
AI-Powered Integrated Data Governance Frameworks:
A Holistic Approach
- Mehul Majethia


Topics to cover:
Data Governance Challenges
How is AI/ML impacting?
A holistic approach meaning
Why AI & ML?
Reimagining Data Governance
DG Frameworks
DG Models
The “Dirty” Data Problem

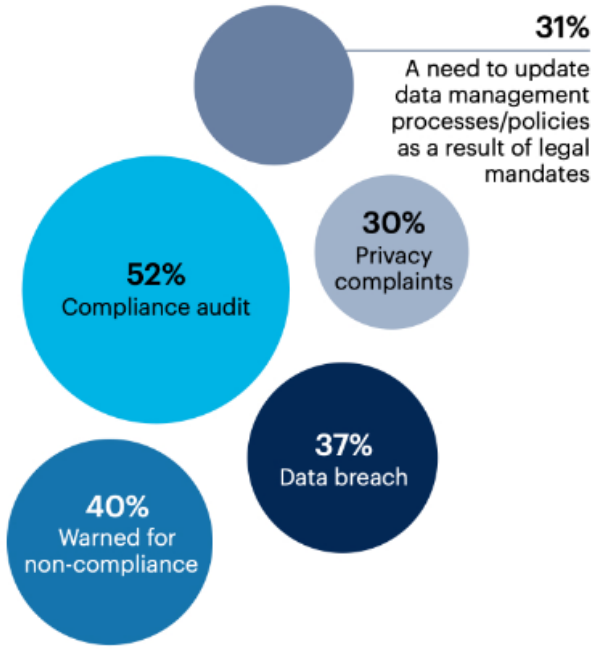
UNOFFICIAL

Data Governance Challenges

As regulations evolve, data governance is becoming essential for proper data management and compliance. What challenges do organizations face in doing so?

The majority have experienced data governance issues and have established or begun to execute a data governance framework

Has your organization experienced any of the following data governance issues?



The **most common** data governance **issues** that leaders have encountered are **compliance audits (52%), warnings for non-compliance (40%)** and **data breaches (37%).**

Data loss **21%**, Fined for non-compliance **20%**, Data leak **19%**, None of these **10%**, Any other data governance issues? **<1%**

n = 300

Data Governance Without AI or ML: Challenges

Here are some common challenges to traditional Data Governance:

Lack of consistency in data across business functions.

Divergent views of data available across business functions

Lack of common data definitions

Lack of documented Data Governance strategy

Misuse of data in self-service analytics or BI platforms

Big Data Governance

UNOFFICIAL

How is AI/ML impacting Data Governance?

Do you know where your data is?

Data cataloging

Metadata management

Data quality

Data modelling

Data policy and life cycle management

Data availability

Humans still needed

UNOFFICIAL

AI Governance vs AI-based data Governance

Artificial Intelligence (AI) governance refers to the strategies and policies that govern the ethical use, development, and deployment of AI technologies within an organization

AI-based data governance refers to the use of artificial intelligence technologies to automate and enhance the management, control, quality, security, and compliance of data within an organization.

Key components of an AI data governance framework :

1. Ethical guidelines and principles
2. Data quality management
3. Compliance and legal framework
4. Transparency and documentation
5. Data privacy and security
6. Accountability and oversight
7. Data ownership and access control
8. Stakeholder engagement
9. Continuous monitoring and improvement
10. Training and awareness programs

OFFICIAL

UNOFFICIAL

A holistic data governance approach :

A : Leverages elements of the published ISO standard on data governance [ISO-IEC] applied across three general enterprise levels:

1.Executive
2.Management
3.Operations

B: Employs a set of six data accountability areas from the ISO standard as a consistent means of assessing data governance across the three enterprise levels:

1.Collect
2.Store
3.Report
4.Decide
5.Distribute
6.Dispose

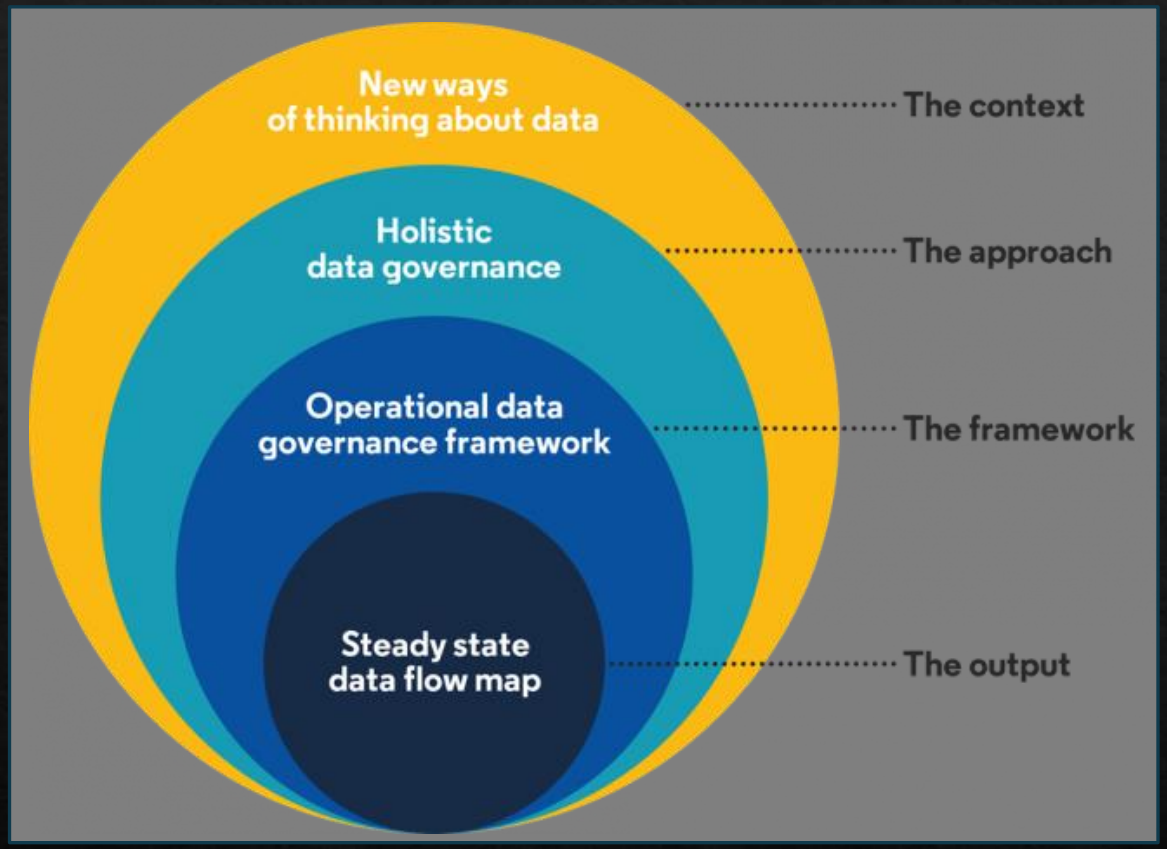
C: Each enterprise level accountability area is measured in terms of three data-specific aspects of governance as delineated in the ISO standard:

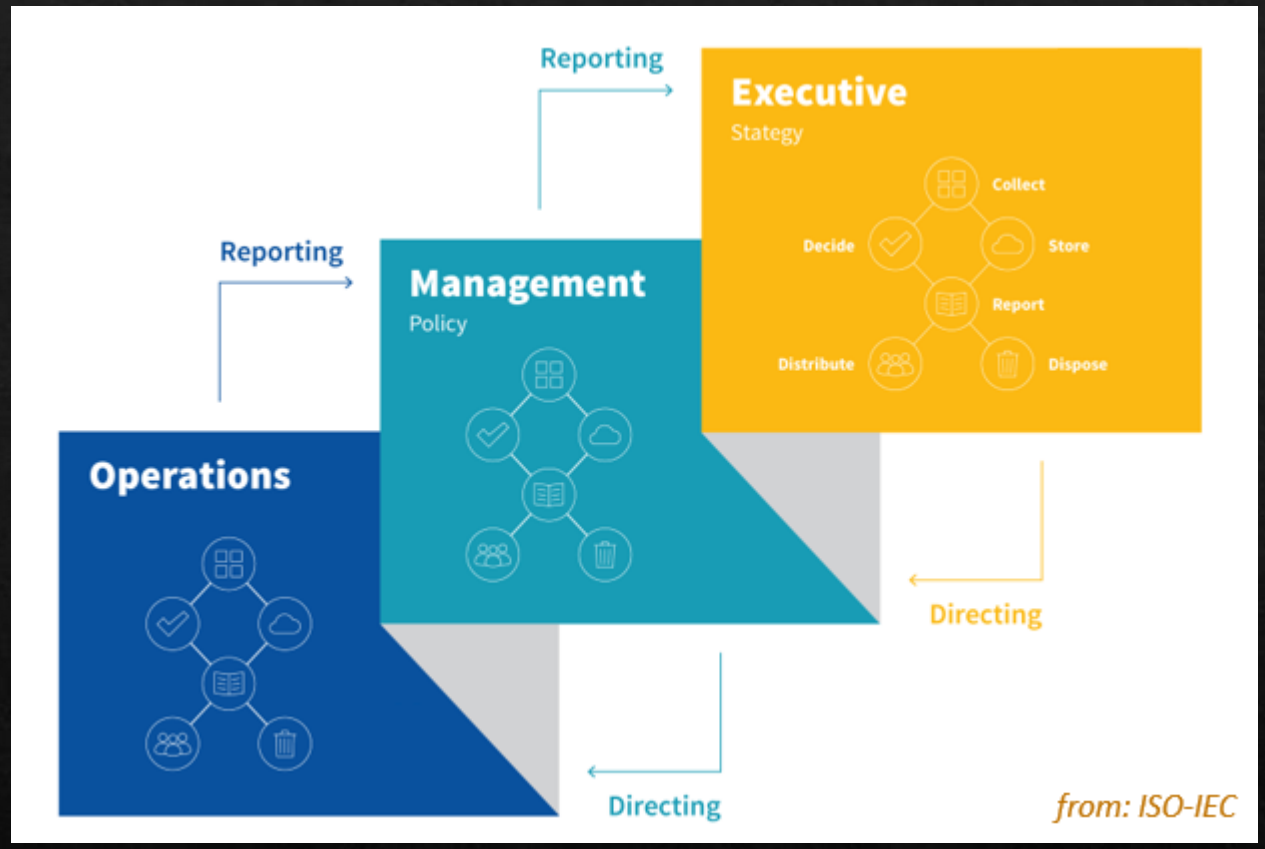
1.Value
2.Risks
3.Constraints

UNOFFICIAL

UNOFFICIAL

Holistic approach to Data Governance (With or without use of AI/ML)





Focus on three areas for a holistic data governance approach:

1. Establish a data governance framework to empower, not restrict, your organization
2. Provide tooling that promotes trust in governed data
3. Educate and engage your workforce with your governance strategy

UNOFFICIAL

UNOFFICIAL

Why AI for data governance?

Automate data compliance

Strengthen data security

Democratize data

Improve the quality of your data

UNOFFICIAL

UNOFFICIAL

3 Ways AI Can Impact Data Governance

Recent breakthroughs in artificial intelligence present a new opportunity to use AI to enhance data governance and in turn, use data governance to enhance AI.

AI Can Reduce Risk by monitoring Data Governance Policies

AI Can Save Time and Point People to the Right Policies

AI Can Automate Data Governance Processes

UNOFFICIAL

UNOFFICIAL

How AI Is Reimagining Data Governance

- Data Classification and Tagging & Cleansing:
- Data Security:
- Data Governance Analytics & Automation:
- Data Governance Decision Making:

As businesses increasingly rely on AI and machine learning to drive operational efficiencies, it's important to understand that the value of these technologies is only as high as governance standards that are in place. The implementation of a proper **data governance** framework is essential to enable organizations to fully unlock the value of their data and keep their businesses moving in the right direction.

UNOFFICIAL

UNOFFICIAL

Data Governance Framework Models

The models are based on how data governance decisions will flow through your organization.

Top-down

Company leadership implements data governance policies that are then passed down to individual business units and shared with the rest of the company.

Bottom-up

Company Employees at the lower levels implement data governance practices, such as standardizing naming conventions, which spread to the higher levels of the organization.

Hybrid

Data governance decisions involve different levels of the organization. For example, a company uses a center-out model to suggest a course of action but employs a top-down model to make the final decision.

Center-out

The team or individual responsible for data governance sets data standards that the entire organization follows.

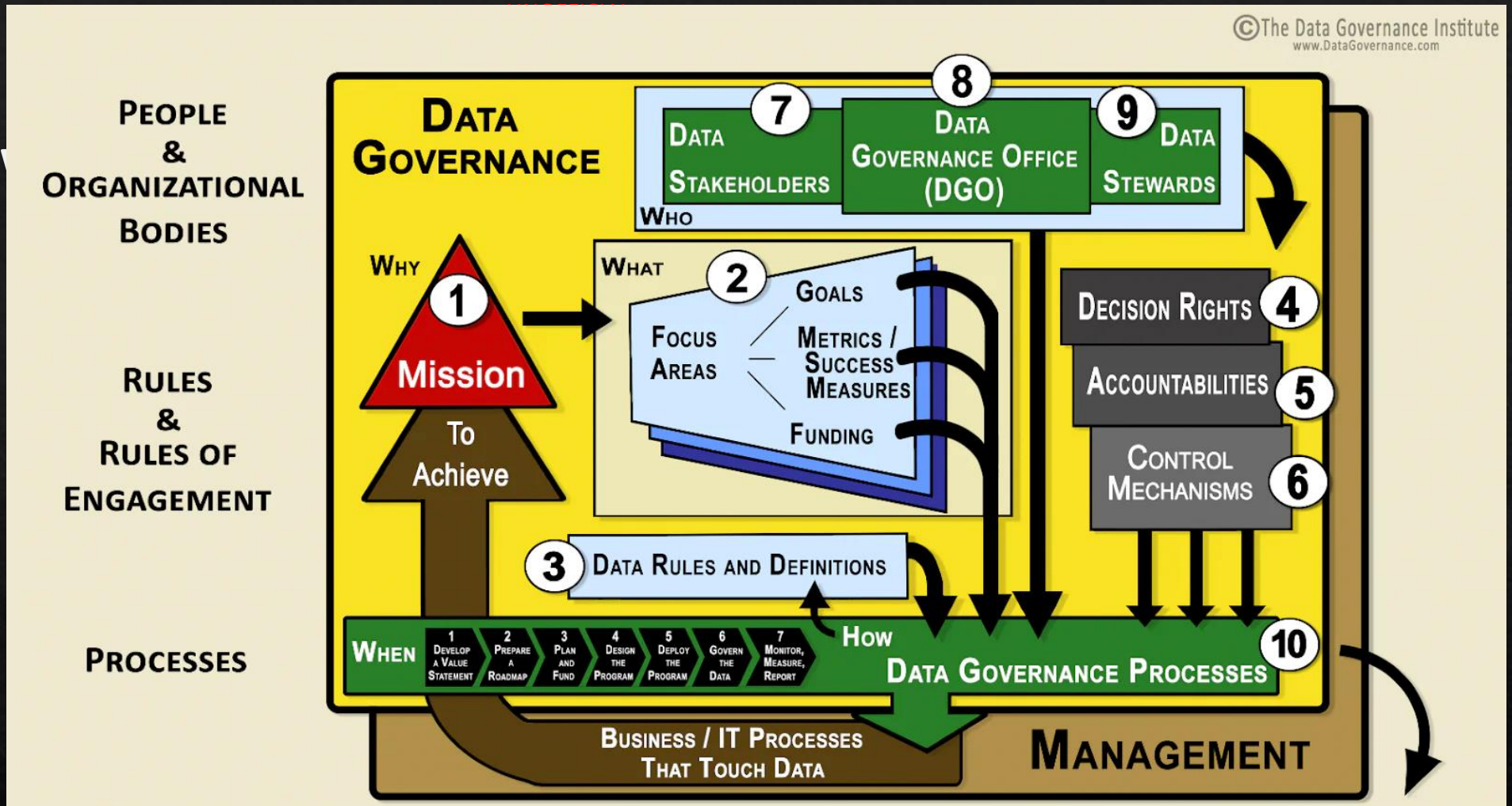
Silo-in

Various departments come together to align on data governance while keeping in mind the needs of each group.

UNOFFICIAL

Data Governance Framework from DGI :

Other leading data governance frameworks to consider include the Data Management Body of Knowledge (DAMADMBoK) and the Data Management Capability Assessment Model (DCAM).



Definition:

Data Governance is the exercise of decision making and authority for data-related matters.

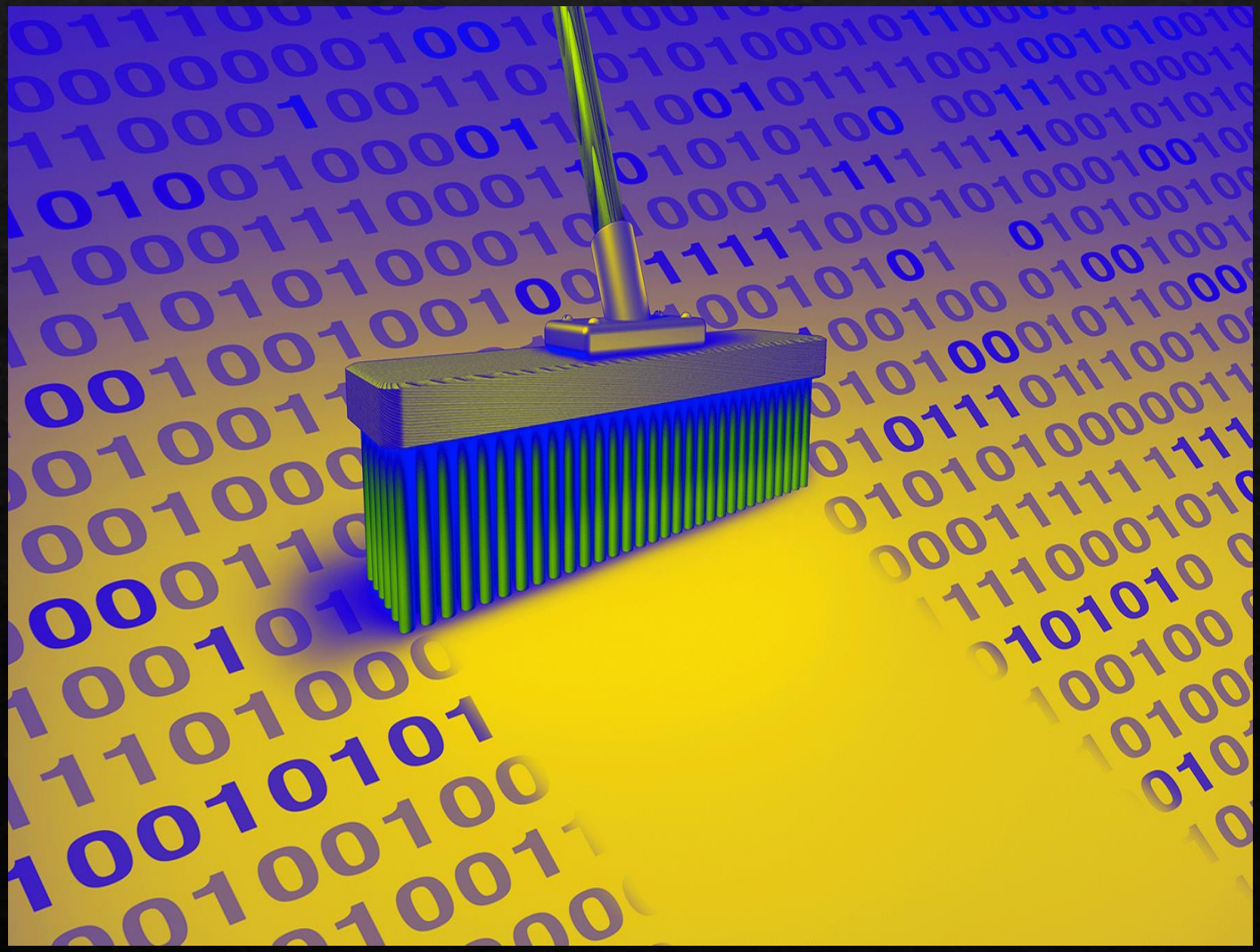
It's a system of decision rights and accountabilities for information-related processes, executed according to agreed upon models which describe who can take what actions with what information and under what circumstances, using what methods.

Processes for governing how data is used, and when, and by whom

1. Aligning Policies, Requirements & Controls
2. Establishing Decision Rights
3. Establishing Accountability
4. Performing Stewardship
5. Managing Change
6. Defining Data
7. Issue Resolution
8. Specifying Data Quality Requirements
9. Building Governance into Technology
10. Stakeholder Care and Support
11. Stakeholder Communications
12. Measuring and Reporting Value

UNOFFICIAL

The “Dirty” Data Problem: A Case of AI and Data Interdependency



Dirty data refers to any erroneous, misleading, duplicate, non-integrated, incorrect, and non-compliant data that is not likely to give proper insights.
Dirty data can also be the information that is in computer memory but has not yet been put into a database.

- ✓ ML algorithms can detect missing and outlier values.
- ✓ ML algorithms can spot duplicate entries with slightly different terminology that describe the same thing.
- ✓ ML algorithms can normalize data to the standard format and vocabulary.

Walk cautiously on the road of the AI-predictive model. The value to the business increases with each progression through the analytics maturity model:

1.Start with process and data mapping.
2.Then, move to descriptive analytics.
3.After that, predictive analytics.
4.Finally, to prescriptive analytics.

To conclude…

The integration of AI and ML in data governance- is it just an enhancement?

It’s a

GAME CHANGER!