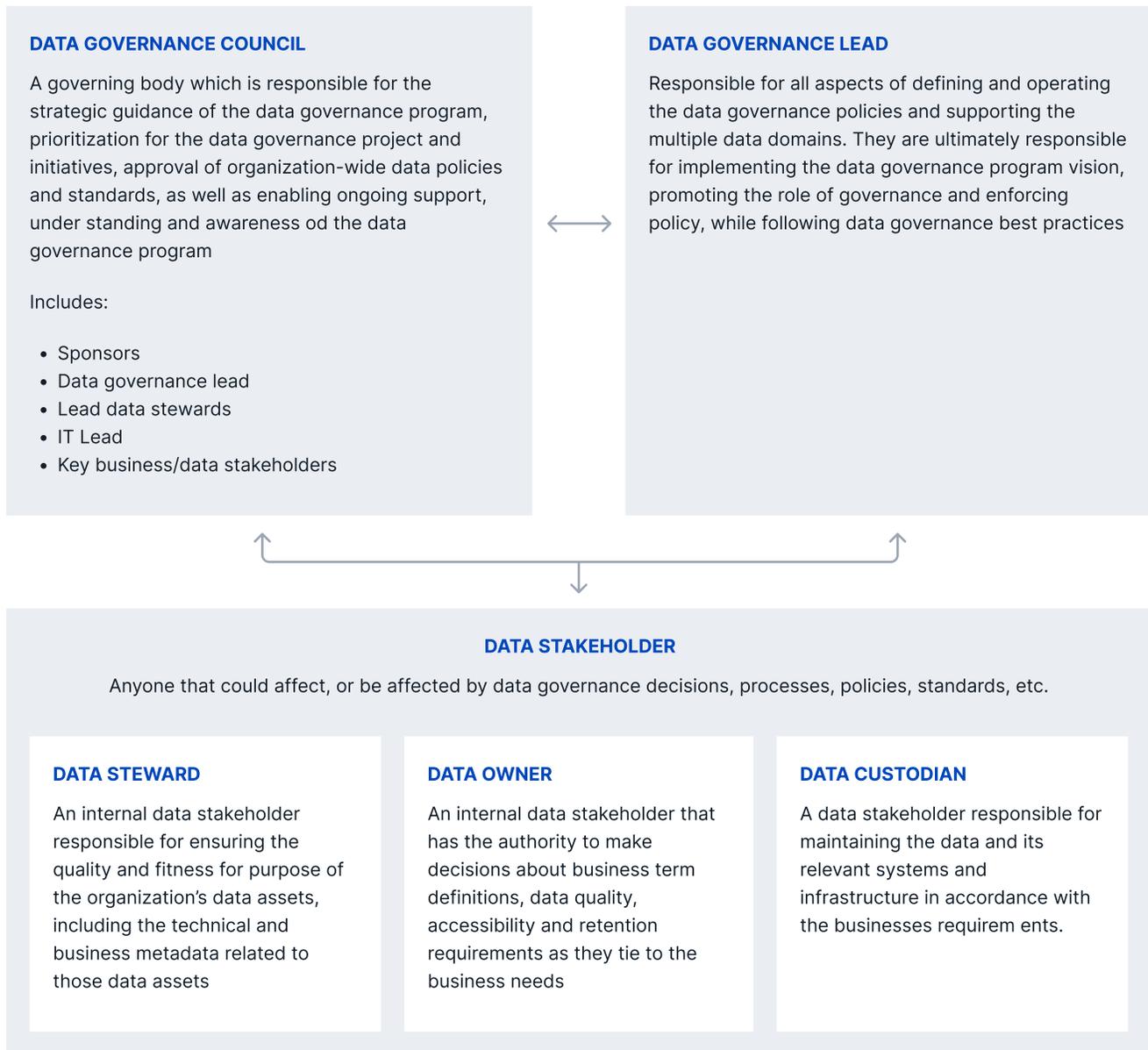
Data governance council - A governing body that is responsible for the strategic guidance of the data governance program, prioritization for the data governance projects and initiatives, approval of organization-wide data policies and standards, as well as providing ongoing support, understanding and awareness of the data governance program.

Data stakeholder – Anyone that could affect, or be affected by, data governance decisions, processes, policies, standards, etc.

Data owner – An internal data stakeholder that has the authority to make decisions about business term definitions, data quality, accessibility and retention requirements as they tie to the business needs.

Data steward – An internal data stakeholder responsible for ensuring the quality and fitness of the organization's data assets, including the technical and business metadata related to those data assets.

Data custodian – A data stakeholder responsible for maintaining the data and its relevant systems and infrastructure in accordance with the businesses requirements.

DATA GOVERNANCE COUNCIL

A governing body which is responsible for the strategic guidance of the data governance program, prioritization for the data governance project and initiatives, approval of organization-wide data policies and standards, as well as enabling ongoing support, under standing and awareness od the data governance program

Includes:

- Sponsors
- Data governance lead
- Lead data stewards
- IT Lead
- Key business/data stakeholders

DATA GOVERNANCE LEAD

Responsible for all aspects of defining and operating the data governance policies and supporting the multiple data domains. They are ultimately responsible for implementing the data governance program vision, promoting the role of governance and enforcing policy, while following data governance best practices

DATA STAKEHOLDER

Anyone that could affect, or be affected by data governance decisions, processes, policies, standards, etc.

DATA STEWARD

An internal data stakeholder responsible for ensuring the quality and fitness for purpose of the organization's data assets, including the technical and business metadata related to those data assets

DATA OWNER

An internal data stakeholder that has the authority to make decisions about business term definitions, data quality, accessibility and retention requirements as they tie to the business needs

DATA CUSTODIAN

A data stakeholder responsible for maintaining the data and its relevant systems and infrastructure in accordance with the businesses requirem ents.

A couple of other notable roles that are not necessarily tied to data governance:

- **Insights Creator** – Any user, user interface, automation, service or device that creates or collects data relevant to a business, and turns it into actionable insights and analytics
- **Insights Consumer** - Any user, application, or system that uses data collected or produced by another user or system or is stored in a data repository

In order for the data insights creator to effectively communicate to the data consumer, we also need the following elements of data governance: data policies, standards, procedures, and defined metadata.

Data Policies

A policy is a statement of a selected course of action and high-level description of desired behavior to achieve a set of goals. Data governance typically defines policies related to privacy, security, access, usage, analytics (algorithms), compliance, and quality. Guidelines also cover the previously discussed roles and responsibilities of those implementing policies and compliance measures.

In the end, the purpose of these policies is to ensure that organizations are able to maintain and secure high-quality data. Governance data policies form the base of your larger data governance strategy and enable you to clearly define how data governance is carried out.

A few common areas covered by data governance policies are:

- **Data quality** – Ensuring data is correct, consistent, and free of "noise" that might impeded usage and analysis.
- **Data accessibility and availability** – Ensuring that data is available, accessible and easy to consume by the business functions that require it.
- **Data usability** – Ensuring that data is clearly structured, documented and labeled, that it enables easy search and retrieval, and that it is compatible with tools used by business users.
- **Data integrity** – Ensuring data retains its essential qualities even as it is stored, converted, transferred, and viewed across different platforms.
- **Data security** – Ensuring data is classified according to its sensitivity and defining processes for safeguarding information and preventing data loss and leakage.

Data Standards

A data standard is an agreement on representation, format, definition, structure, tagging, transmission, manipulation, use, and management of data. We need standards to create, share, integrate, and use data. Standards aid in data cleansing and data transformation, but they also support data policies and adherence to them. Data standards range from anything and everything on how to record master data, reference data, transactional data, and analytics data such as measures/metrics.

Data Processes

A lot of the data governance processes will have the purpose of carrying out data governance policies and ensuring standards are followed and met. The remaining ones will ensure that the activities of data governance are carried out. The typical processes address the planning, designing, managing, operating, and sustaining of:

- Regulatory compliance
- Standards and policies
- Master data support
- Data analytics and ML/AI
- Security and privacy
- Enterprise data model
- Technical and business metadata
- Data governance program's plan and management

Defined Metadata

There are two types of metadata for which data governance sets the roles and responsibilities, processes, and standards to create and maintain:

1. **Business metadata** – i.e. the business concepts for an organization or industry. Business metadata defines things like what a customer is, sale conversion rate, credit, and funds received. It is mostly information authored and controlled by data stewards.
2. **Technical metadata** – i.e. the specifications about a field in a database. These specifications typically include data type, allowed values, default values, constraints, relations to other data elements, meaning and purpose. This information is mainly handled by data custodians.

Both of these types of metadata, if properly defined and maintained, provide the necessary information and context for data stakeholders and data consumers to use the data without making incorrect assumptions.

The Main Reasons for Implementing Data Governance

Understanding why your organization is implementing data governance is crucial for several reasons. But most importantly you need to know the 'why' in order to guide your data governance strategy and change management adoption strategy within your organization. Moreover, it will provide you with a direction on what should be tackled first. After all, you need to start somewhere and you can't just tackle everything at once.

While not every organization will implement data governance for exactly the same reasons, there are some general 'whys' that can provide a prioritization guide on what areas of the data and the data governance program you should focus on first. These are:

Scalability

The immense potential of data is well recognized by organizations across all industries. But for any data initiatives to succeed and scale, the data must first be accessible and available, compliant, defined and understood, and of high quality. This is where data governance comes in to set the right foundations for scalable data initiatives. Because in order to scale up one needs to ensure that:

- Data is not siloed, but made available to the right people for the right usage
- Data is put into context, described, and understood through common business verbiage and data dictionary
- Data quality is at a level needed to meet the business requirements
- There are clear rules and roles for data creation/acquisition, maintenance, usage/ dissemination, and archival/destruction
- Mechanisms are in place for managing the data lifecycle
- The same data can be used for different purposes by different teams

Access and availability

Data governance helps with the mapping, lineage and organization of a company's data and ensures organizations have more visibility and control over the data being gathered across it. This allows data stakeholders, data producers and data consumers to share insights and eliminate data silos. In turn, it often establishes a consistent and complete single source of truth of critical data and metrics that business stakeholders can agree upon to make better cooperative decisions.

Proper governance enables collaboration across departments, fostering broader insights, fueling better decisions, and promoting a more data-driven organization. In the end, data needs to be accessed and available by the right people at the right time.

Compliance

Data governance allows organizations to have clear control processes over their data to align with pre-set business rules around regulatory compliance, data privacy and security. Data governance allows organizations to ensure they have policies, standards, and processes in place to identify and control data covered under specific regulations and assure that all relevant compliance regulations are met in all your organization's data practices. This makes it easier to stay compliant - and avoid big fines.

Actionability

Data governance is the first step in creating an organization which is driven to make decisions that are based on undisputable data. And it's a simple fact that organizations that take action based on their data are more likely to achieve growth than organizations that continue to operate in data silos. In order to be actionable, data first needs to comply with data quality requirements and second it needs to be defined and understood.

In many organizations the data producers are actually not fully aware of the data quality requirements of the data consumers. Or they're not aware of the different business processes and data products such as: data visualizations and dashboards, data warehouse, recommender systems, unstructured data classifications, etc. A data governance program factors in those clearly defined roles and responsibilities to allow data stakeholders and in particular, data stewards, to measure, monitor, and improve the data quality dimensions that are relevant to their line of business.

Data governance also allows users to define and understand the data, its context, its usage, while allowing them to better troubleshoot and prevent data issues. Data governance helps create, establish, and socialize common verbiage around metrics and datasets that provides all stakeholders with a common data language and consistent terminology that's easily understood across the organization.

Without quality data that's defined and understood, organizations can't drive the correct actions out of their data. Users will make poor decisions because they are either based on bad quality data or incorrect assumptions derived from a lack of common data definitions.

ACCESS & ACCESSIBILITY

Shared data	Data is no longer siloed	Single source of truth

ACTIONABILITY

Incrreased confidence in data quality	Common data dictionary & business verbiage

COMPLIANCE

Clear rules and data processes	Data policy alignment to regulations

BUSINESS OUTCOMES

Reduced costs

Actionable results
Increased efficiency

Data driven decisions

It, data, business teams mode agile

Higher compliance and reduced risks

Traditional vs. Modern Data Governance

Let's refer to traditional data governance as data governance practices done in a pre-cloud environment and modern data governance as those practices done in a post-cloud environment.

Traditional data governance

Due to the gargantuan efforts of breaking down technology and people silos and tackling data management and data governance organization-wide, data governance used to be focused on either one of the following two:

1. A core system – usually an Enterprise Resource Planning (ERP) or a Customer Relationship Management (CRM) system
2. A particular business unit – such as marketing, procurement, finance, privacy and security

As noted before, starting a data governance program can be challenging, so setting the initial focus on a core system or a particular business unit can get things off the ground quicker and easier. Afterall, the scope is bound by the data housed within that core system or pertaining to that particular business unit.

If the focus was set on a core system, some of the data governance definitions, standards, and processes were simply inherited from it, or in better cases served as a starting point. The downside of this pathway was that data governance was mainly led by IT with a lot of the requirements dictated by the technical limitations of that core system.

The business usually lacked sufficient representation as there would have been data stakeholders that were not also system stakeholders. Therefore, these data stakeholders, although key in helping establish an enterprise-wide data governance program, were often omitted from consultations and decisions. Moreover, these omitted data stakeholders often included the data analysts and data scientists as their work would not necessarily yield data that had to be fed back into the system. This meant that their work often resided outside the boundaries and the scope of a data governance program, which started with a focus on a core system.

If the focus was on a particular business unit, there were some benefits coming out of already knowing the stakeholders and having the relationships within the business unit to gain support and influence adoption. This was much easier than having a business unit's data governance office reach into the unknown of other units or lines of business. The good thing was that the data from multiple systems could have been in scope, if those systems had the business unit as an owner or a key stakeholder. Furthermore, if data analysts and data scientists created data products for that business unit, those would also fall under the data governance umbrella.

The common challenges with both choices on how to focus this traditional data governance program were:

- Organization-wide business requirements were not captured
- Organization-wide level data needs would not be met
- Scaling beyond the core system or the business unit was difficult and costly
- Conducting redundant and sometimes conflicting efforts if different data governance initiatives were started around separate systems and business units

Typically, there are three operating models of traditional data governance: decentralized, centralized, and federated.

Decentralized Data Governance

In a decentralized model we would find multiple, concurrent data governance initiatives and groups that aim to govern the data value creation activities addressed by siloed teams belonging to different business units.

The business usually lacked sufficient representation as there would have been data stakeholders that were not also system stakeholders. Therefore, these data stakeholders, although key in helping establish an enterprise-wide data governance program, were often omitted from consultations and decisions. Moreover, these omitted data stakeholders often included the data analysts and data scientists as their work would not necessarily yield data that had to be fed back into the system. This meant that their work often resided outside the boundaries and the scope of a data governance program, which started with a focus on a core system.

If the focus was on a particular business unit, there were some benefits coming out of already knowing the stakeholders and having the relationships within the business unit to gain support and influence adoption. This was much easier than having a business unit's data governance office reach into the unknown of other units or lines of business. The good thing was that the data from multiple systems could have been in scope, if those systems had the business unit as an owner or a key stakeholder. Furthermore, if data analysts and data scientists created data products for that business unit, those would also fall under the data governance umbrella.

The common challenges with both choices on how to focus this traditional data governance program were:

Organization-wide business requirements were not captured

Organization-wide level data needs would not be met

Scaling beyond the core system or the business unit was difficult and costly

Conducting redundant and sometimes conflicting efforts if different data governance initiatives were started around separate systems and business units

Typically, there are three operating models of traditional data governance: decentralized, centralized, and federated.

Decentralized Data Governance

In a decentralized model we would find multiple, concurrent data governance initiatives and groups that aim to govern the data value creation activities addressed by siloed teams belonging to different business units.

The result may yield value for the given unit, but not necessarily for the broader organization, which will likely be receiving conflicting reports of its master data, redundant efforts creating the same metrics multiple times, and inconsistencies on how data was defined, created, maintained, and consumed.

Centralized Data Governance

Typically, as the data governance program would try to scale up and include newer core systems or other business units, a common data governance program would aim to emerge, led by a single data governance council. Its membership included representation of all those with their data in scope. The data, analytics, and data value creation were no longer handled in siloed teams, but they were mainly done centrally. Consensus, accuracy, and consistency were achieved, but because data, analytics, and data value creation were practically gated by this central team, progress was slow and often impeded innovation and the ability to act quickly to changing business needs.

Federated Data Governance

Federated data governance aimed to be the best of both worlds. It still provided a centralized structure that oversaw the enterprise-level data, analytics and data value creation while allowing the flexibility and self-governance for anything that was unique to each business unit.

Modern Data Governance - Data Democratized and Federalized

All of these operating models were still plagued by the traditional way of setting the focus of data governance. A new way had to be established in order to enable an organization-wide data governance program to emerge. One that enables the delivery of actionable insights at scale.

Enter modern data governance

Defining Modern Data Governance

Modern Data Governance is defined as the use of cloud-based technology and tools to govern the use of data effectively and continuously focus on the following five key capabilities:

1. Modern Cloud-based Data Platforms
2. Data Democratization
3. Data as a Product
4. Federalized, Hub-and-Spoke Delivery
5. Data Observability / Accountability

MODERN DATA PLATFORMS	DATA DEMOCRATIZATION	DATA AS A PRODUCT	HUB-AND SPOKE DELIVERY MODEL
Cloud-based data platforms and tools	Rapid data access and self-service enablement	Create and manage data as a product	Decentralize insights creation - centralize data management
Data Observability / Accountability - data usage, decisions, actions and compliance			

Modern Data Platforms and the Cloud

Data cloud environments and the tools that power them are starting to make it much easier to integrate data from multiple systems, and start handling and consuming the data from the point of view of the entire organization and not just isolated business units or solitary systems.

The cloud architecture enables organizations to open the flood gates of their data and make it cheaper and easier to start democratizing the data. Suddenly, business users with little knowledge of the data space could leverage modern data platforms to get close-to-real-time insights about the business without always having to rely on the Business Intelligence team every time they needed an answer. The data science team could start creating their AI products and yielding new measures/metrics without having to wait for new infrastructure to be in place. The data governance team could set their scope on the cloud and not have to worry as much about the source systems that data came in; they could focus more on the needs of the data consumers.

Data Democratization

Now that companies are able to store and provide access to data centrally in the cloud, insights creators and consumers are increasingly demanding that data be democratized. There is a common misperception that data democratization is a synonym for data access. But in reality, it is much more than that. Data democratization is a process that enables all data and business stakeholders within an organization to work with the data they need in order to deliver actionable insights at scale and make data-informed decisions. Here are the major elements and differences between a modern data governance program within a data democratization environment and a traditional one:

TRADITIONAL	MODERN DATA GOVERNANCE - DEMOCRATIZED AND FEDERALIZED
Data access provided according to each project need	Data access provided according to business role
Only technically skilled people can work with data	Any stakeholder can work with data
Analytical tools not designed for product teams	Analytical tools designed for product teams
Know-how and context of the data gatekept by data experts	The necessary metadata and context is available for all data consumers
New data products are created by dedicated BI and analytics teams	New data products can be created by any stakeholder
New data products are mainly available to their creators	New data products are available to anyone that needs them
Complex value creation model bottlenecked by central IT and data teams	More agile value creation through self-service analytics

With these changes, data democratization yields both caution and optimism for the potential outcomes. Those who are both supportive or weary of this new world of data democratization identify some common opportunities and challenges.

Data democratization opportunities

- Empowering employees by providing quicker access to data that is both defined and understood
- Switching from a siloed approach on accountability over the organization's data to a more all-encompassing, centralized one
- Gaining more insights by having a diverse set of teams collaborate on the same measures/ metrics and data

Data democratization challenges

- More users could now have access to more data, creating risks
- Duplication of efforts if different users prepare the same measures/metrics without them knowing about it
- More prevalent misuse of data by misinterpreting it if there was a lack of context and definitions

Modern Data Governance Delivered via a Federalized, Hub-and-Spoke Delivery

To alleviate the challenges and capitalize on the data democratization opportunities, a modern data governance has to come into play. The best practice is therefore to govern this data by data domain (also called subject area).

By data domain, I'm referring to "a logical grouping of items of interest to the organization, or areas of interest within the organization". You can think of data domains as high-level categories of data for the purpose of assigning accountability and responsibility for the data. Such categories or subject areas such as:

- Customer
- Product
- Service
- Location
- Vendor/ Supplier

The benefit of governing data this way is that you have organization-wide representation from any data stakeholder handling any data, analytics, and data value creation within each domain. The needs of the whole organization as well as each business unit are now considered and evaluated, and the overall roles and responsibilities, policies, definitions, standards, processes tend to be agnostic from the systems in which the data reside.

This method brings even more benefits when coupled with a federalized, hub-and-spoke delivery of data governance. For each data domain, all data stakeholders come together to define any policies, processes, standards, and metrics related to it, but these are created in the central data governance office, or the hub.

The artifacts and information then flows to its spokes for adoption and implementation. The spokes could be represented by individual business units and departments, but even geographically different offices and sites. The spokes differ greatly from one another from a market, product, or functional need perspective, but they are free to create their own addendums or specific artifacts that only concern their area and do not impact others. The hub-and-spoke model allows organizations to analyze and process more forms of data, for different business needs, while still offering a structured delivery of data governance.

But the reality is that even though governing data with a focus on data domains through a hub-and-spoke model ensures the enablement of a data-driven organization, data governance is not keeping up the pace with the data product creation. There's still a missing piece.

Data governance hand in hand with data product creation

Modern data consumers and producers tend to have the same data challenges they've always had:

- Ensuring they are working with trusted, accurate, reliable data
- Understanding the data and its context
- Having the data readily available and accessible

Only now, these challenges are at a higher magnitude as there are more stakeholders who need to work with data and produce data products. In turn, these data products also need to be thrown under the same data governance umbrella so that it can be used as quickly as possible by others.

The ideal state of data democratization that organizations are trying to achieve is to be able to govern the data centrally so that all the benefits of data governance come to fruition, but at the same time not create a slow queue for new data products to be created or once they are created to be allowed to be in use.

Let's step back for a second and remind ourselves that data democratization comes with a big cultural shift when it comes to working with data.

Deriving insight, both hindsight and foresight out of our data, is no longer gate-kept by IT and specialized data teams such as BI and data analytics. Instead these capabilities are made available to all relevant stakeholders throughout the organization. Within a democratized data environment, analytics and reporting is embedded throughout the organization from top to bottom. One can immediately see the value in this as those who know the business well should be provided with a way to also know and use the data to ask and answer questions relevant to their business units.

Unfortunately, all these benefits also come with challenges. The truth is that many companies utilize a manual approach to data governance for creating standards and processes to vet data, compare reports, map upstream and downstream dependencies, and so on. So, any new data products like new measures/metrics have to wait in line, sometimes for too long, before the data governance team can get to them.

Modern data governance ensures data traceability and data quality for these new data products as well. It also makes data and business stakeholders more efficient at finding and understanding their data so that they can ask the right questions as well as answer them properly. Modern data governance also introduces the ability to reuse data and data processes, thus reducing repeated and redundant work. Afterall, a more productive stakeholder maximizes the income generation of potential data.

Modern data governance can't be overly restrictive to the point of allowing down progress. It needs to be more flexible and agile and quickly adapt to the increased volume, variety, and velocity of data and data products that organizations consume and create. Modern data governance needs to expand its scope to the entire value creation chain.

So, what's the solution? What's the missing piece to unlock the full potential of modern data governance?

Our modern data governance needs the help of a semantic layer.

Semantic Layer Enables the Full Potential of Modern Data Governance

A Semantic Layer is a business representation of data that helps data stakeholders and, in particular, data consumers access data using common business terms. A Semantic Layer maps business data into familiar terms to offer a unified, consolidated view of data across the organization, ideally also grouped by data domain. In simple words, the Semantic Layer provides the context and information needed for actionable analytics.

The Semantic Layer needs to offer:

- **A common understanding of data** - Data consumers, irrespective of their data or business knowledge, need to find the right data, understand its technical and business metadata, and gain the context to use it correctly.

- **Consistency in data usage** – As organizations continue relying on different skills and teams to perform different BI and data analytics functions and implement different applications with the same data, consistency in the data usage is critical.
- **Decentralized data value creation** – As data products get created, things like new metrics/ measures, and AI results should be quickly instantiated for available access by other data consumers such as the BI team.
- **Data availability, auditable access and data lineage** – As data is made available to more and more stakeholders, its access needs to be controlled in relation to the type of data, business roles, and business use. As data gets transformed into new data products, and consumed and used in different mediums and applications, the ability to trace that lineage will remain instrumental.

Through this type of a Semantic Layer, a federated data governance operating model would function best as it would offer a centralized data governance function for enterprise-level data, while offering a decentralized governance for data product creation.

The data governance team could still exert control over the Semantic Layer by requiring any user on-boarded onto the Semantic Layer to be trained on how to contribute to and maintain the roles and responsibilities, policies, standards, processes, and metadata related to the available data.

Semantic Layer Rising

Over the past two years, there has been a tremendous resurgence among enterprises in using the Semantic Layer. This traces to their recent experience migrating to modern data platforms. Enterprises now have the need to improve speed, scale and cost savings for AI and BI and are able to generate actionable insight from newly available data sources for many new users and use cases.
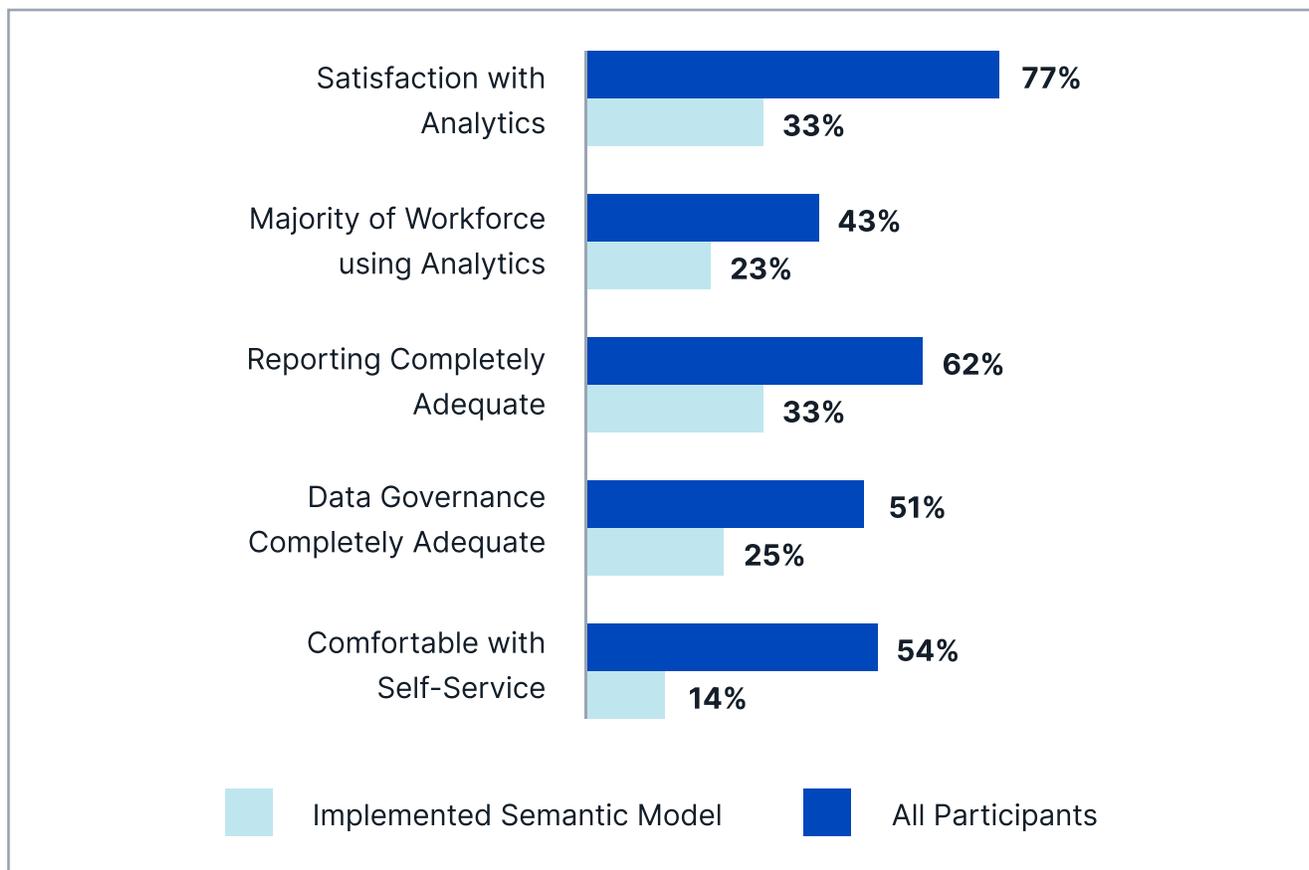
The good news is that recent research affirms the value of using a semantic layer. Companies realize the promise of successful, impactful data and analytics programs using a semantic layer, including to deliver effective data governance - and stand in stark contrast to those companies that don't use a semantic layer.

According to recent survey of 300 respondents from Ventana Research:

Organizations that have successfully implemented a semantic model/layer:

- Are significantly more satisfied with analytics (77% compared with 33% of total respondents)
- Have more of the workforce engaged in analytics (43% compared with 23% have more than one-half the workforce using analytics)
- Find analytics capabilities completely adequate (62% vs. 33% of total respondents)
- Say data governance capabilities are completely adequate (51% vs. 25% of total respondents)
- Are more comfortable with self-service: (54% very comfortable vs. 14% of total respondents)

VALUE OF SEMANTIC MODELS

Satisfaction with Analytics — 77%, 33%
Majority of Workforce using Analytics — 43%, 23%
Reporting Completely Adequate — 62%, 33%
Data Governance Completely Adequate — 51%, 25%
Comfortable with Self-Service — 54%, 14%

Implemented Semantic Model   All Participants

George is a passionate advocate for the importance of data, a frequent conference speaker and a YouTuber, being ranked among Top 5 Global Thought Leaders and Influencers on Big Data, Digital Disruption and Top 15 on Innovation.

His innovative approach to addressing data challenges received international recognition through award-winning program and project implementations in data governance, data quality, business intelligence and data analytics. He advises organizations on how to treat data as an asset, and he shares his practical takeaways on social media and various industry sites and publications.

George has been a data professional for more than ten years. One of George's passions is to create informative, practical and engaging educational content to share with individuals such as yourself, and help organizations get more visibility on social media. George is the proud founder of LightsOnData.com and its YouTube channel and is a co-host of the Lights On Data Show.

George Firican, Founder of LightsOnData