MAKING THE JOURNEY INTO DATA GOVERNANCE

IRENA A. ZADONSKY
PERSONAL REFLECTIONS

June 2022

My path to DG was non-linear. Stared as a programmer. Over the years, moved to data management, data and enterprise architecture and D&A strategy
DG was the common denominator
Most of my work - in the US Federal sector

Disclaimer: The views and opinions expressed in this presentation do not reflect the views or positions of any of my current or past employers. They reflect only my personal perspective on DG

Why is Data Governance Important

Computational Governance is the New Black

3

Some organizations have been in the cloud for a while. Others are starting to plan their move to the cloud. **Computational Governance is the talk of the town.**

Data democratization removes the gatekeepers to data. It's a new frontier, wild west.

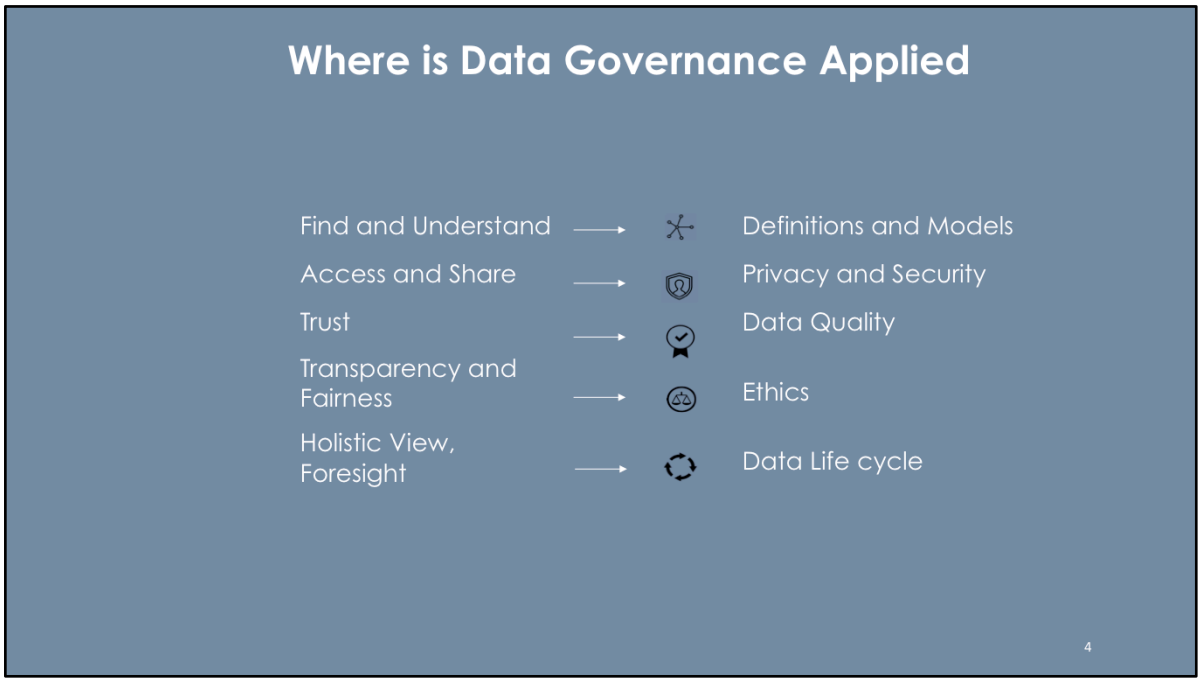
The authority for data is getting more distributed. Organizations need guardrails to control the cost and risk of moving to cloud and democratizing data and analytics.

The data rules and controls need to be automated to avoid bottlenecks

Cloud or not. Data is everywhere, Data Governance is not. People are looking to data governance to help answer questions like:

- Where can I find the data I need? Where did the data come from? How can get access to data?
- Can I trust the data? Who is using the data? Who are the experts in this data domain/dataset?
- Can I use it? What are the permitted uses and rules for sharing for data? (some purchased data is subject to rules set by the vendor)

Key decisions about data in multi-cloud environment: where (which domains) should the data be



Key DG focus areas are defined by issues that come up again and again-whenever you deal with data

1) The basic need – where do I **find the data that I need and figure out how to use it?** - enabled by **Definitions and Models**

- Metadata, lineage, glossaries, taxonomies, classification and categorization of data - Every time you capture, change, share, use data – you create metadata. Tracking all that allows you to understand what org is doing w/ data to support automation of data preparation and integration components

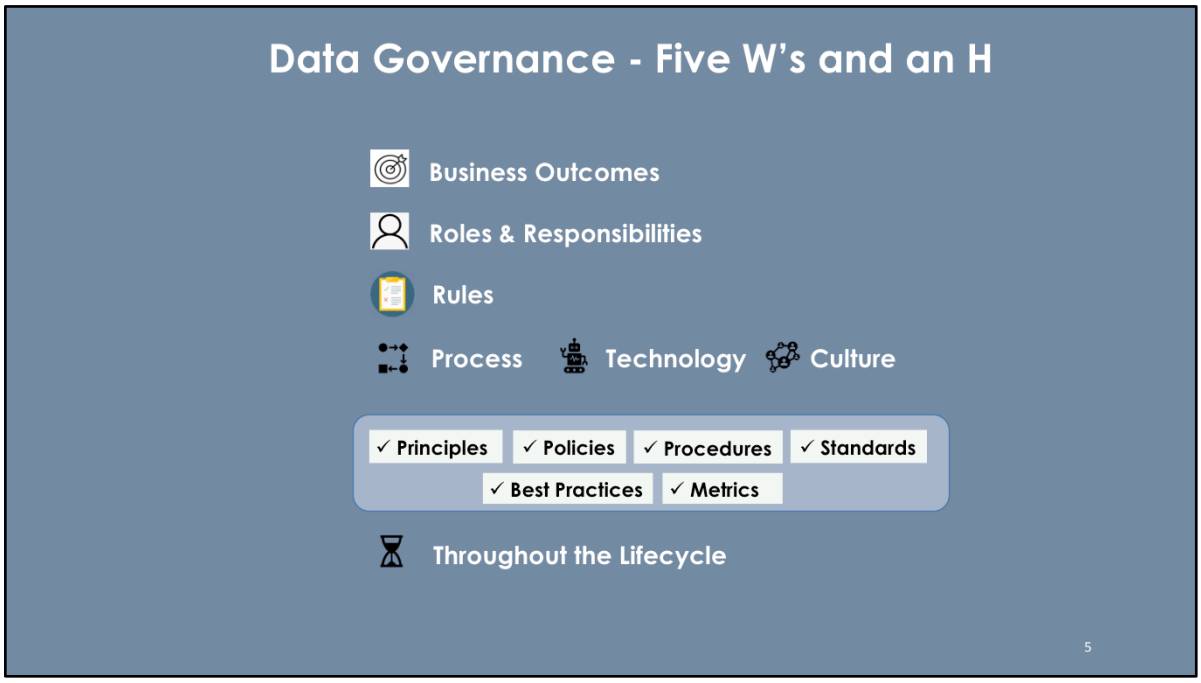
- Data Architecture, data flow -backbone of discoverability, interoperability and analysis

2) How do I get a**ccess and how can I share data and do it safely?** – need to define **Privacy and Security** rules

- Who has access to what, how does access change when data is merged/aggregated?

3) can I **trust this data?** – need **Data Quality** : defining data quality – which data elements, rules, thresholds

4)To **reduce I bias** in algorithms used with AI/ML and automation– that's where ethics rules are needed



DG defines WHO decides WHAT and WHEN about DATA in each phase of the data lifecycle to maximize value of data and support data driven decision making. DG is about defining **ACCOUNTABILITY** for data

What **business outcome** this initiative or project seeks to achieve? (Different drivers for Commercial vs Government)

What are the data related **roles and responsibilities** for this initiative or project? Are the **right people** making decisions/involved (RACI) : authority, aptitude, wiliness – time+ability to engage across business and IT

What are the **rules and requirements** for defining the needs for data, categorization, access, sharing, protection, quality and proper use?

Principles, policies, and procedures are mechanisms to build these rules into processes and technology

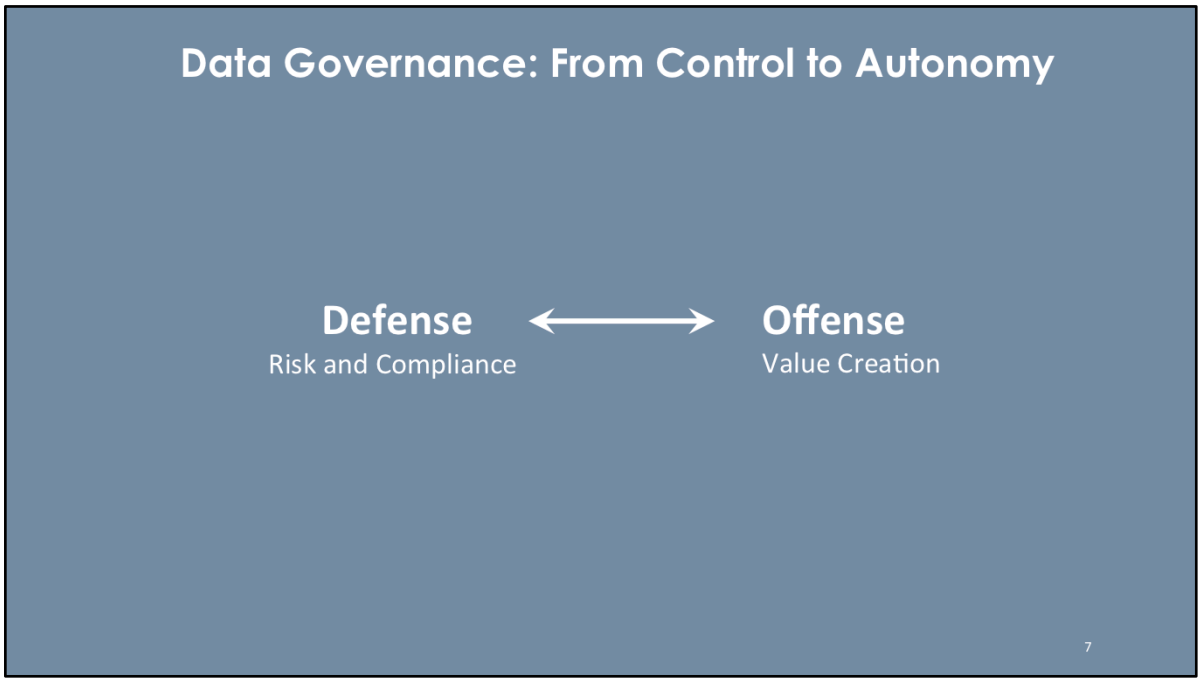
Use of standards and best practices is important to improve consistency of managing and using data across the organization. Metrics are key to measure the progress of DG efforts

Poll

Which DG aspects are most critical in your organization?

1 Roles and Responsibilities
2 Definitions and Models (Architecture and Metadata)
3 Data Privacy and Security
4 Data Quality
6 Ethics
7 Data Lifecycle
8 Other

6

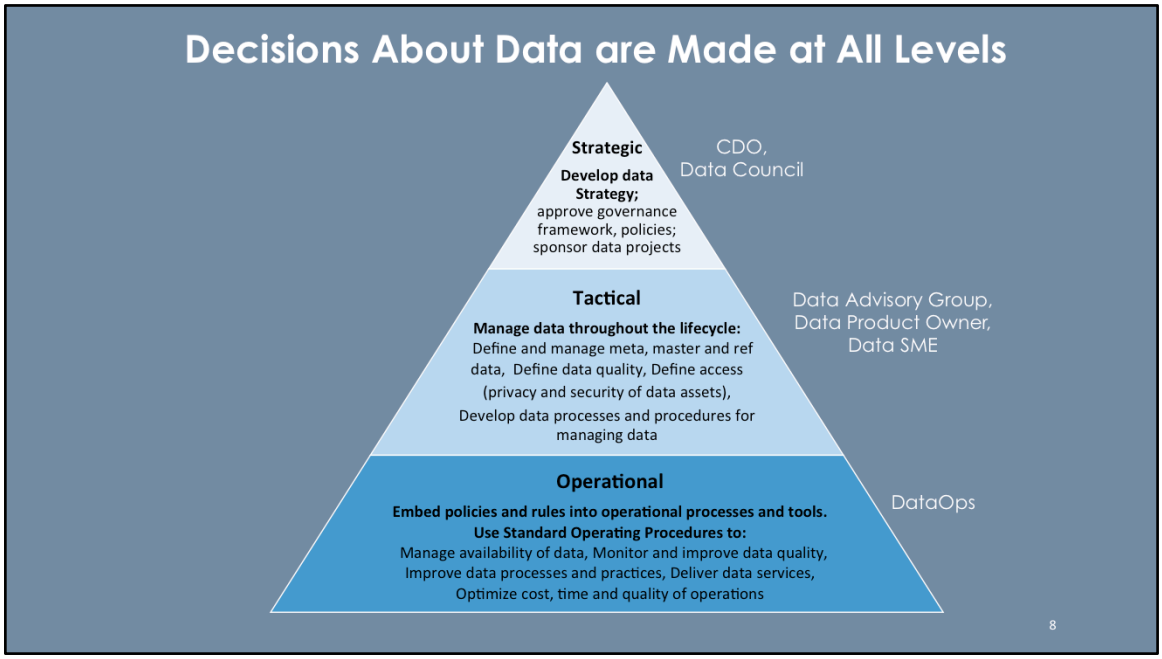


DG emerged as a discipline to meet regulatory requirements with control-oriented approach (gdpr in EU or ccpa in CA). Risk and Compliance will continue to be one of the key drivers in banking, insurance, also MDM, SOR and DM for specific systems with traditional focus on processes. That's the Defense.

Emmerging – driven by the cloud and need to be data driven - is the Offence - value creation, ROI, new data and novel use of data - social media, IoT, blockchain, AI/ML. Change drives need for exploration, POCs, more agility- move to Autonomous (ppl and things) – algorithmic data governance decisions

Move from the traditional model of Command and Control to Collaborate/Communicate enables and accelerates innovation and calls for minimum guardrails to safeguard exploration and enable automation

One of the biggest misconceptions about DG – along w/ equating it with DM - is limiting its scope to operations. Data decisions happen at all levels, incl. operations.



Data decisions are made at all levels – Need to orchestrate the flow between the levels and align accountability for data decisions

Are the rules for data governance supporting strategic objectives?
Do the stakeholders clearly understand the principles used for their decision making?
Start with FAIR + trusted, secure, etc.
Are these principles used to guide data decisions?

Are the right people (knowledge, ability, authority) involved at each level of the decision making?
Is there clarity and accountability for these decisions?

How are these decisions enforced?

What could happen if you don't have it

9

To some extent You will always have a few cats to chase, regardless whether you have DG or not! But Having DG allows you to avoid the complete chaos

When roles and rules are not clear - Decisions are made by the seat of someone's pants, most of the time…

Results into

Siloes -> Duplication of work –> Everyone is doing everything
Data breaches
Poor data quality -> Poor business decisions/Conflicting decisions
Noncompliance
Impact on strategic planning, resource allocation…
Lack of insights…

How to get started

Think Big, Start Small, Scale Fast -Jim Carroll

10

How do you get started?

Identify key stakeholders (men, squirrels, dogs here) What is important to them(where did I put that bone? Who touched my acorns? Can I dig here?)

Agree on the basics: guiding principles and scope. What **business outcomes** each of them wants to achieve with DG.

DG is change management! How will you manage political process of making this change happen? who can champion / support you /who is on the fence/ hostile

Like w/ any CM – develop a strategy for engaging w/ each stakeholder group

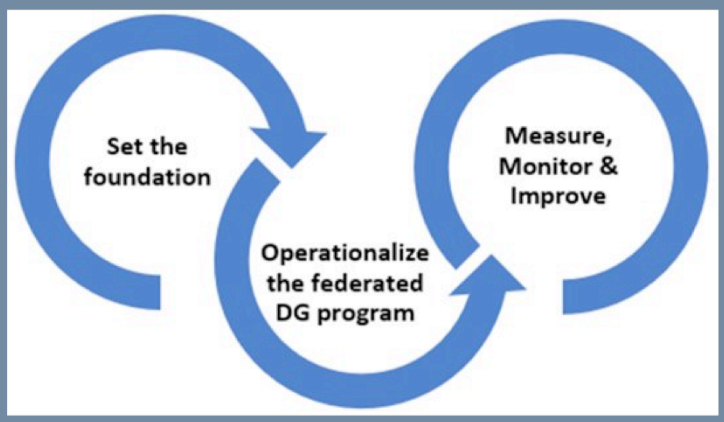
Secure business sponsorship, executive level sponsorship
Adaptive, tailored approach is key

Approach

Participatory
Identify and engage stakeholders and partners

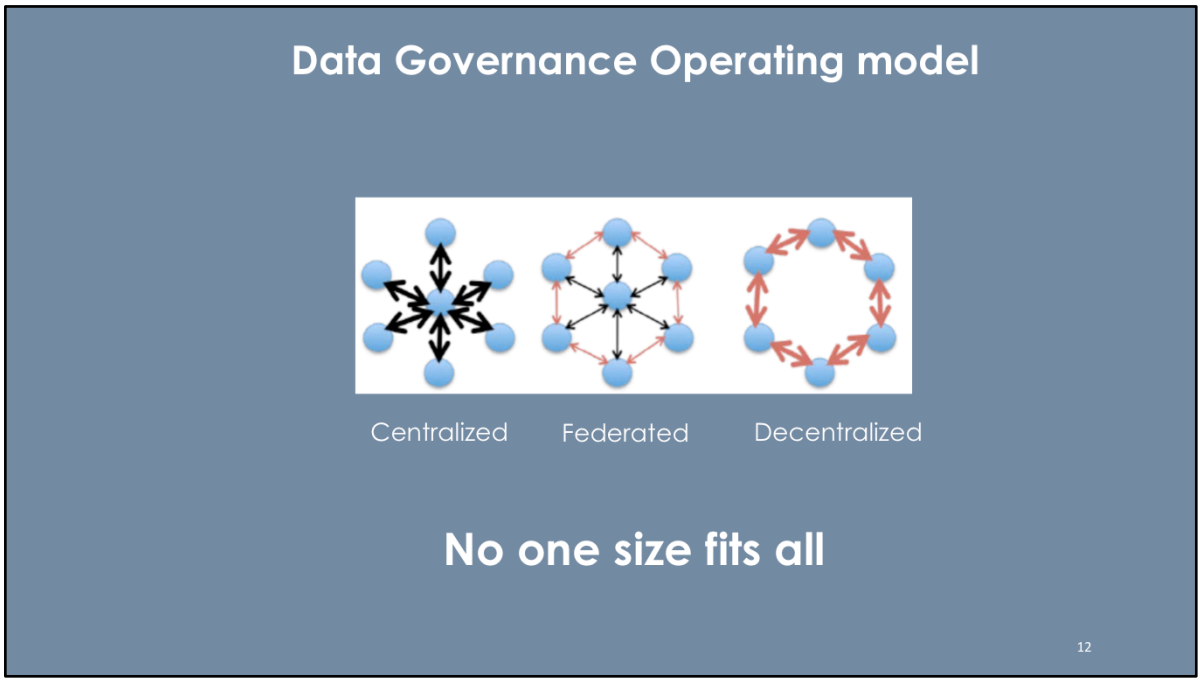
Adaptive
Align with divisional data strategies and priorities

Iterative
Optimize and Repeat



11

Engage your stakeholders to agree upon a common vision for DG - co-create the vision to secure buy-in
Adjust based on the feedback
Establish metrics and monitoring and Iterate through small experiments in partnership with your stakeholders



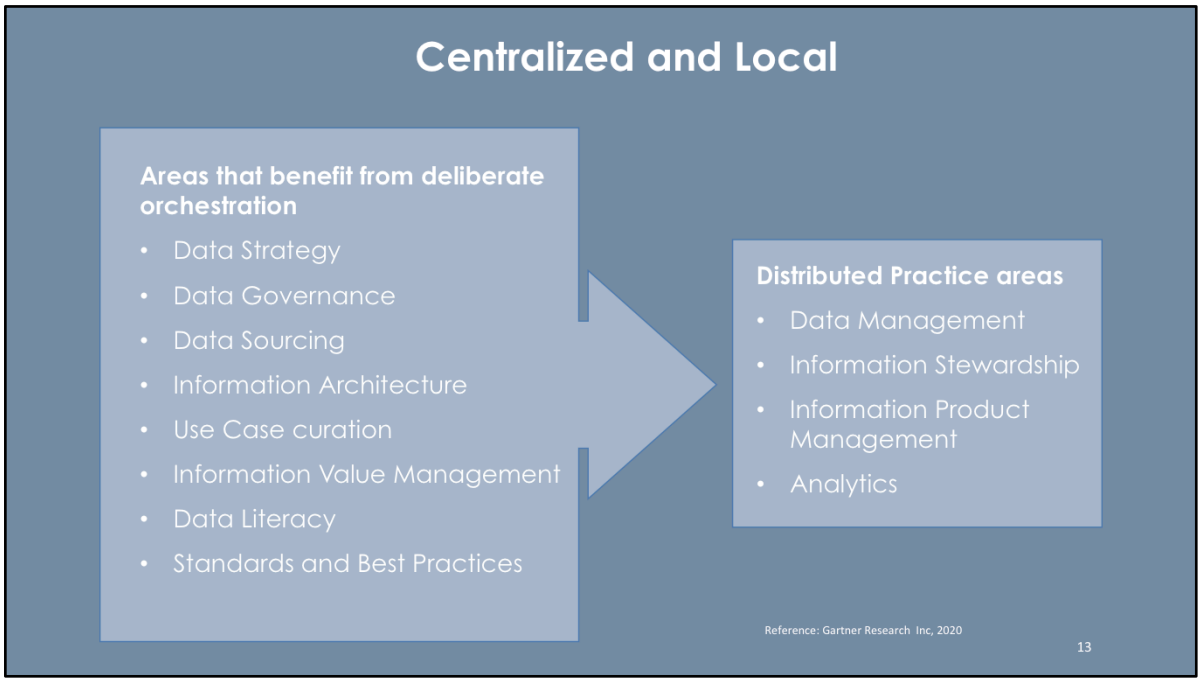
No one size fits all – **Governance needs to be adaptive**
Tailor to the organization, its structure, culture and resources. Centralized – bottleneck for complex distributed organizations. Decentralized – chaos -> pendulum

Use adaptive governance to distribute decision making across teams (project, product) and functions (services) for planning and delivery

Define the boundaries between central unit (CDO, CDAO, CoE) and business units, iterate using small well-defined experiments (POC) and adapt/ **focus on value to customer**

Monitor for culture rejection signs: lack of empowerment, do folks get back to control based decisions? What are the obstacles that need to be addressed? W rules are needed?

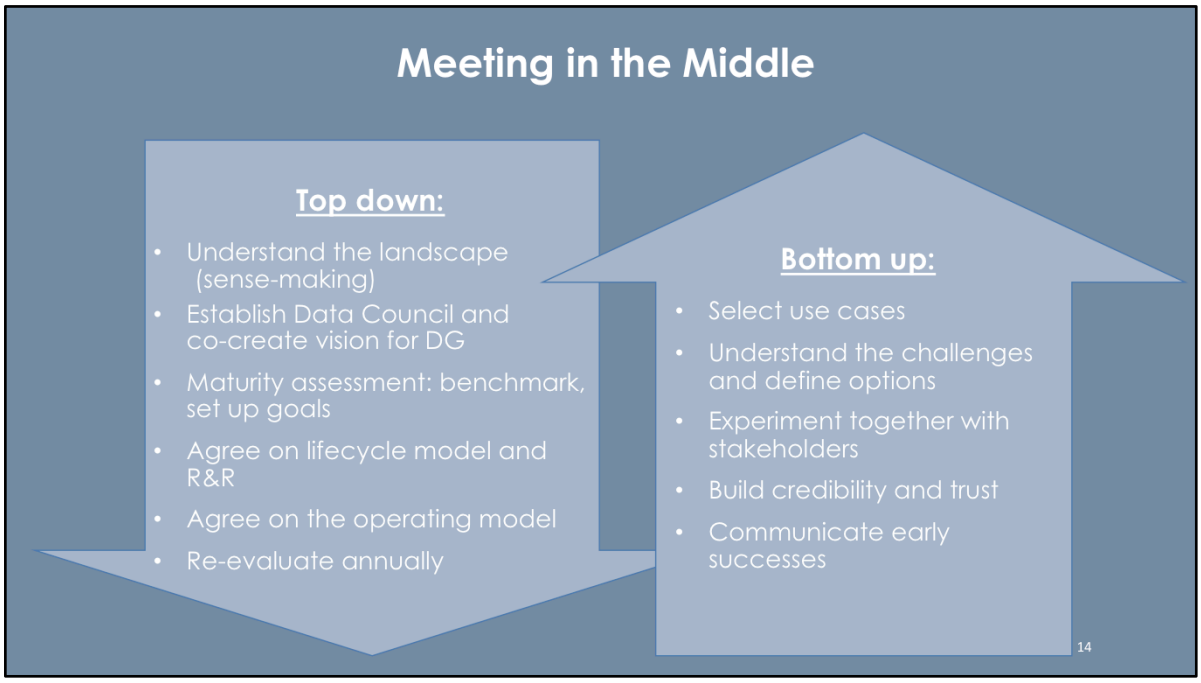
Continually review, refine and strengthen governance mechanisms. Communicate progress!



What should be centralized/distributed?

Some of the common patterns of centralized vs. localized areas

Need to strike a balance between safeguarding and enabling



Hybrid approach : start the data governance work at macro and micro level at the same time

Top Down -Understand the data Landscape: key data stakeholders, existing data, roles, policies, skills, identify gaps and opportunities
Stand up governance structures. Identify the point of controversy and strive to build shared understanding
Focus on people behavior, **create incentives**, **build capabilities,** prepare for disruption **(develop resilience)**
You may want to use a maturity model to assist with shared understanding of good data practices and agree

Sponsorship is key to success!

Lack of commitment – Shying away from taking ownership, perceive DG as tax
Whatever the steward is – i don't want to be one **– focus on value to them and make it easy for them**

Skills – augment and upskill staff /start **Data literacy program**

Poll on the biggest challenges

What are the top two challenges with DG progress in your organization?

1. Lack of sponsorship
2. Lack of business buy-in
3. Limited resources
4. Lack of skills
5. Lack of tools
6. Lack of shared understanding of what DG is
7. Other

16

Business folks are knocking on your door. Business stakeholders want to engage with you. Your use cases log grows.

When DG practices are embraced, folks are embedded in business units data teams, building relationships and trust

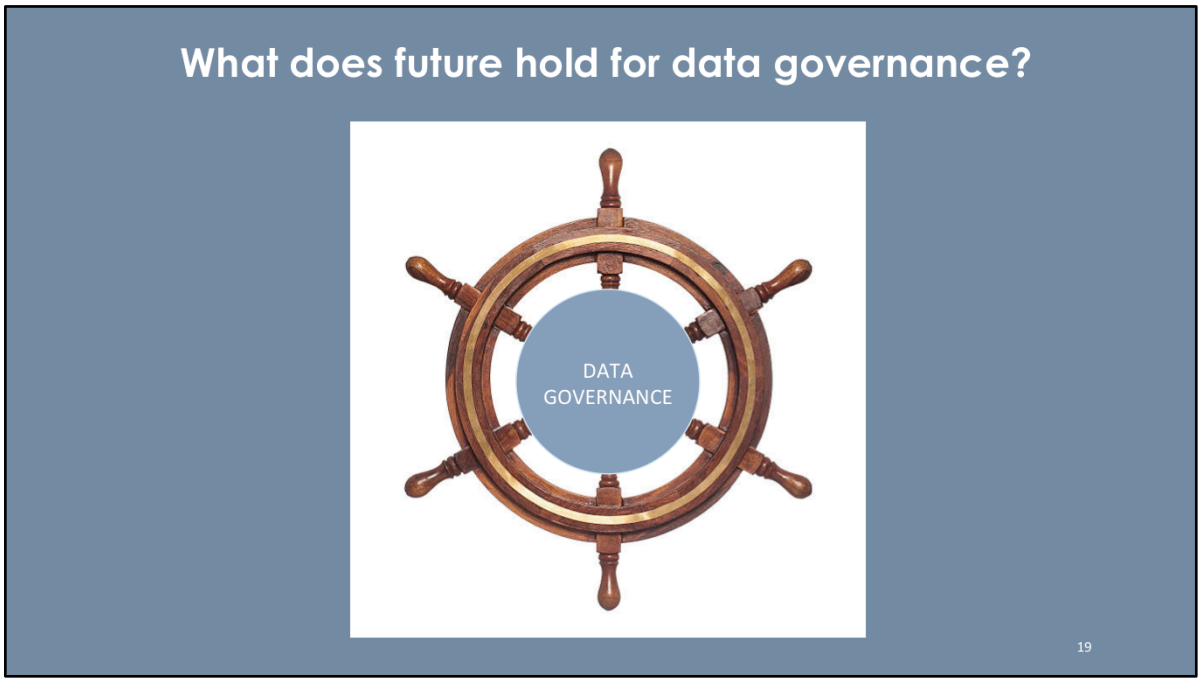
When data perspective has a seat at the table, Data perspective is included in decision making from planning to operations

Benefits of a Mature DG program

- Clear understanding of who is doing what and where to get support
- Increased value and reuse of data due to consistent and effective use of best practices for using and managing data
- Improved data quality and trust in data, reduced risk
- Improved data discoverability, better understanding of data, its lineage and provenance
- Improved access, reuse and sharing of data
- Better decision making and business insights

18

key areas:
- Business Impact,
- better efficiency
- better risk containment



W/ the move to computational DG what does the future hold for DG? will there be a need for DG team when the DG is automated?

Business metadata annotation is built into schema, DQ and access rules -into data ingest, to allow automation. DG is no longer a chore. The execution is automated. Computational part is all about execution. You will still need to define.
1)The world does not stand still and nothing is constant but change. the DG becomes more about defining the "good enough" (borrowed from Shirshanka Das, DataHub)
2)It will be a while before all the DM is automated away completely

Also semantics- continue to facilitate development of common vocabularies, categorization, etc. Convergence of DM and KM.

Steering into the future, charting the course and defining the roles and rules for what is to come in D&A&RPA&ML&AI&?

Thank you!

Irena.A.Zadonsky@gmail.com

20

Appendix

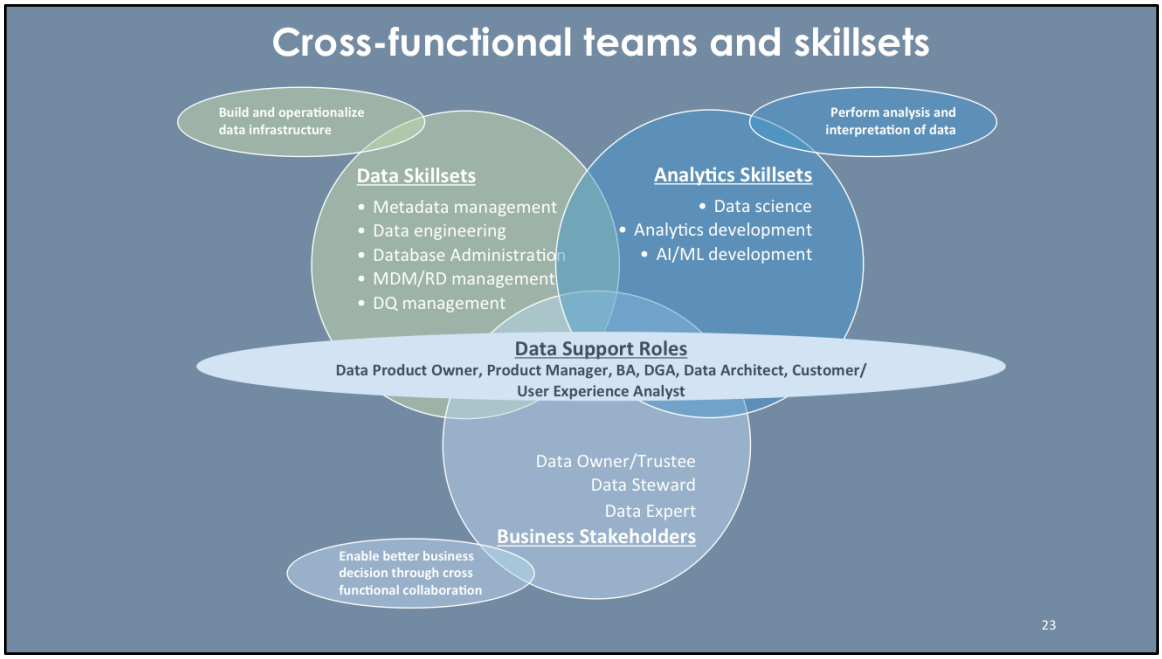
Data Roles

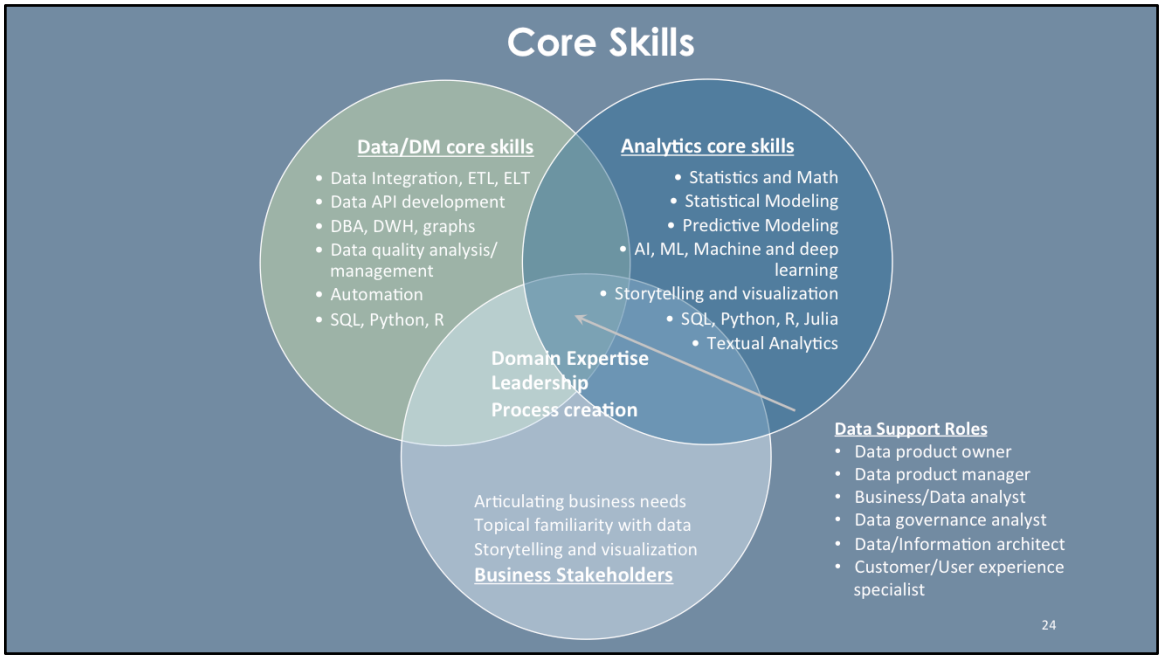
Every data asset needs to be accounted for

Key data roles:

- Data owner/trustee/sponsor
- Data Steward/Guardian/Custodian
- Data SME
- Data User

22


Cross-functional teams and skillsets
Build and operationalize data infrastructure
Perform analysis and interpretation of data
Data Skillsets
• Metadata management
• Data engineering
• Database Administration
• MDM/RD management
• DQ management
Analytics Skillsets
• Data science
• Analytics development
• AI/ML development
Data Support Roles
Data Product Owner, Product Manager, BA, DGA, Data Architect, Customer/ User Experience Analyst
Data Owner/Trustee
Data Steward
Data Expert
Business Stakeholders
Enable better business decision through cross functional collaboration
23


Core Skills
Data/DM core skills
• Data Integration, ETL, ELT
• Data API development
• DBA, DWH, graphs
• Data quality analysis/ management
• Automation
• SQL, Python, R
Analytics core skills
• Statistics and Math
• Statistical Modeling
• Predictive Modeling
• AI, ML, Machine and deep learning
• Storytelling and visualization
• SQL, Python, R, Julia
• Textual Analytics
Domain Expertise
Leadership
Process creation
Articulating business needs
Topical familiarity with data
Storytelling and visualization
Business Stakeholders
Data Support Roles
• Data product owner
• Data product manager
• Business/Data analyst
• Data governance analyst
• Data/Information architect
• Customer/User experience specialist
24