MANAGING BIG DATA

PROF. DR. FLORIAN STAHL

Managing Big Data

Types of Data

Data Architecture

Master Data Management

Data Quality

Data Governance

Data Security

What Is Master Data Management?

Master data management (MDM) provides a **trusted, single version of the truth** on which to base decisions to ensure that organizations do not use multiple, potentially inconsistent versions of the same concept in different divisions.
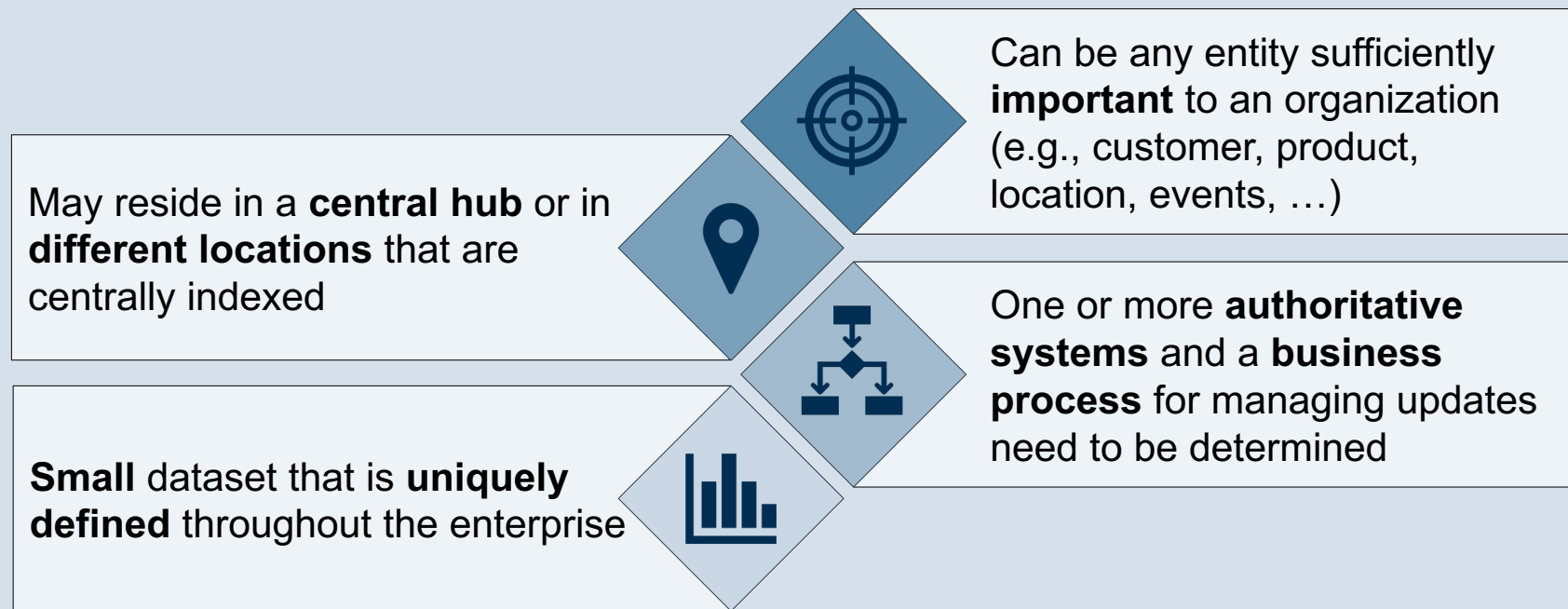
Eliminate duplicates

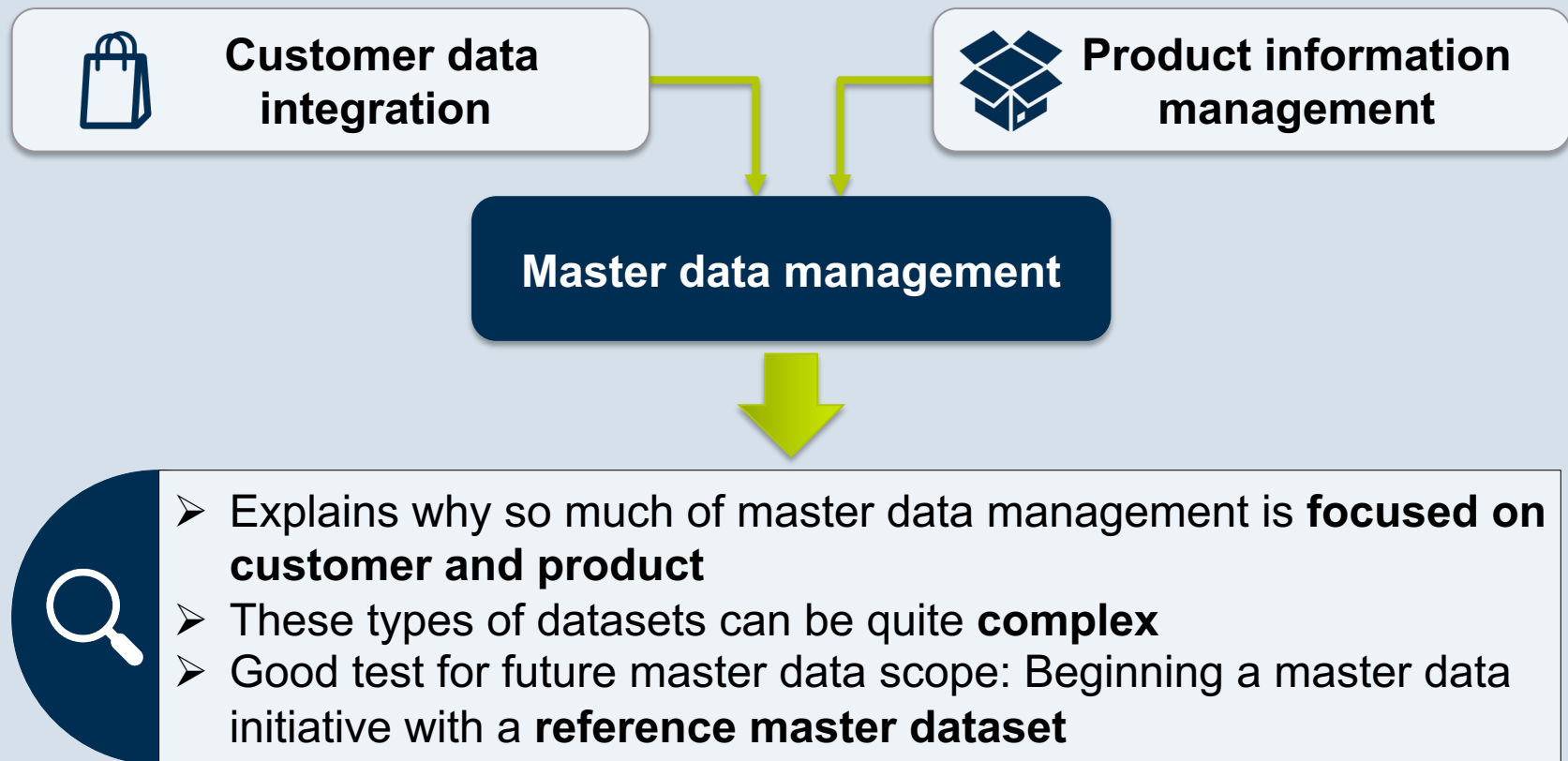Share a **consistent master dataset** across the organization

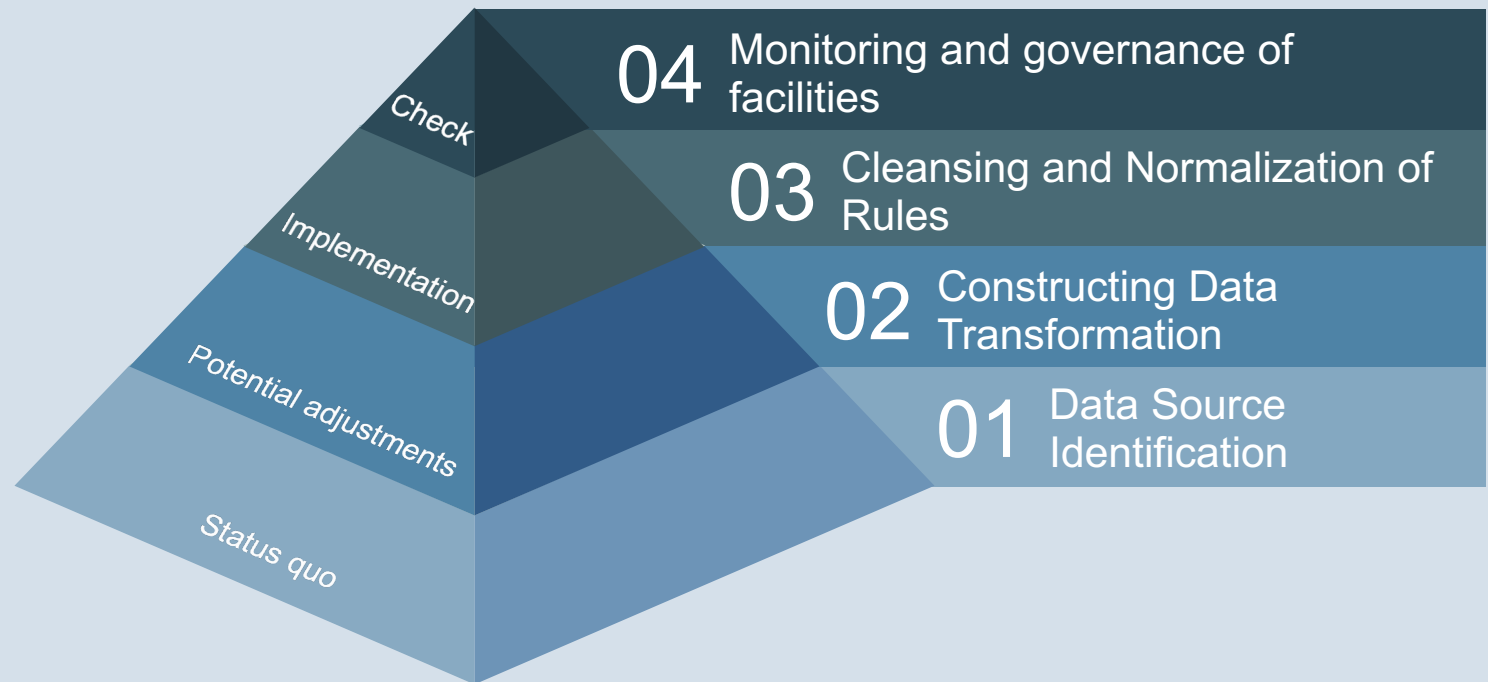Allow for **collaborative updates** to the master dataset

UNIVERSITY OF MANNHEIM
AACSB ACCREDITED
EQUIS ACCREDITED
ASSOCIATION AMBA ACCREDITED

What Is Master Data Management?

Can be any entity sufficiently **important** to an organization (e.g., customer, product, location, events, …)

May reside in a **central hub** or in **different locations** that are centrally indexed

One or more **authoritative systems** and a **business process** for managing updates need to be determined

Small dataset that is **uniquely defined** throughout the enterprise

What Is Master Data Management?

Customer data integration

Product information management

Master data management

- ➢ Explains why so much of master data management is **focused on customer and product**
- ➢ These types of datasets can be quite **complex**
- ➢ Good test for future master data scope: Beginning a master data initiative with a **reference master dataset**

Master Data Management Initiative

! A **Master Data Management Initiative** requires a centrally governed data model and a corresponding metadata repository in order to be successful.

04 Monitoring and governance of facilities

03 Cleansing and Normalization of Rules

02 Constructing Data Transformation

01 Data Source Identification

Check

Implementation

Potential adjustments

Status quo

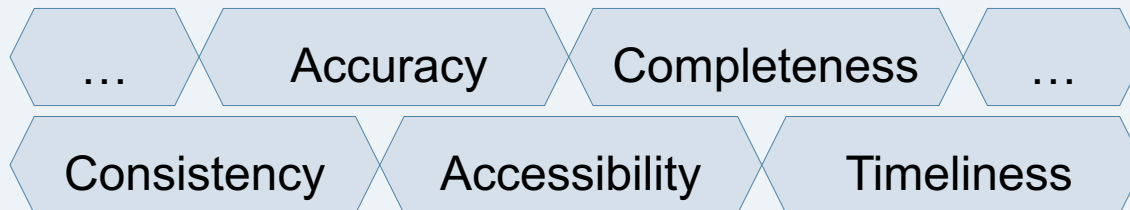# Data Quality and Master Data Management

Data Quality

- Defined as "**fitness for use**"
- Required level of quality of data depends on the **context**
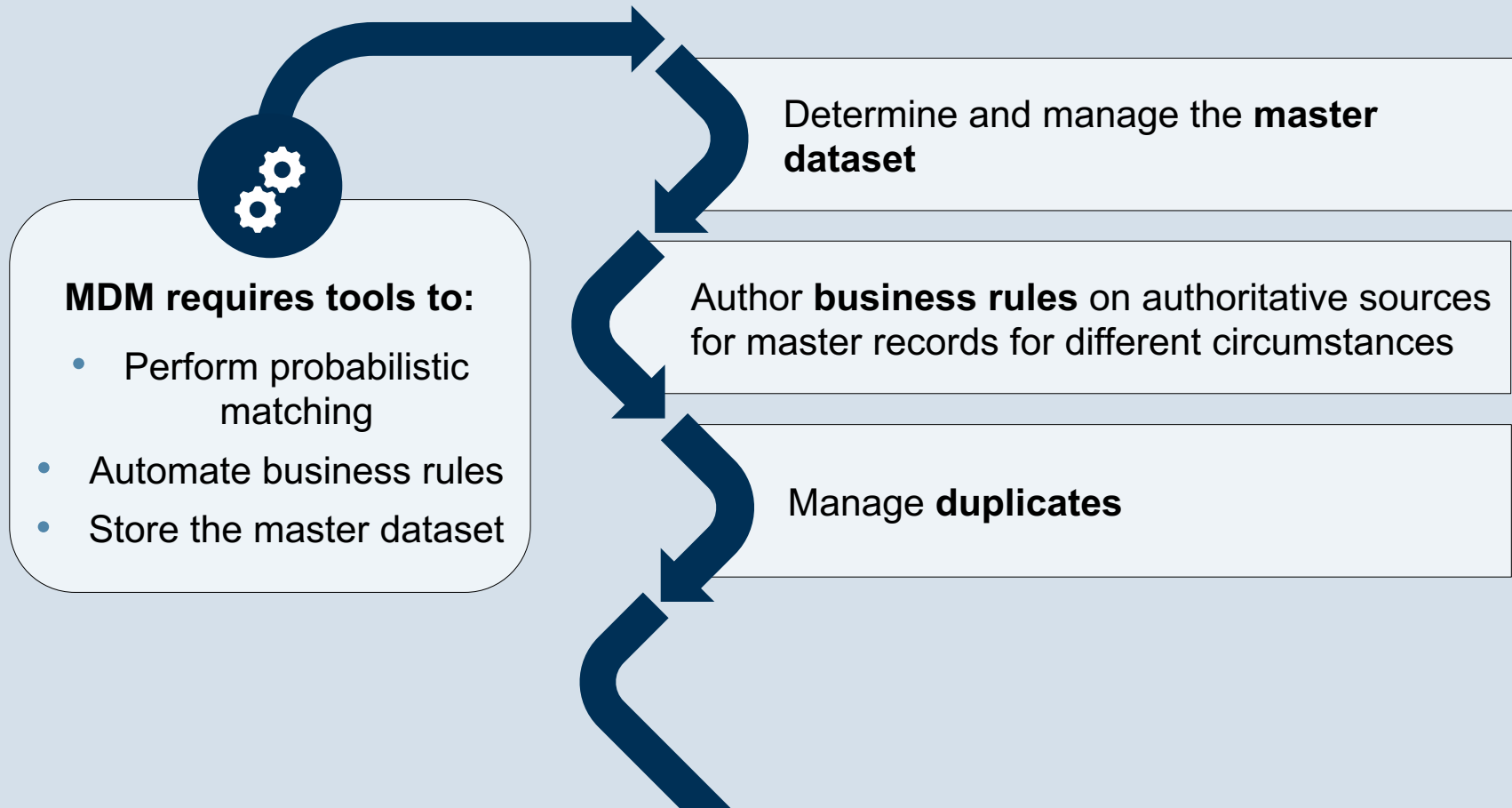- Encompasses **various aspects** or **criteria** for assessment

Data Integration

- Heavily **related** to **data quality**
- **Influences** data quality both **positively** and **negatively**

…	Accuracy	Completeness	…
Consistency	Accessibility	Timeliness	

Process of Master Data Management?

MDM requires tools to:

- Perform probabilistic matching
- Automate business rules
- Store the master dataset

Determine and manage the **master dataset**

Author **business rules** on authoritative sources for master records for different circumstances

Manage **duplicates**

Benefits of Master Data Management

Reduction or **elimination of duplicate records** in mission critical datasets

Availability of current, cleansed, **consistent master data** throughout the enterprise

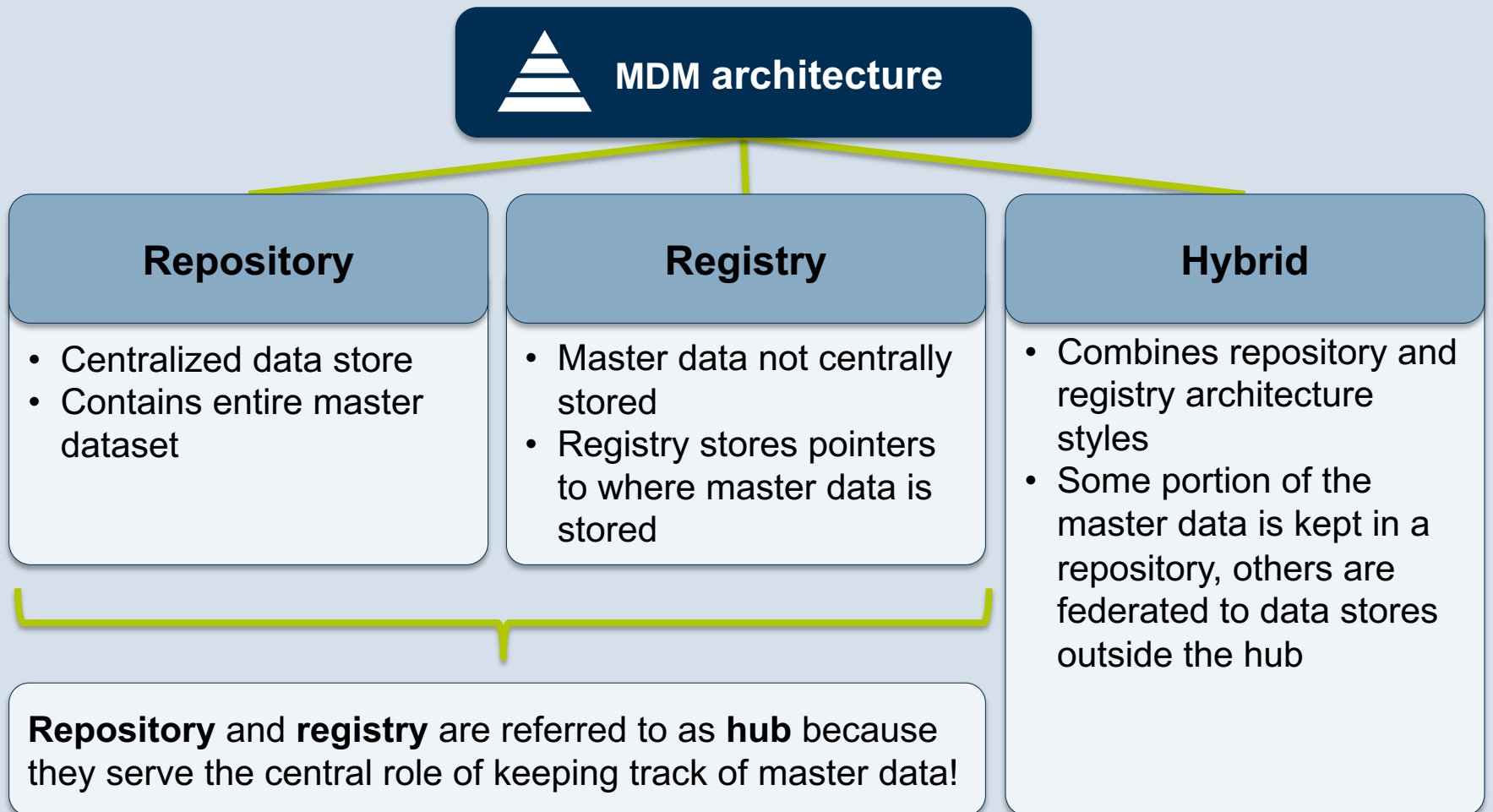Automated **probabilistic matching** and **application of business rules**

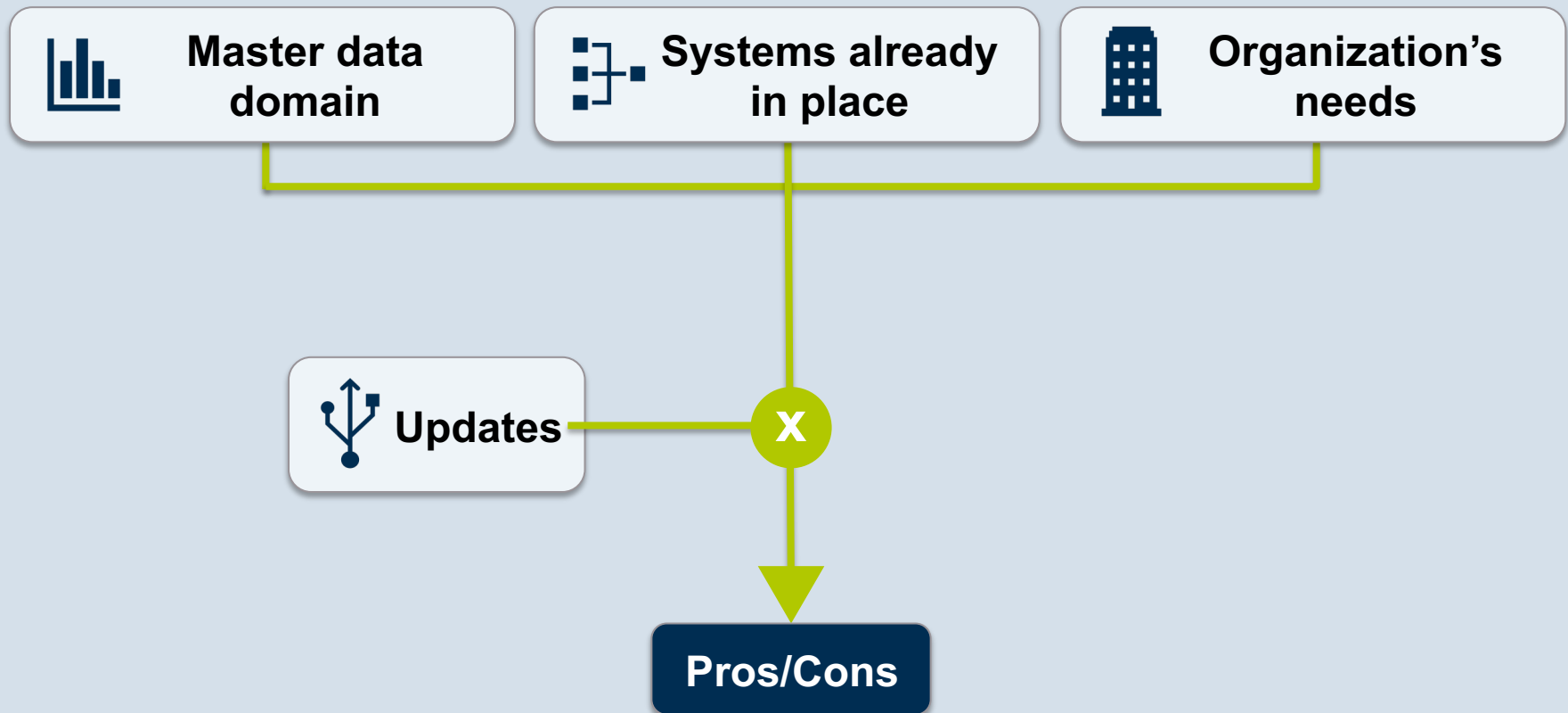Known, managed **authoritative sources** for master data

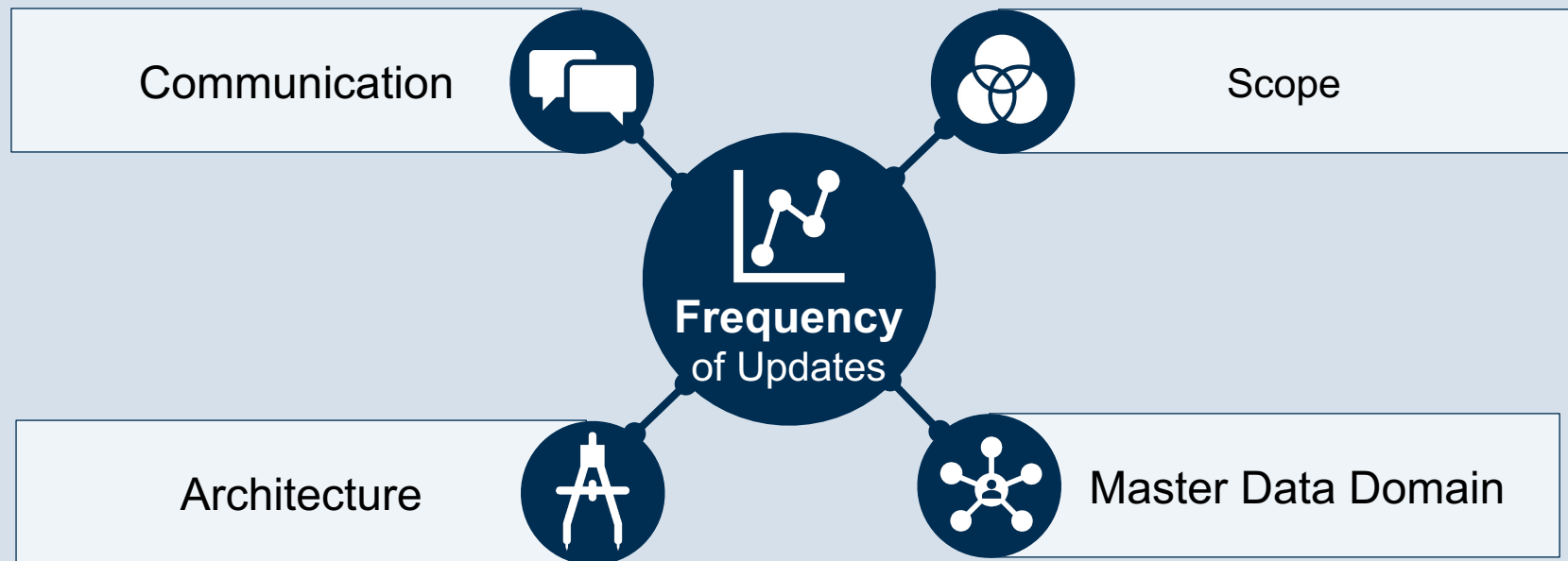Potential **collaborative management** of the master dataset
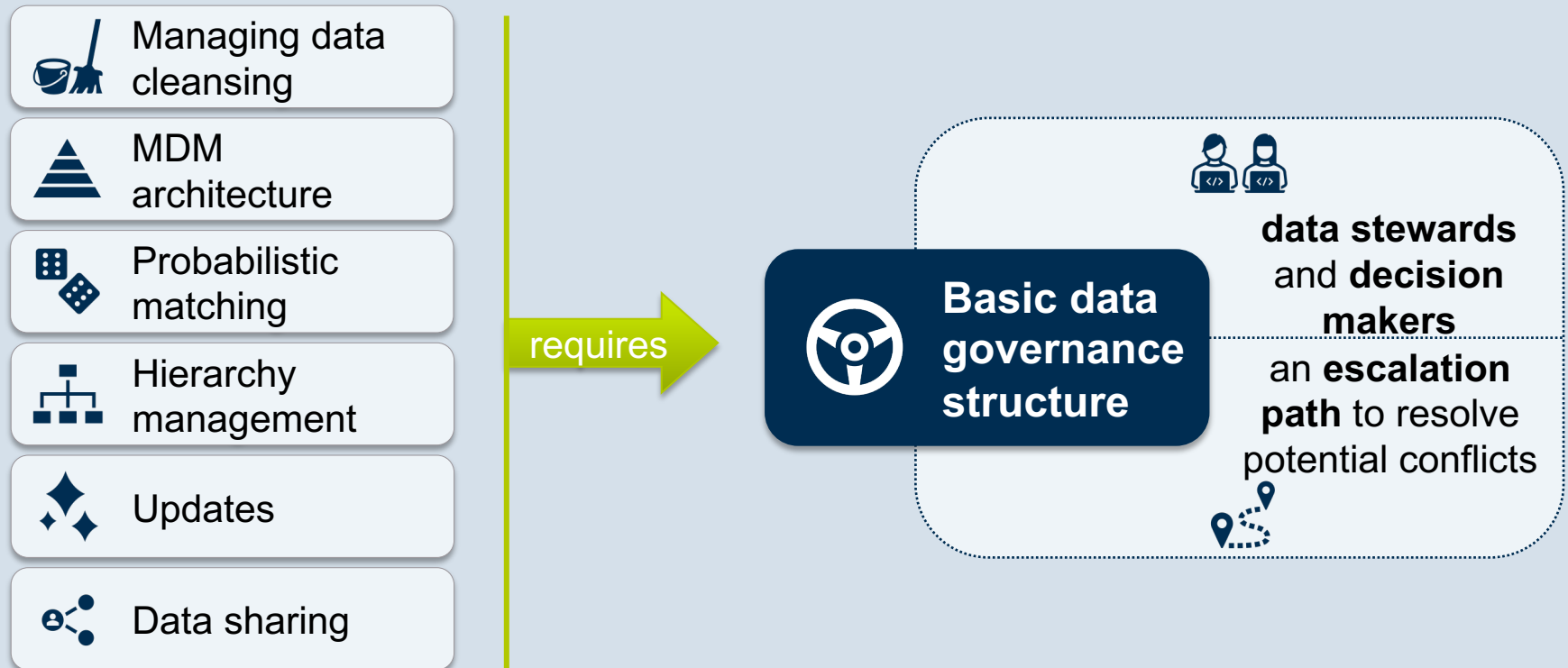
Master Data Management Framework

MDM architecture

Repository

- Centralized data store
- Contains entire master dataset

Registry

- Master data not centrally stored
- Registry stores pointers to where master data is stored

Hybrid

- Combines repository and registry architecture styles
- Some portion of the master data is kept in a repository, others are federated to data stores outside the hub

Repository and **registry** are referred to as **hub** because they serve the central role of keeping track of master data!

Master Data Management Framework

Master Data Management Framework

Master Data Management Framework
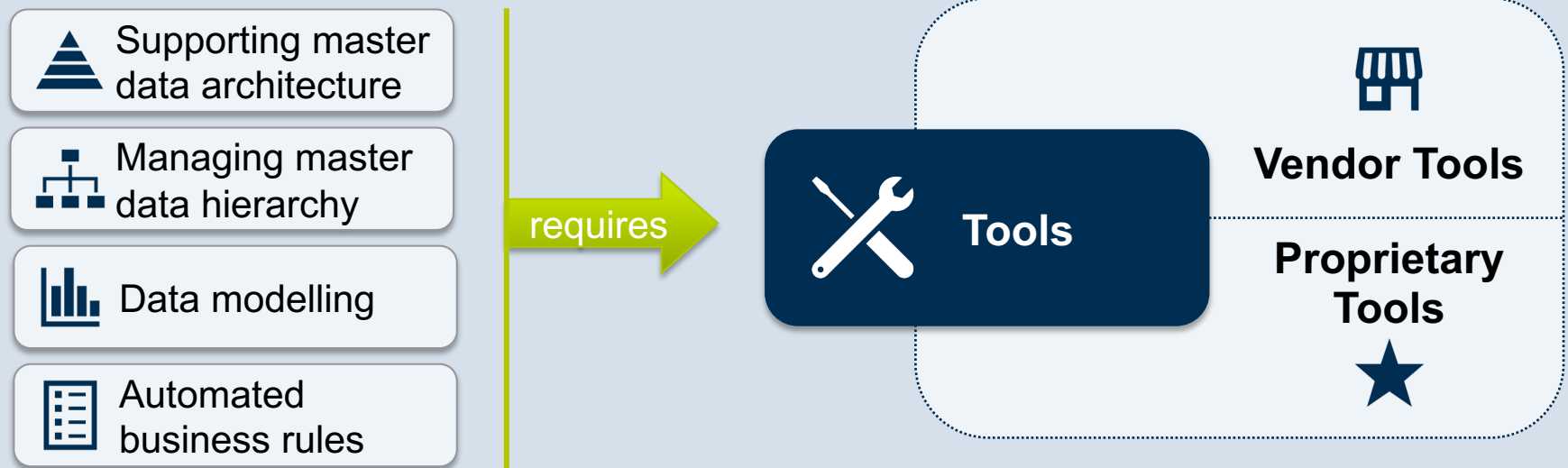

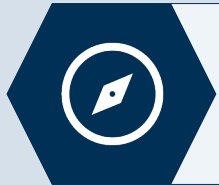Managing data cleansing

MDM architecture

Probabilistic matching

Hierarchy management

Updates

Data sharing

requires

Basic data governance structure

data stewards and **decision makers**

an **escalation path** to resolve potential conflicts

Implementing Master Data Management

Master Data Management Tools

Supporting master data architecture

Managing master data hierarchy

Data modelling

Automated business rules

requires
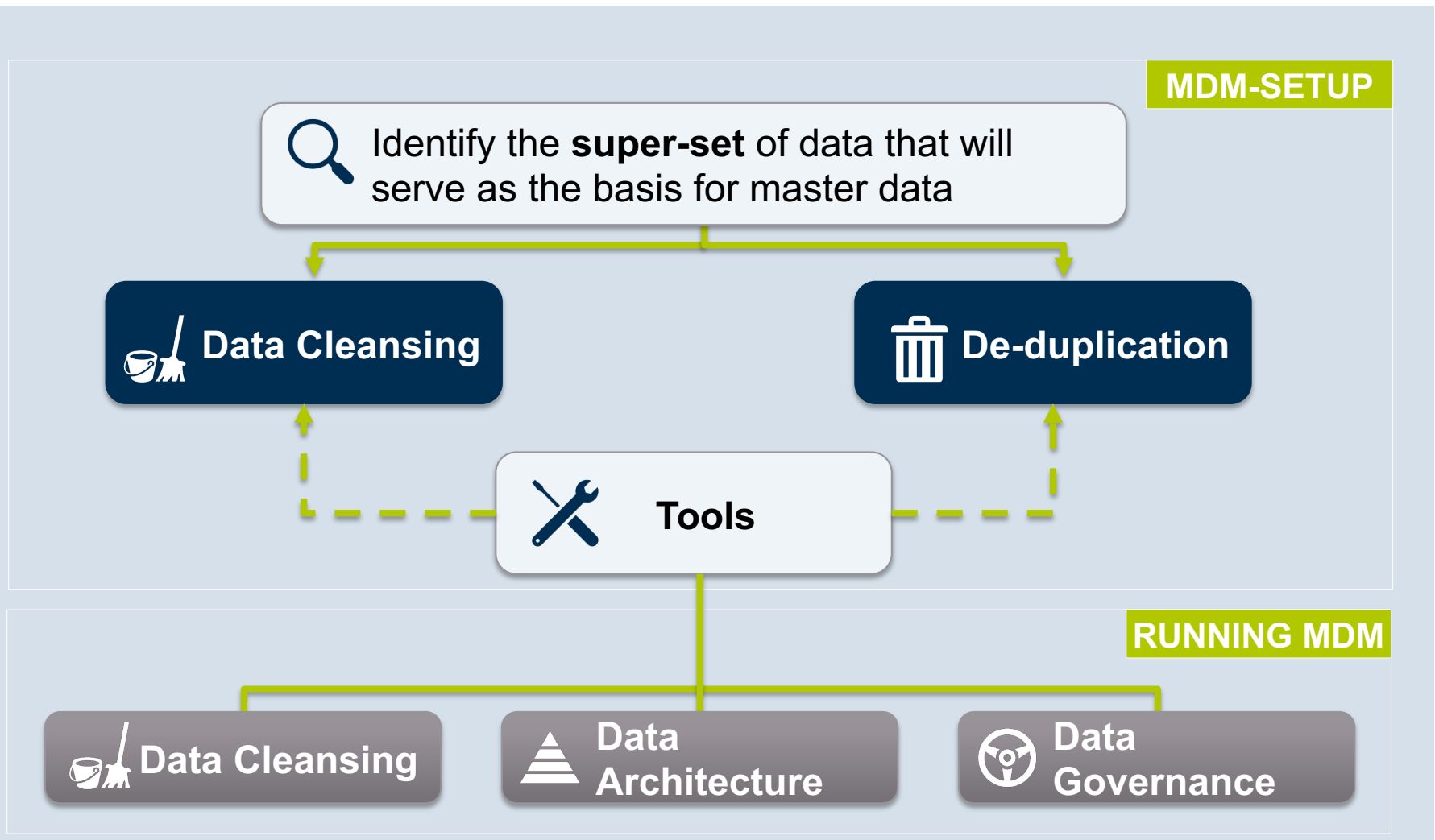
Tools

Vendor Tools

Proprietary Tools

Master Data Management Tools

Vendor tools are typically oriented **toward a particular domain** (most common orientations are customer and product)

Tools are trying hard to become multi-domain oriented

Tools typically **support all master data architecture styles** (repository, registry, and hybrid)

Some tools support **cloud integration** or include data from nontraditional sources

Other functionalities supported to differing degrees: data cleansing, data stewardship, and real-time MDM

Reference Data

Reference data is a special subset of master data that is **used for classification** throughout your entire organization.
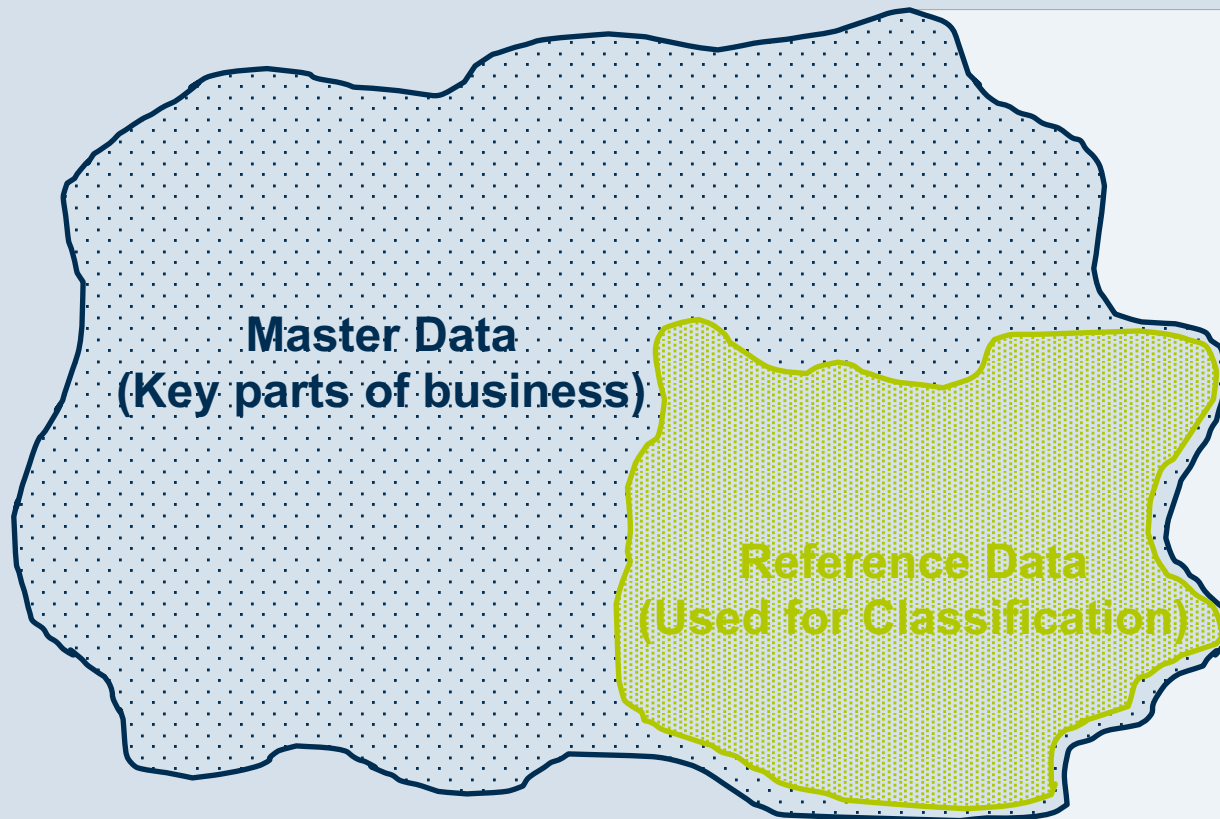
Multidomain Reference Data

- **Non-industry specific**
- Contains **multiple functional areas**
- Includes **various content types**

Real-time Reference Data

- More **industry specific**
- Typically used in **capital markets**
- Increasing importance with **IoT applications**

Reference Data vs. Master Data

**Master Data
(Key parts of business)**

**Reference Data
(Used for Classification)**

- **Reference Data** is a **subset** of **Master Data**
- **Changes** in **reference data** occur **as a reflection** of **business processes**
- **Master data** changes occur **within** standard **business processes**
- **Reference data** is **rarely changed**

Examples of Reference Data

Why is Reference Data important?

Domains

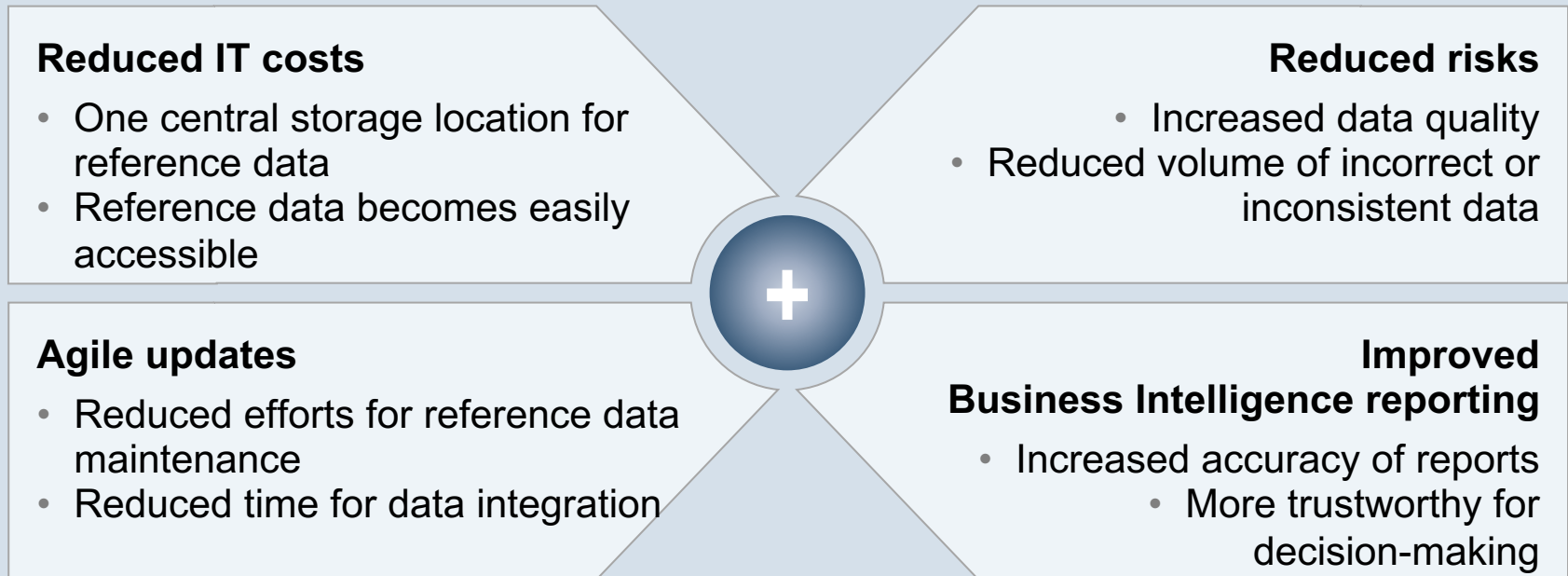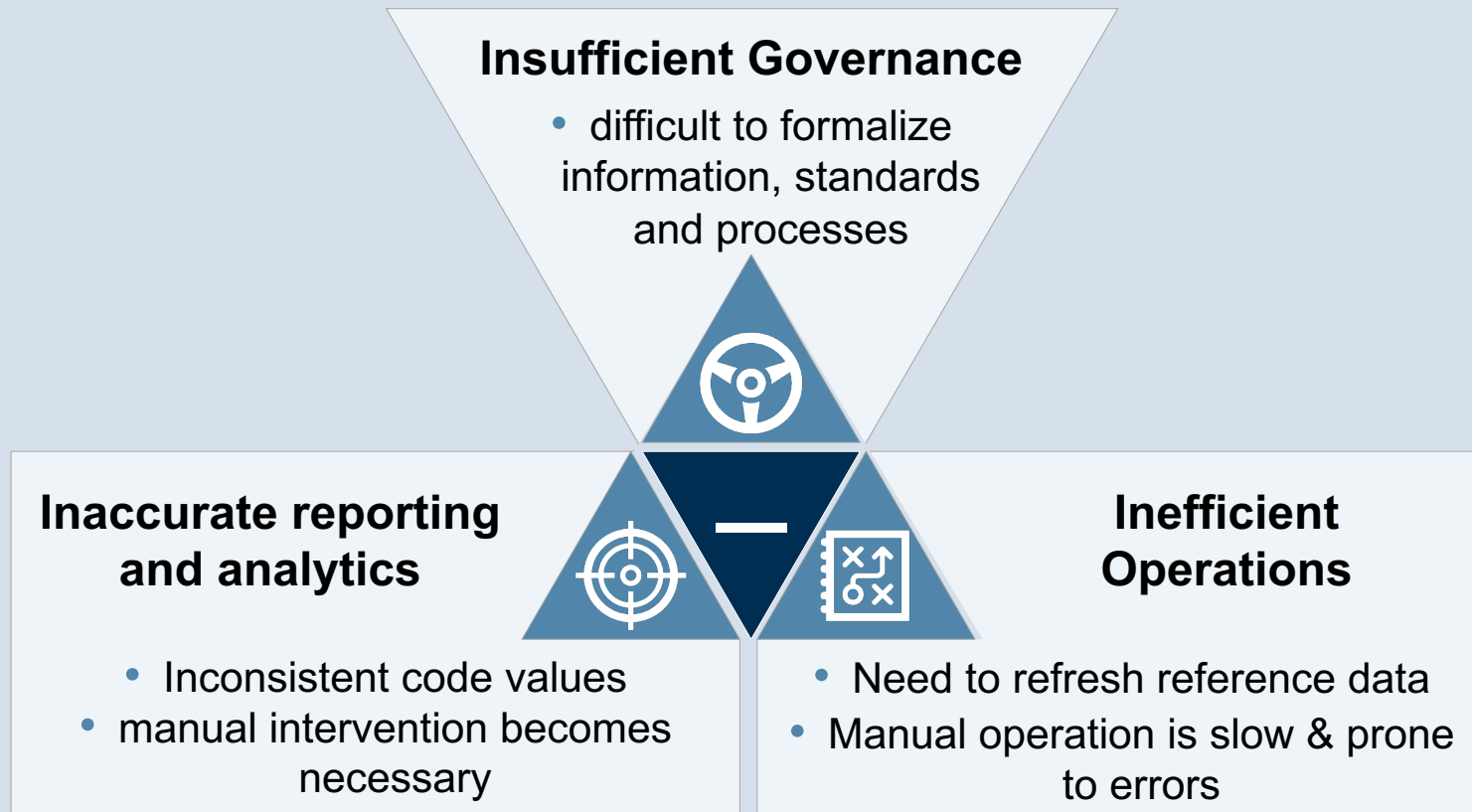**Applications**

As **reference data connects domains** and **applications**, these connections can be **retraced through reference data management**, e.g., to handle:

- Governance issues
- Versioning control
- Custom hierarchies

Benefits of Managing Reference Data

Reduced IT costs
- One central storage location for reference data
- Reference data becomes easily accessible

Reduced risks
- Increased data quality
- Reduced volume of incorrect or inconsistent data

Agile updates
- Reduced efforts for reference data maintenance
- Reduced time for data integration

Improved Business Intelligence reporting
- Increased accuracy of reports
- More trustworthy for decision-making

Risks of Poor Reference Data Management

Insufficient Governance
- difficult to formalize information, standards and processes

Inaccurate reporting and analytics
- Inconsistent code values
- manual intervention becomes necessary

Inefficient Operations
- Need to refresh reference data
- Manual operation is slow & prone to errors

Capabilities for Managing Reference Data

Data Governance

Data Lineage

Workflows

Stewardship

Policy Management

Who is dealing with Reference Data?

Banking Financial Service	• Complementing trades or transactions • Essential to the industry
Government/ Public Sector	• Tracking of public records • Maintaining consistency across several organizations
Healthcare	• Ensuring accurate associations (e.g., between lab results and patients)
Travel and Hospitality	• Organization of large operations • Maintaining exceptional customer experiences
eCommerce	• Consistent linkage of transaction codes, customer data, etc. • Ensuring security and efficiency

Key Take Aways

Master Data Management

| MDM and its links to data quality | MDM framework and tools | Reference Data - Relevance, benefits, and risks | Capabilities of Reference Data management |