MANAGING BIG DATA

PROF. DR. FLORIAN STAHL

Managing Big Data

- Types of Data
- Data Architecture
- Master Data Management
- Data Quality
- **Data Governance**
- Data Security

Insufficient Policies
The case of Uber

https://www.youtube.com/watch?v=pJaQDk0B6dY

US Data laws requires to comply with state **consumer protection laws** in safeguarding personal information and to immediately notify authorities in case of a breach; to establish methods to protect user data stored on third-party platforms; and to create strong password-protection policies.

A **transparent, well-organized** DG framework with established **policies** and clear **responsibilities** may have prevented Uber from this infringement.

No clear responsibilities
The case of British Airways

In 2018, Data from millions of customers were stolen. This was only possible due to the insufficient security measures in place to protect the personal data of the customers.

With a well-organized DG framework with people responsible for data security and a centralized, secured data storage this could have been prevented.

British Airways fined £20m over data breach

🕐 16 October 2020

GETTY IMAGES

British Airways has been fined £20m ($26m) by the Information Commissioner's Office (ICO) for a data breach which affected more than 400,000 customers.

Data Governance

https://video.sas.com/detail/video/3161708635001/what-is-data-governance
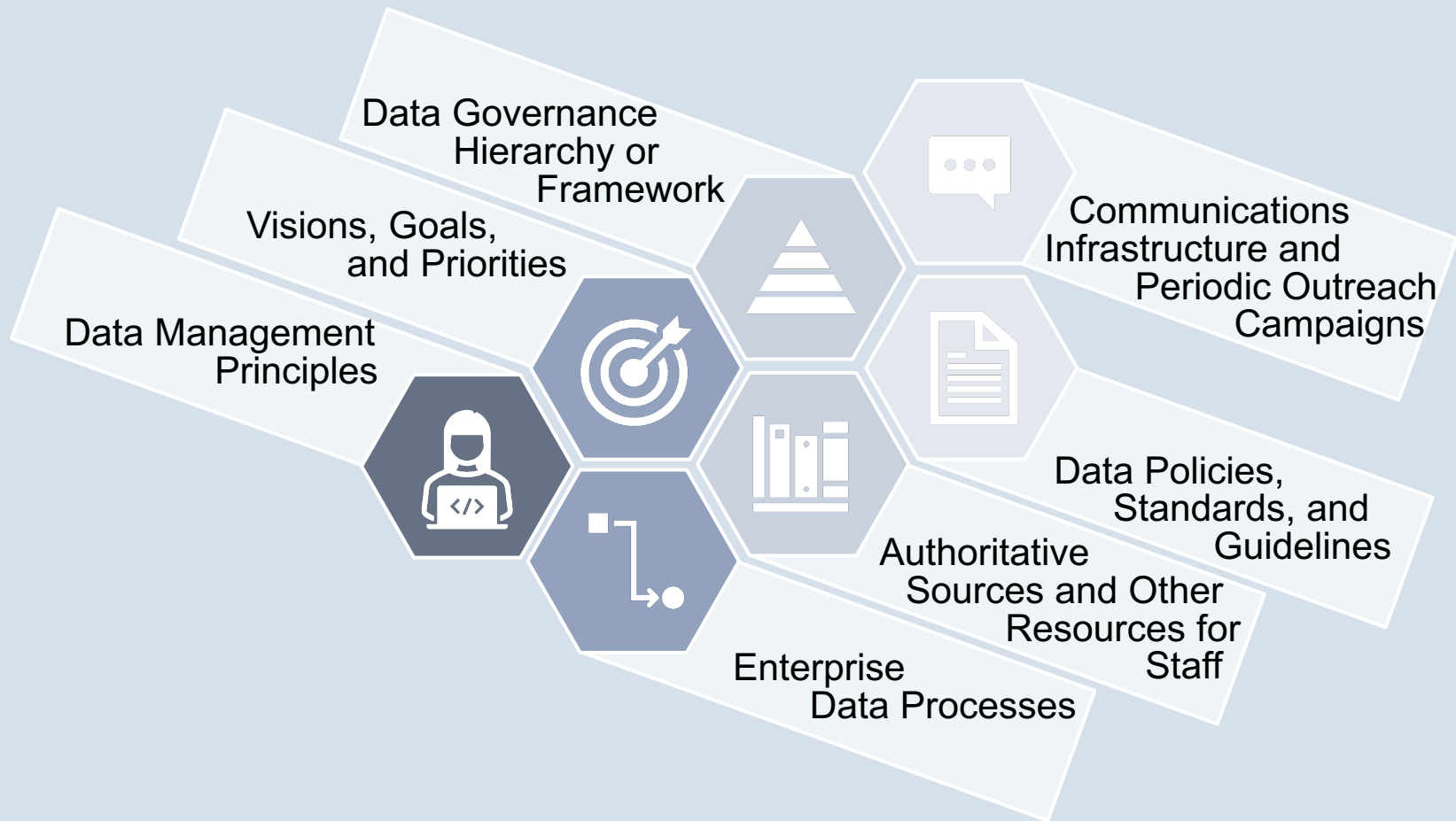
What is Data Governance?

Data governance is the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets. The data governance function guides how all other data management functions are performed.

Data governance describes the development of the ownership, stewardship, and operational structure needed to ensure that corporate data is managed as a critical asset and implemented in an effective and sustainable manner.

Data governance is the discipline of administering data and information assets across an organization through formal oversight of the people, processes, technologies, and lines of business that influence data and information outcomes to drive business performance.

Components of a Data Governance Program

Data Governance Hierarchy or Framework

Visions, Goals, and Priorities

Data Management Principles

Communications Infrastructure and Periodic Outreach Campaigns

Data Policies, Standards, and Guidelines

Authoritative Sources and Other Resources for Staff

Enterprise Data Processes

Vision, Goals, and Priorities

Business units affected

Existing and forthcoming policies, strategies and constraints

SCOPE

Data that **will be governed** and **required** data management **domains**

! A **vision statement, goals**, and **priorities** align a data governance program to the overall organizational mission and help focus both the program and corporate leadership on data management priorities.

Data Management Principles

Data management principles can be described as the **foundational statements of organizational values.**
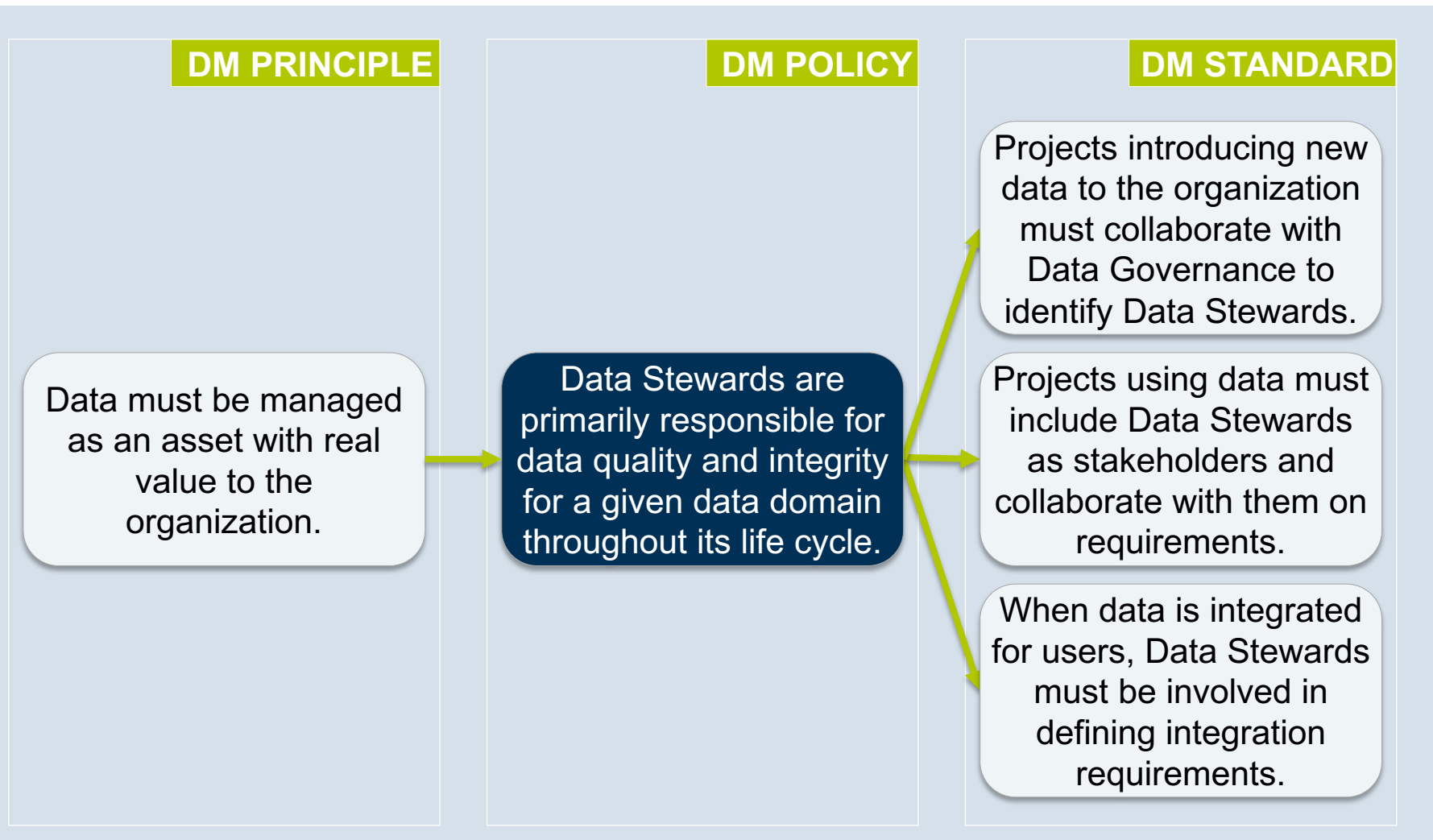
Rationale

- Business decisions are based on **accurate** and **timely information**
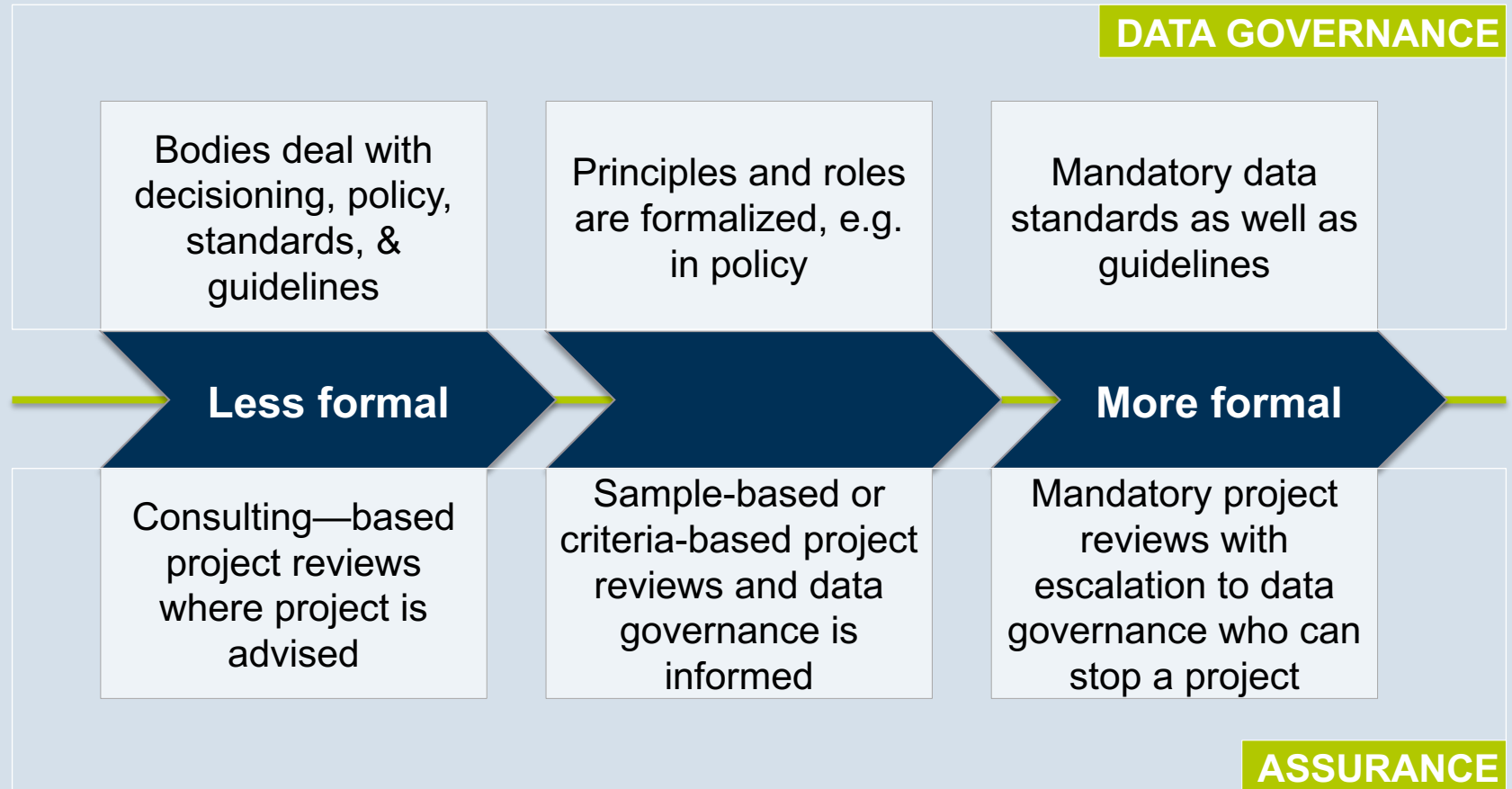- Data and metadata have **real, tangible value** and should be carefully managed

Implications

- Data is owned by the enterprise
- **Data governance framework**
- **Governing** data **cooperatively**
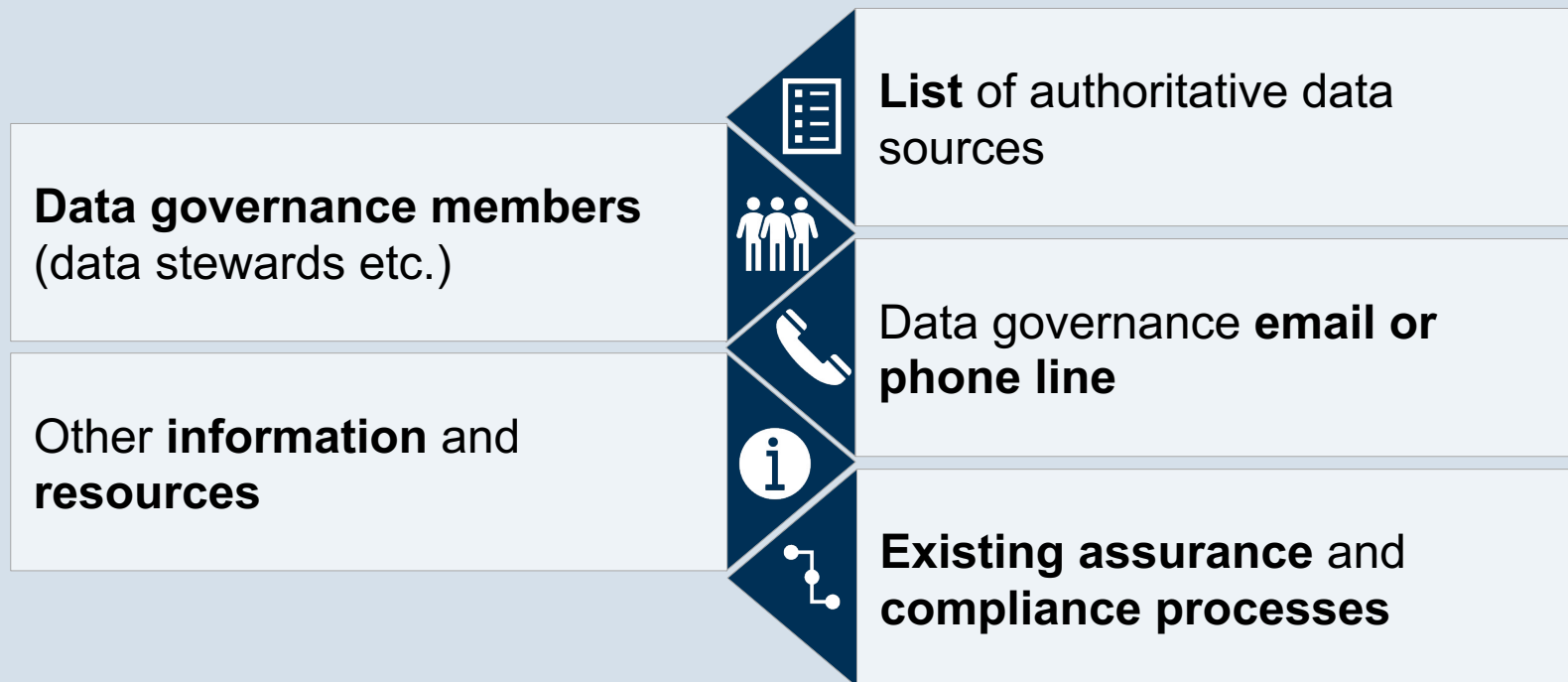- **System of record** and other authoritative sources
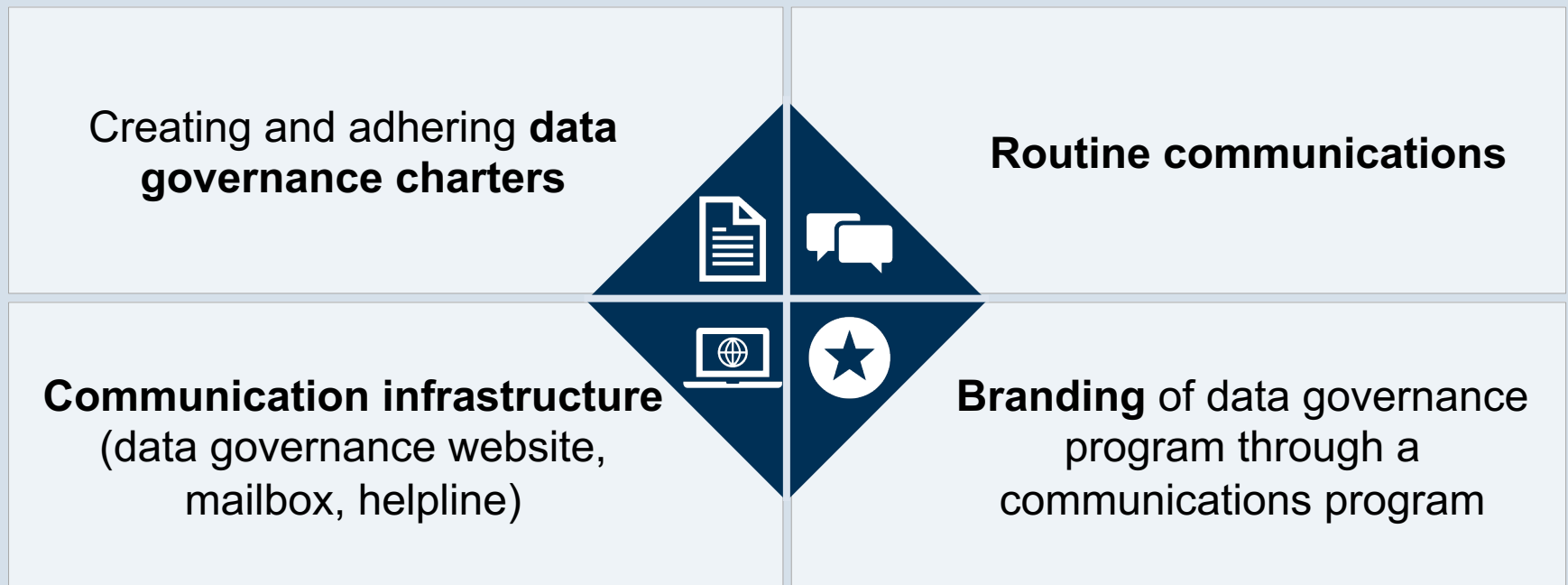
Data Policies, Standards and Guidelines

DM PRINCIPLE

DM POLICY

DM STANDARD

Data must be managed as an asset with real value to the organization.

Data Stewards are primarily responsible for data quality and integrity for a given data domain throughout its life cycle.

Projects introducing new data to the organization must collaborate with Data Governance to identify Data Stewards.

Projects using data must include Data Stewards as stakeholders and collaborate with them on requirements.

When data is integrated for users, Data Stewards must be involved in defining integration requirements.

Data Governance and Assurance Spectrum

DATA GOVERNANCE

| Bodies deal with decisioning, policy, standards, & guidelines | Principles and roles are formalized, e.g. in policy | Mandatory data standards as well as guidelines |

Less formal → → **More formal** →

| Consulting—based project reviews where project is advised | Sample-based or criteria-based project reviews and data governance is informed | Mandatory project reviews with escalation to data governance who can stop a project |

ASSURANCE

Authoritative Sources and Other Resources for Staff

Data governance members (data stewards etc.)

Other **information** and **resources**

List of authoritative data sources

Data governance **email or phone line**

Existing assurance and **compliance processes**

Communications Infrastructure and Periodic Outreach Campaigns

Creating and adhering **data governance charters**

Routine communications

Communication infrastructure
(data governance website, mailbox, helpline)

Branding of data governance program through a communications program

Who is Data Governance?

INFORMAL DATA STEWARDS

Lack of formal network to coordinate their efforts with others in working on data management

Difficulties to **determine priorities**
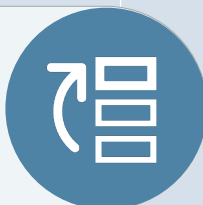
Data management only as a **part-time effort** of overall responsibilities

FORMAL DATA STEWARDS

Engages people from all parts of the organization with **clearly stated roles** and **responsibilities**
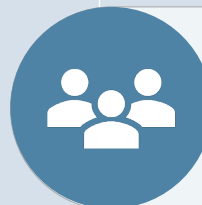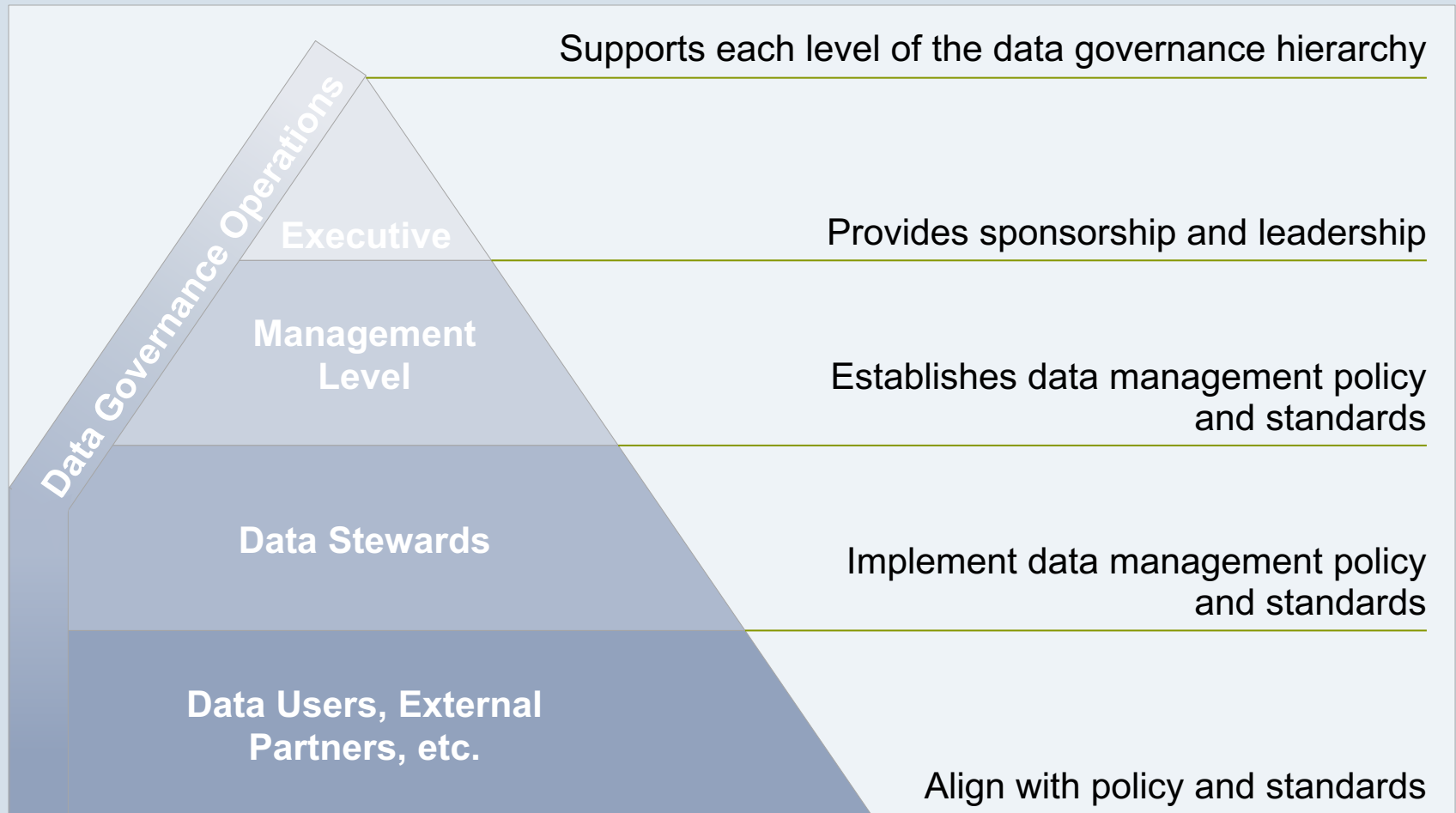
Aligns **data management priorities** with **business needs** and existing policies and procedures
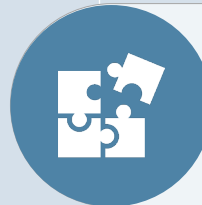
Provides **metrics** and **maturity scale**

Data Governance Framework

Supports each level of the data governance hierarchy

Data Governance Operations

Executive — Provides sponsorship and leadership

Management Level — Establishes data management policy and standards

Data Stewards — Implement data management policy and standards

Data Users, External Partners, etc. — Align with policy and standards
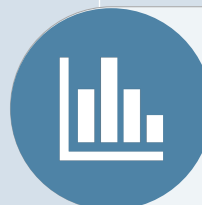
Data Governance Operations (DGO)

Data Governance Operations

➢ Small number of individuals **fully dedicated** to the **overall data governance** initiative

➢ Ensures **coordination across the levels** of the data governance framework and the force that drives data governance forward

➢ Drafting material for the data governance discussion

➢ Providing **meeting support** for the executive- and management level

➢ **Research issues** and escalation of issues if necessary

➢ **Assessment** of existing **capabilities**

Executive Level

Executive
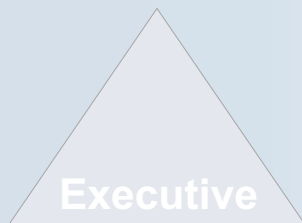
- ➢ **Decision makers** who set the enterprise vision and tone for governing data as an asset in line with business concerns
- ➢ **Identifies members of the management level** to represent their organizations
- ➢ **Approves data management principles, policies**, and **standards** as well as **data-related strategies** and roadmaps
- ➢ **Decision point** for **escalated situations**
- ➢ **Commercial sector:** membership includes the executive of each line of business as well as others (heads of legal, privacy, risk management, …)
- ➢ **Creation of Charter**
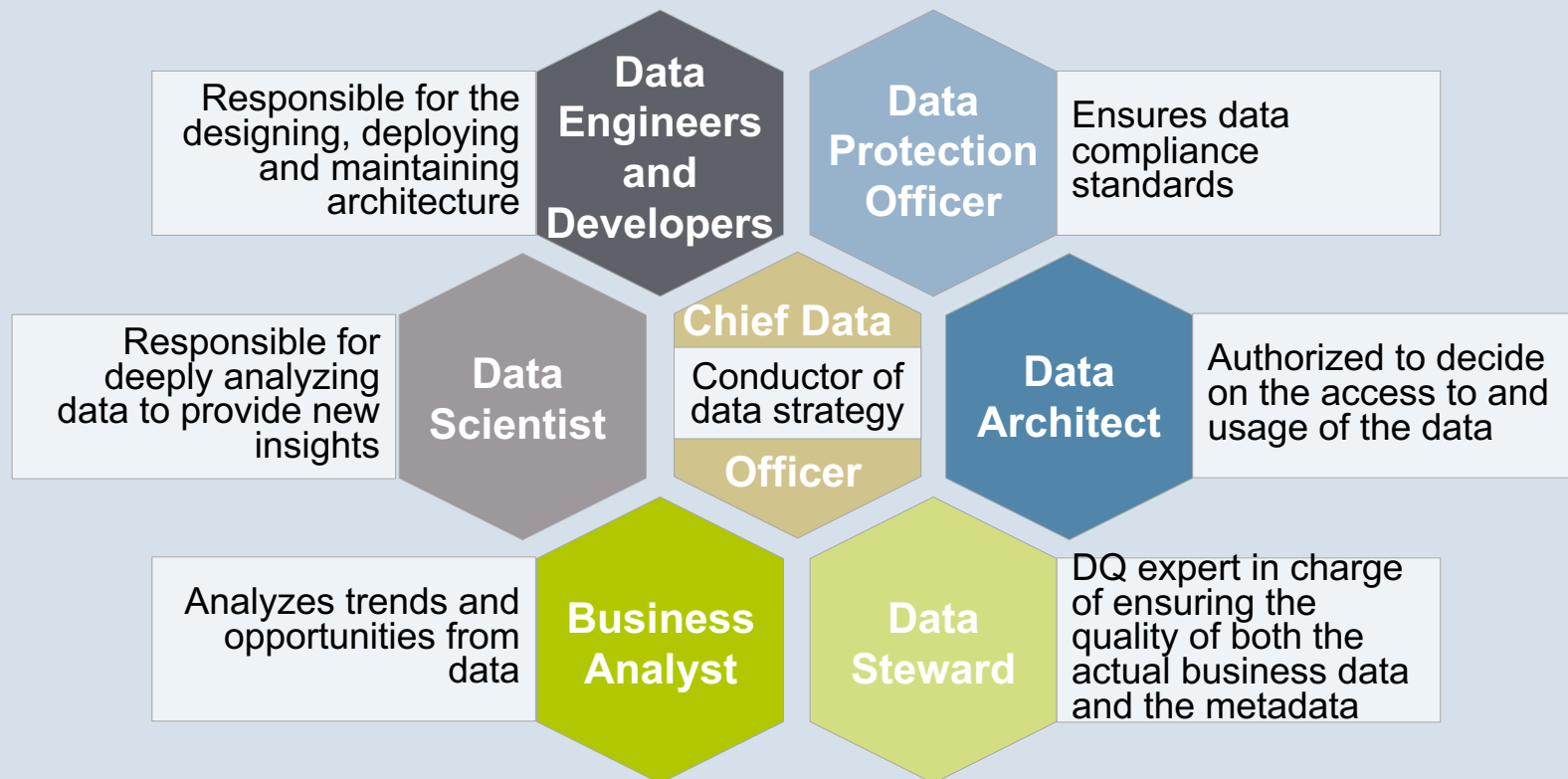
Management Level

Management Level

- ➢ Executes the **enterprise vision**
- ➢ Trusted **advisor** to the executive level: frames options and provides recommendations
- ➢ Strategic perspective: **defines data management principles**, policies, and standards for **executive approval**
- ➢ **Collaborates with executive level** on **data-related strategies** and roadmap priorities
- ➢ Often **identifies data stewards**
- ➢ Might **determine criteria** for what constitutes an authoritative data source
- ➢ May decide **issues** or frame options to executive level
- ➢ Might be **chaired by the director of the DGO group** and should include **one or more business-side representatives** from each major data producer or consumer in the enterprise
- ➢ Should **meet** with executive-level **regularly**
- ➢ **Data management charter** at this level is also common

Data Stewards Level

Data Stewards

➢ Maintain **quality of data** for their respective data domain

➢ Help **resolving data issues**, conflicts, and inconsistencies within their authority and coordinate with their management level representatives

➢ Manage **day-to-day business processes that create data** and supply it (in)directly to downstream business users

➢ Maintain and publish associated **metadata** to provide transparency

➢ Working with known consumers to define **data quality requirements**

Roles in Data Governance

Data Engineers and Developers — Responsible for the designing, deploying and maintaining architecture

Data Protection Officer — Ensures data compliance standards

Data Scientist — Responsible for deeply analyzing data to provide new insights

Chief Data Officer — Conductor of data strategy

Data Architect — Authorized to decide on the access to and usage of the data

Business Analyst — Analyzes trends and opportunities from data

Data Steward — DQ expert in charge of ensuring the quality of both the actual business data and the metadata

Roles in Data Governance

Chief Data Officer

- Responsible for defining, deploying and tracking the data strategy.

- Make sure that data is considered a valuable business asset at the executive level.

- Push investments in data compliance to minimize risks.

Roles in Data Governance

Data Protection Officer

- Ensures compliance standards as defined by authorities.

- Need to consider country-and industry-specific regulations

- Especially concerned about privacy of personal data from employees, customers, …

Roles in Data Governance
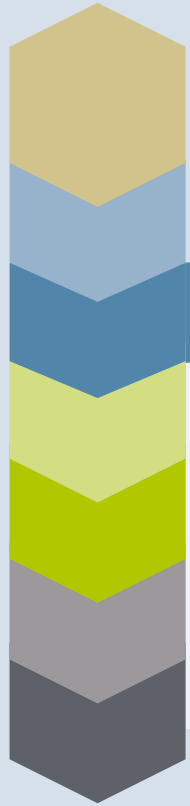
Data Architect

- Make sure that the "data house remains solid".

- Ensure flexibility to help users to own and manage their own datasets while annotating, enriching and certifying data.

- May prioritize information based on measurable outcomes.

Roles in Data Governance

Data Steward

- DQ expert in charge of ensuring the quality of both the actual business data and the corresponding metadata

- Performs extensive and regular data quality checks

- In charge of taking DQ initiatives such as corrective and preventive measures

Roles in Data Governance
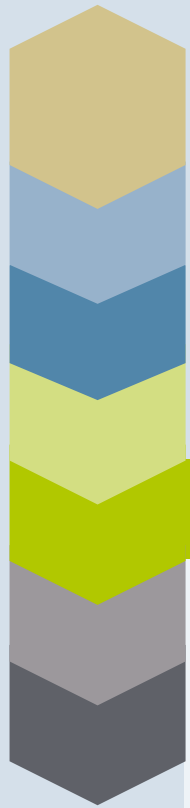
Business Analyst

- Analyze trends and opportunities
- Identify and calculate risks to determine the pulse of the business.
- Users of BI-tools such as Tableau, Power BI and Looker.
- Mission: get insights from data pipelines and present them through comprehensive dashboards.
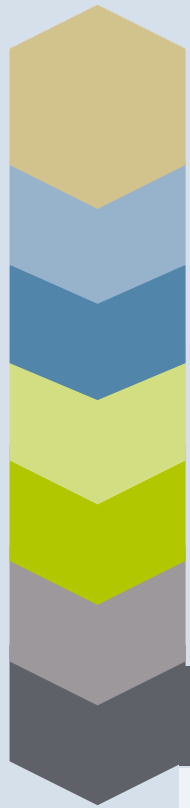
Roles in Data Governance
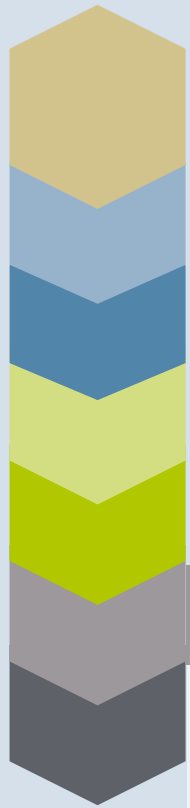
Data Scientist

- Analyzes data using state-of-the-art analytical techniques to provide new insights

- Has a multidisciplinary profile combining ICT skills with quantitative modeling, business understanding, communication, and creativity

Roles in Data Governance

Data Engineers and Developers

- Responsible for designing, deploying, and maintaining the architecture to process complex flows in the organization

- Give autonomy to the whole data community whilst keeping control of access, authorization and product administrations

Benefits of Data Governance

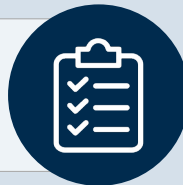|	
Common understanding of data and their value	Improved quality of data
A data map	360° view of customers and business entities
Consistent compliance	Improved data management
Easy access	Efficiency increase throughout the company

Implementing Data Governance – Questions to Keep in Mind

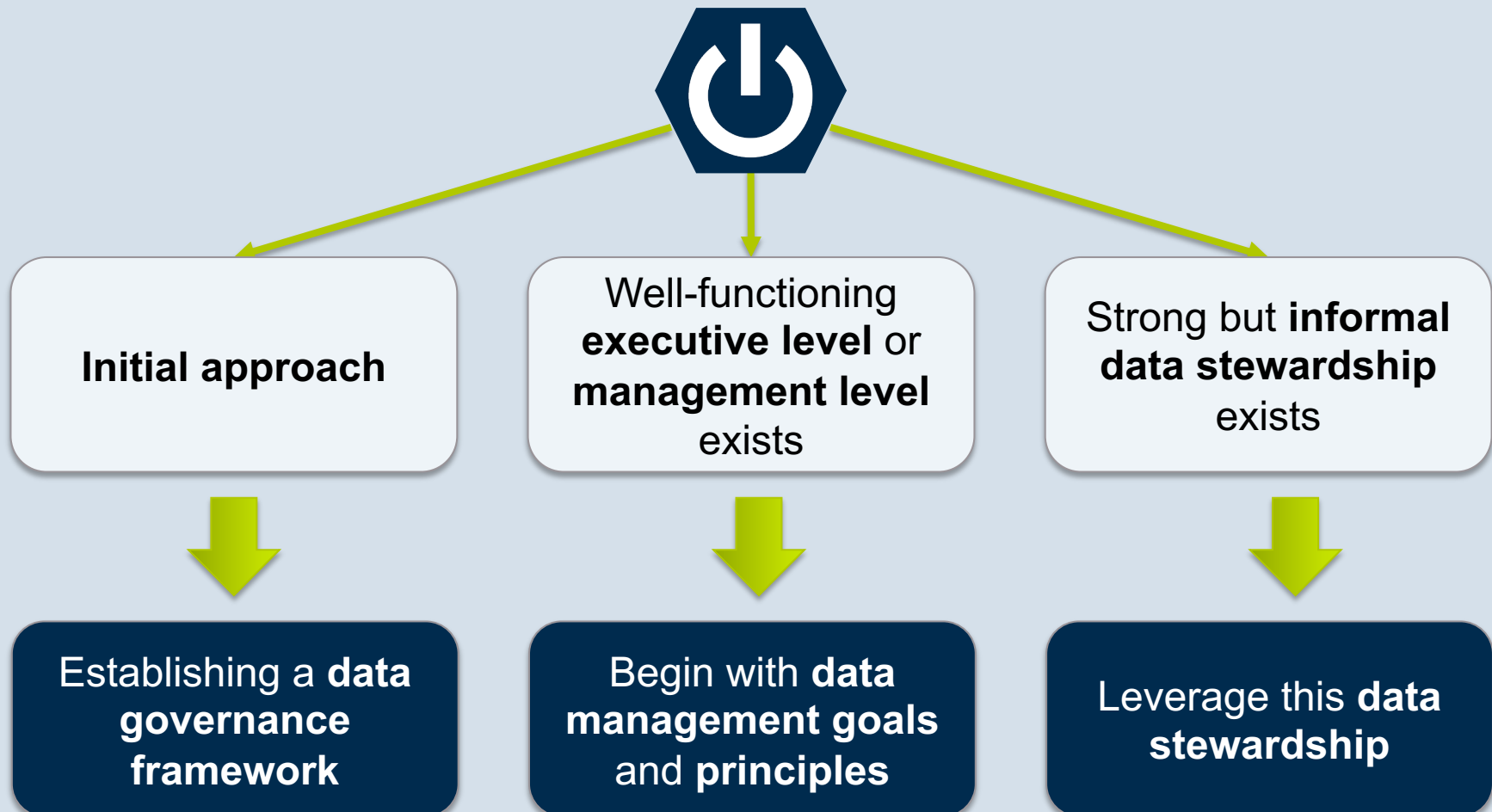**What is the opportunity cost of not getting data governance right?**

1

2 Who is leading governance efforts today, and what would it look like to elevate the conversation to the C-suite?

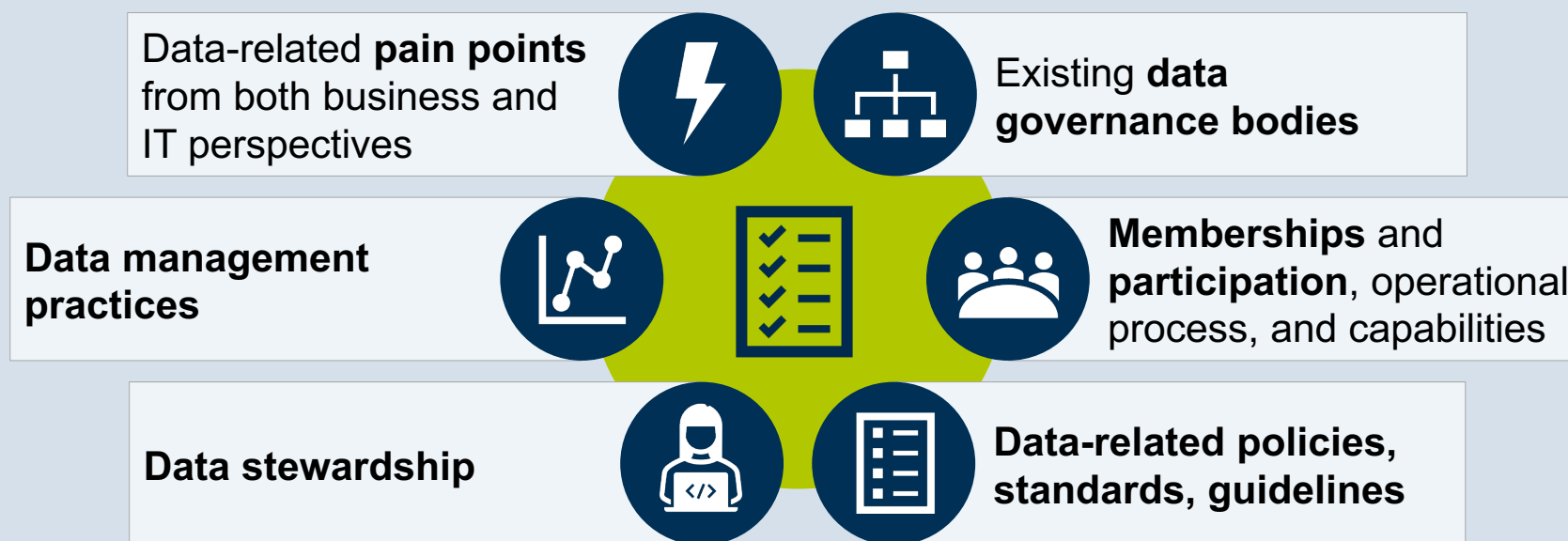**Where is governance most important? What domains and parts of domains does the organization most need right now?**

3

4 What governance archetype best fits the organization, and are current efforts aligned to that level of need?

How can governance be accelerated by adjusting its focus and injecting iterative working concepts?

5

6 Do you have the in-house capabilities to manage such a shift?

Implementing Data Governance

Initial approach	Well-functioning **executive level** or **management level** exists	Strong but **informal data stewardship** exists
Establishing a **data governance framework**	Begin with **data management goals** and **principles**	Leverage this **data stewardship**

Current State Assessment

Data-related **pain points** from both business and IT perspectives

Existing **data governance bodies**

Data management practices

Memberships and **participation**, operational process, and capabilities

Data stewardship

Data-related policies, standards, guidelines

Set up Processes
Single Point of Trust

Save time and resources of your organization once operationalized

Single control point will be the cornerstone of data governance framework

Point of trust

Select datasets and directly assign roles & responsibilities to your single point of control
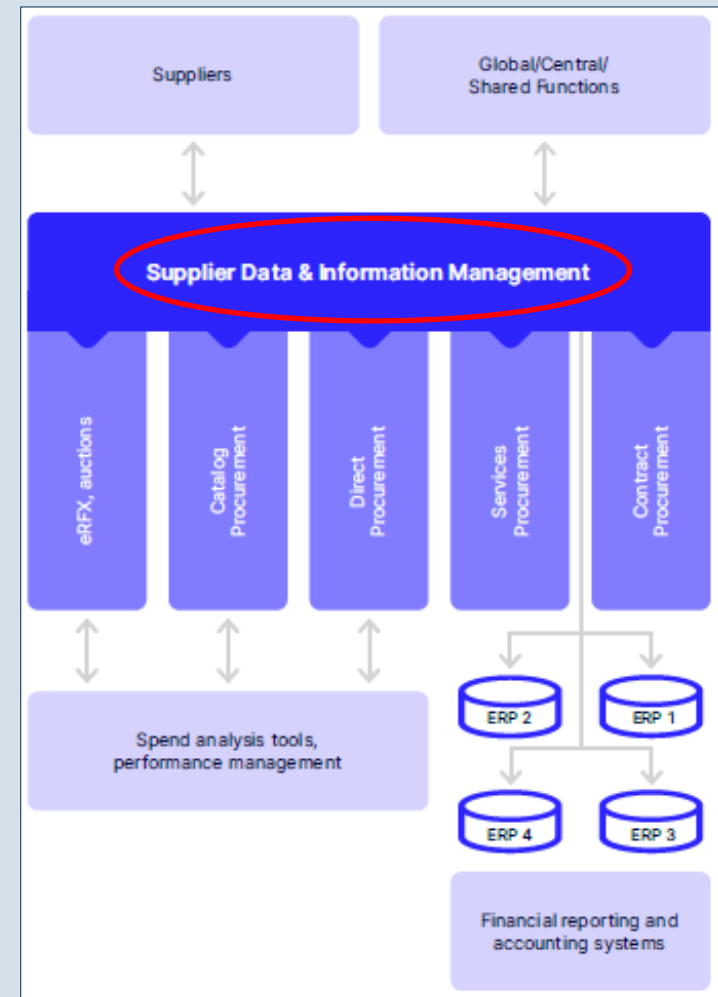
Organize data for massive consumption
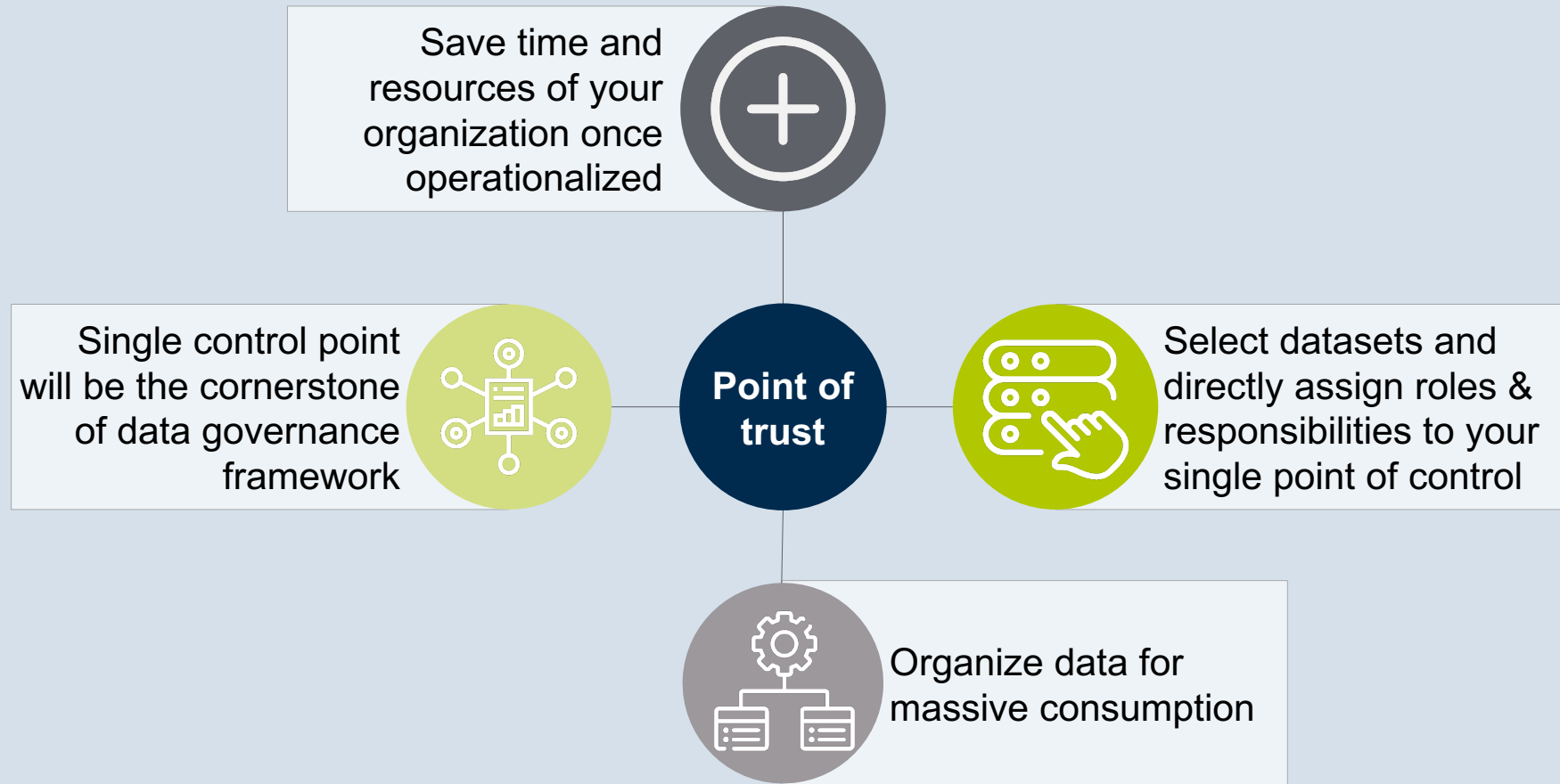
Set up Processes
Example: Supplier Data

!

It is simpler to funnel data through one entry point down into different systems rather than creating multiple different systems and then struggling to bring it all together.

Set up Processes
Automate Data Pipelines

Leverage the power of automation to streamline your dataflows. Use machine learning to learn from remediation and scale faster.
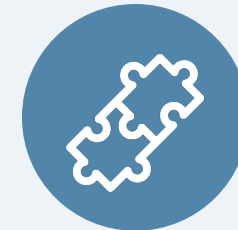
Typical use cases for ML:

Improve developer's productivity by suggesting next best actions.

Capture knowledge from business users and data professionals.
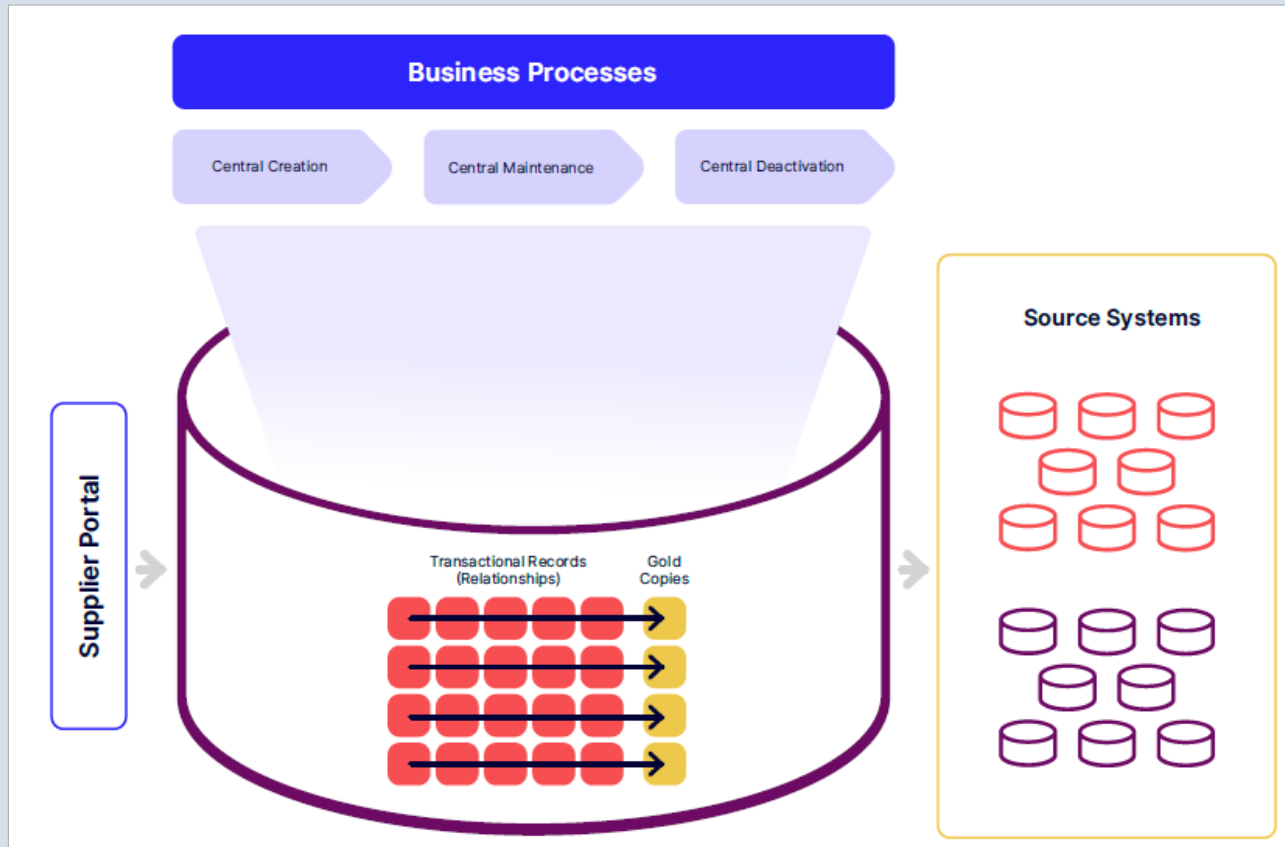
Data error resolution and matching.

! Establish a **platform for all** and leverage **user-friendly apps** for your stakeholder's community. A one-size-fits-all approach is probably not possible.

Set up Processes
Example: Supplier Data

Set up Processes
Business Glossary

Business glossary is used to **define** collections of **terms** and link them to **categories**.

When published, everyone authorized can **access** the glossary.

By making it accessible, it can be **edited, validated** and **enriched**.

Data lineage gives a big picture to **trace** and **track data flows**.

Data lineage quickly **answers to audit trails** as requested by the competent authorities.

Data lineage **supports change management.**

Data Governance Tools

Developed data governance around a specific data management domain

Tools aimed at various aspects of data governance are emerging in the marketplace

Help organize and **support the work required** to design and implement a data governance framework

Serve primarily as a **repository** of **data governance components**
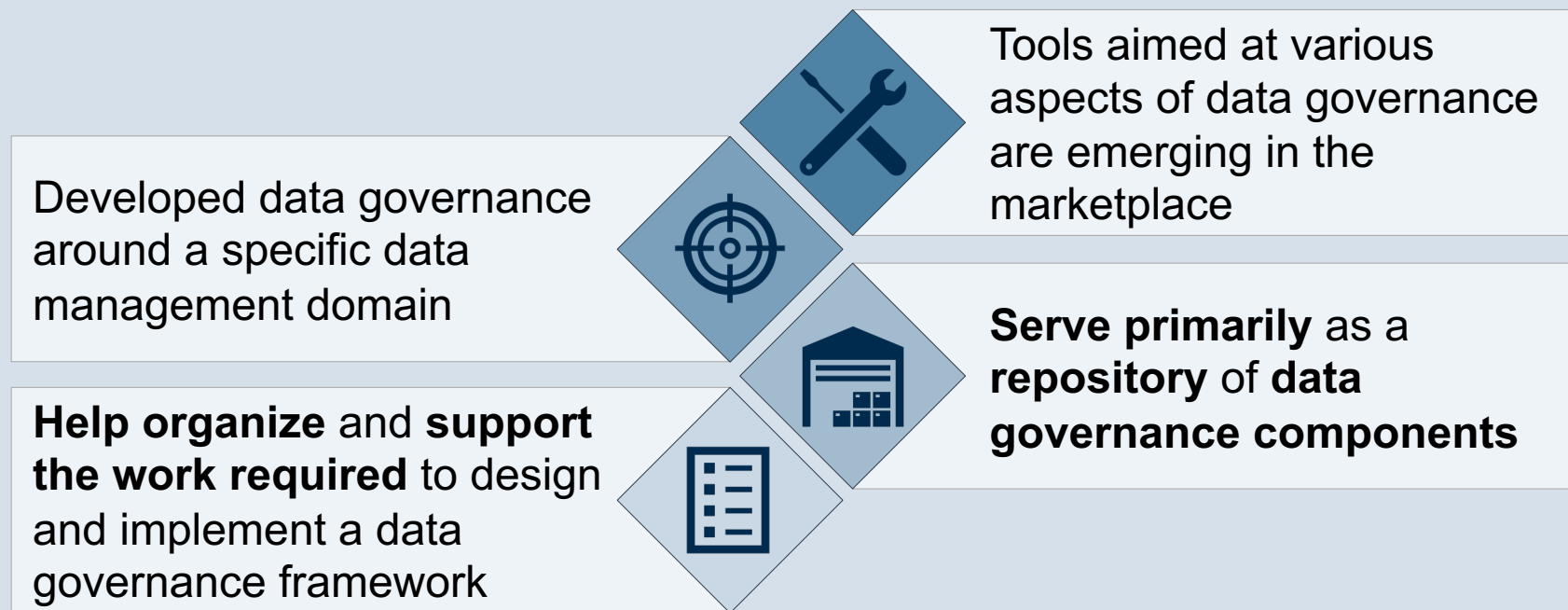
Six Ways to Drive Data-Governance Excellence

Generate excitement for data

6

Secure top management's attention

1

Choose iterative and focused implementation

5

Integrate with primary transformation themes

2

Apply the right level of governance

4

Prioritize data assets and focus data leadership accordingly

3
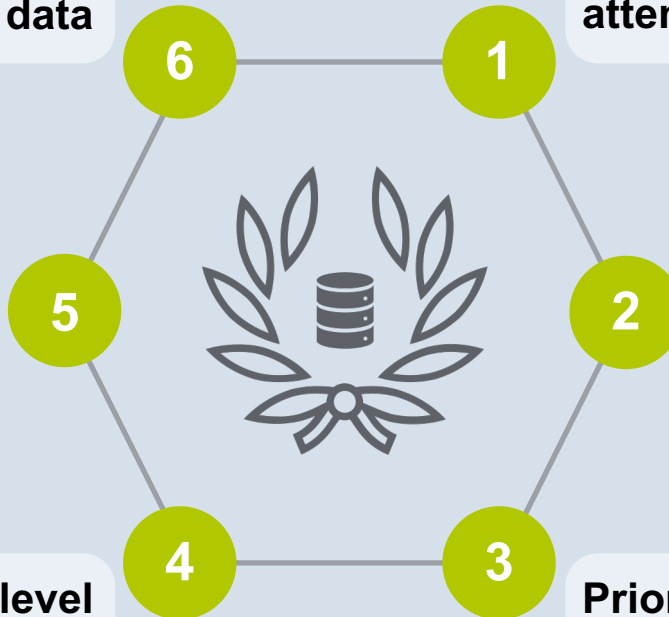
Drive Data-Governance Excellence (6 Steps)
① Secure top management's attention

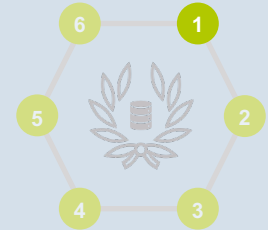Data Management Office (DMO) needs to engage with the C-suite

Step 1

Step 2

Step 3

Form a data-governance council within senior mgmt.

Define a set of data domains & select business executives to lead them

Drive Data-Governance Excellence (6 Steps)
② Integrate with primary transformation themes

Generate excitement for data

Secure top management's attention

6

1

Choose iterative and focused implementation

5

2 **Integrate with primary transformation themes**

Apply the right level of governance

4

3

Prioritize data assets and focus data leadership accordingly
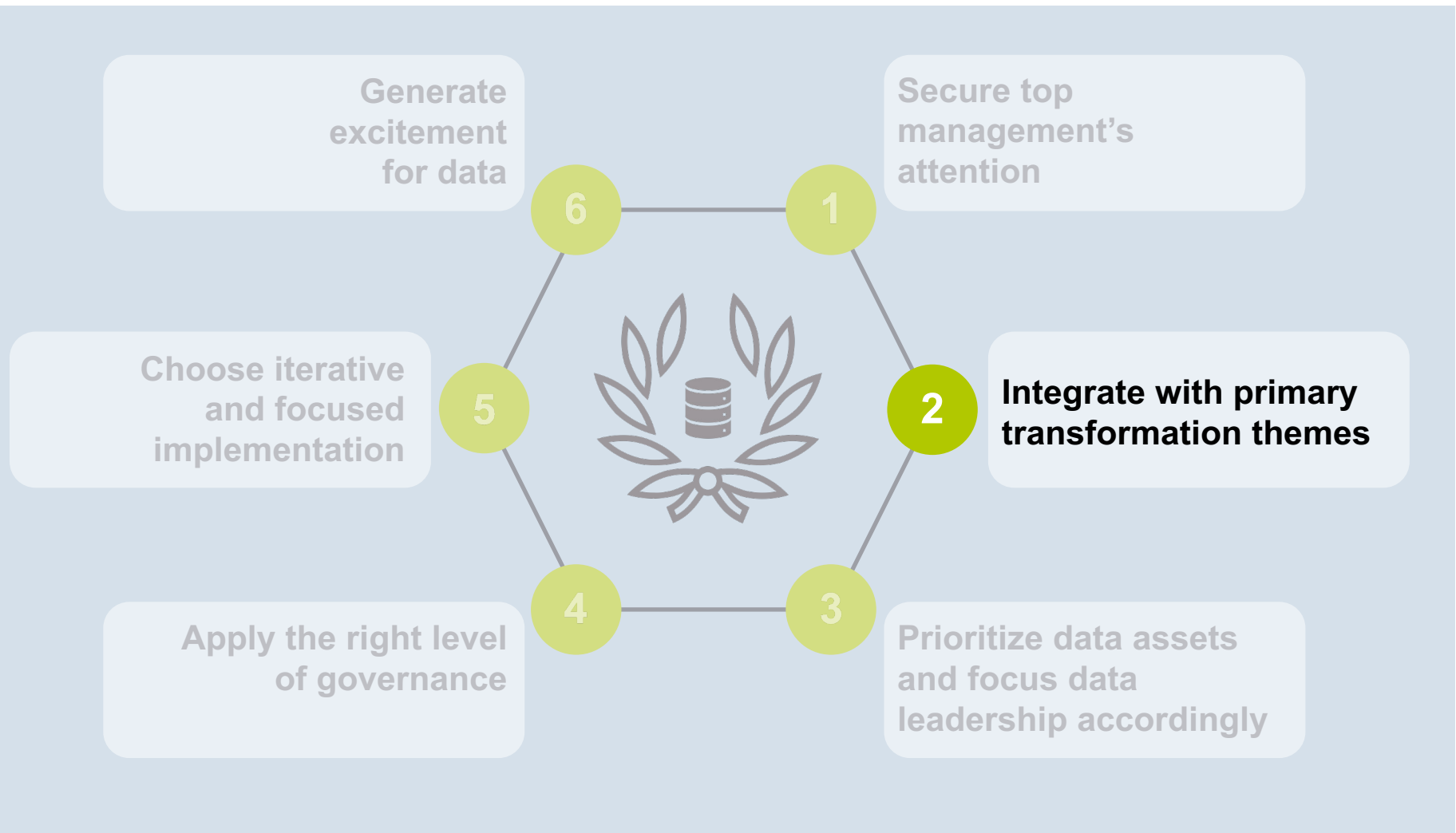
Drive Data-Governance Excellence (6 Steps)
③ Prioritize data assets and focus data leadership

Generate excitement for data

Secure top management's attention

6

1

5

2

Choose iterative and focused implementation

Integrate with primary transformation themes

4

3

Apply the right level of governance

Prioritize data assets and focus data leadership accordingly
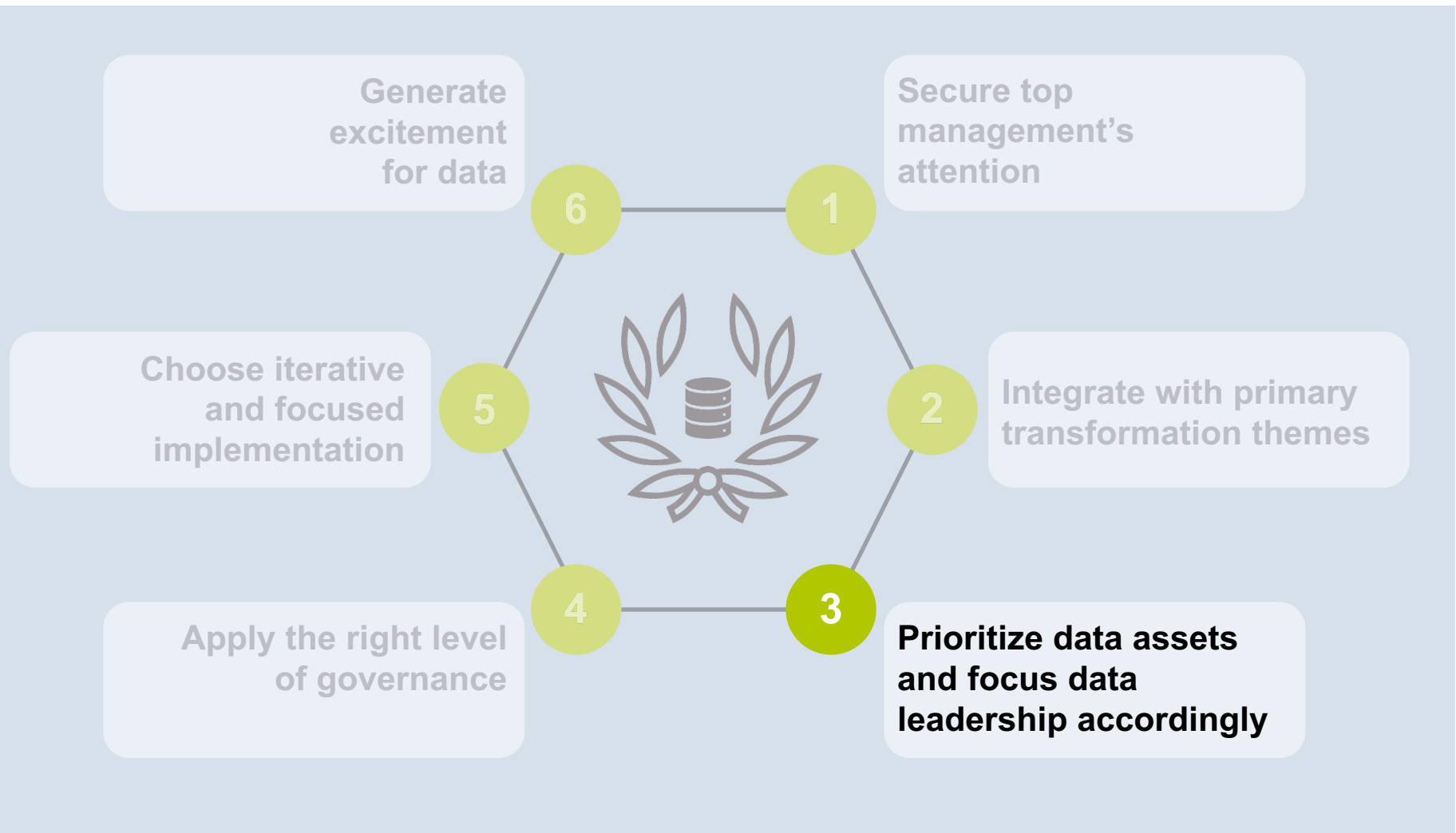
Drive Data-Governance Excellence (6 Steps)
④ Apply the right level of governance

high

Governance focused on mee-ting regulatory requirements

Comprehensive and balanced governance

↑ Level of data-related regulation in the industry

Targeted gover-nance focused on value creation

Governance focused on big value opportuni-ties & standards

low

low   high

→ Level of data complexity

5 Choose iterative and focused implementation

MANNHEIM
BUSINESS SCHOOL

Generate excitement for data

Secure top management's attention

6

1

Choose iterative and focused implementation

5

2

Integrate with primary transformation themes

Apply the right level of governance

4

3

Prioritize data assets and focus data leadership accordingly

Drive Data-Governance Excellence (6 Steps)
⑥ Generate excitement for data

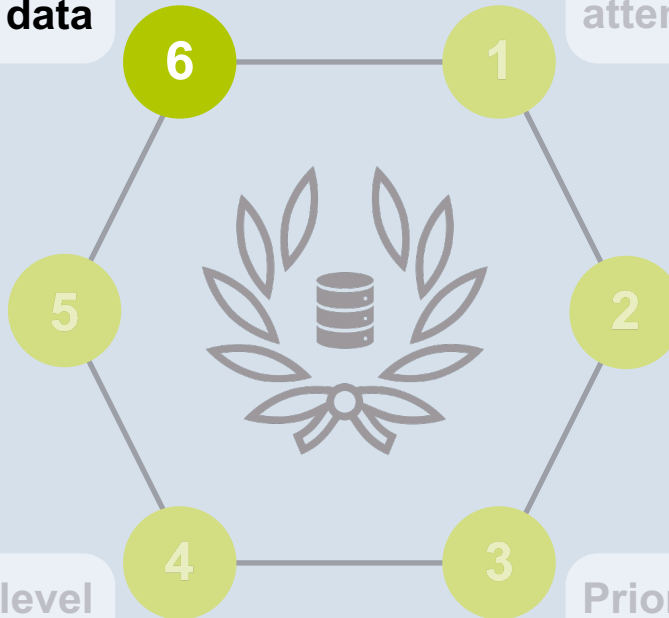**Generate excitement for data**

Secure top management's attention

Choose iterative and focused implementation

Integrate with primary transformation themes

Apply the right level of governance

Prioritize data assets and focus data leadership accordingly

6

1

2

3

4

5

Key Take Aways

Data Governance

Definition, components, scope, etc.

Hierarchical framework and key stakeholders

Implementation of data governance

Data governance excellence