DATA QUALITY

DATA QUALITY FOUNDATIONS

WHAT IS DATA QUALITY?

Data quality (DQ) is the degree to which data **meets** or **exceeds** business requirements (i.e., the extent to which data is “fit for purpose”).

It involves the **planning** and **implementation** of quality management techniques to **measure**, **assess**, and **improve** the fitness of data for use within an organization.

“Improving data quality leads to improved decision-making throughout the enterprise. The more high-quality data you have, the more confidence you will have in your decision-making.”

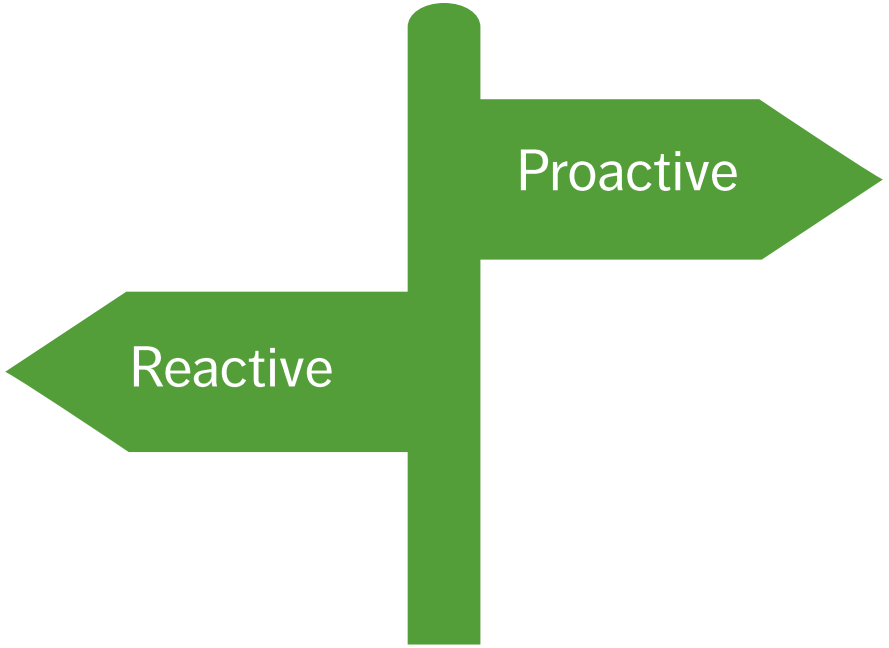
WHY IS DATA QUALITY IMPORTANT?

Data has **intrinsic value** due to its **information content**. The impacts of poor-quality data can include:

- Bad decisions, wasted resources, lowered performance chasing after issues.
- Escalating costs to remediate if issues are not caught early.
- Lack of trust in data for decision-making.
- Outdated and meaningless data bloat.
- Challenges with data sharing, integration and access to historical data.
- Poor-quality data is detrimental to analytics.

Data is the currency of **control:** we can't control what we don't know.

DATA QUALITY IS REACTIVE AND PROACTIVE



Quality Assurance

Preventing data quality issues ***before*** they impact operations and programs

Quality Control

Identifying and fixing data quality issues ***after*** they occur

A data quality program needs elements of both **control** and **assurance** to be fully effective.

DATA QUALITY DIMENSIONS

To implement data quality, we typically introduce categories called "**dimensions**".

These dimensions are **measurement attributes** of data, which can be individually **assessed**, **interpreted**, and **improved**.

We can use them to help build **data quality dashboards**.

They can also help us group **issues** and **risks**, helping us:

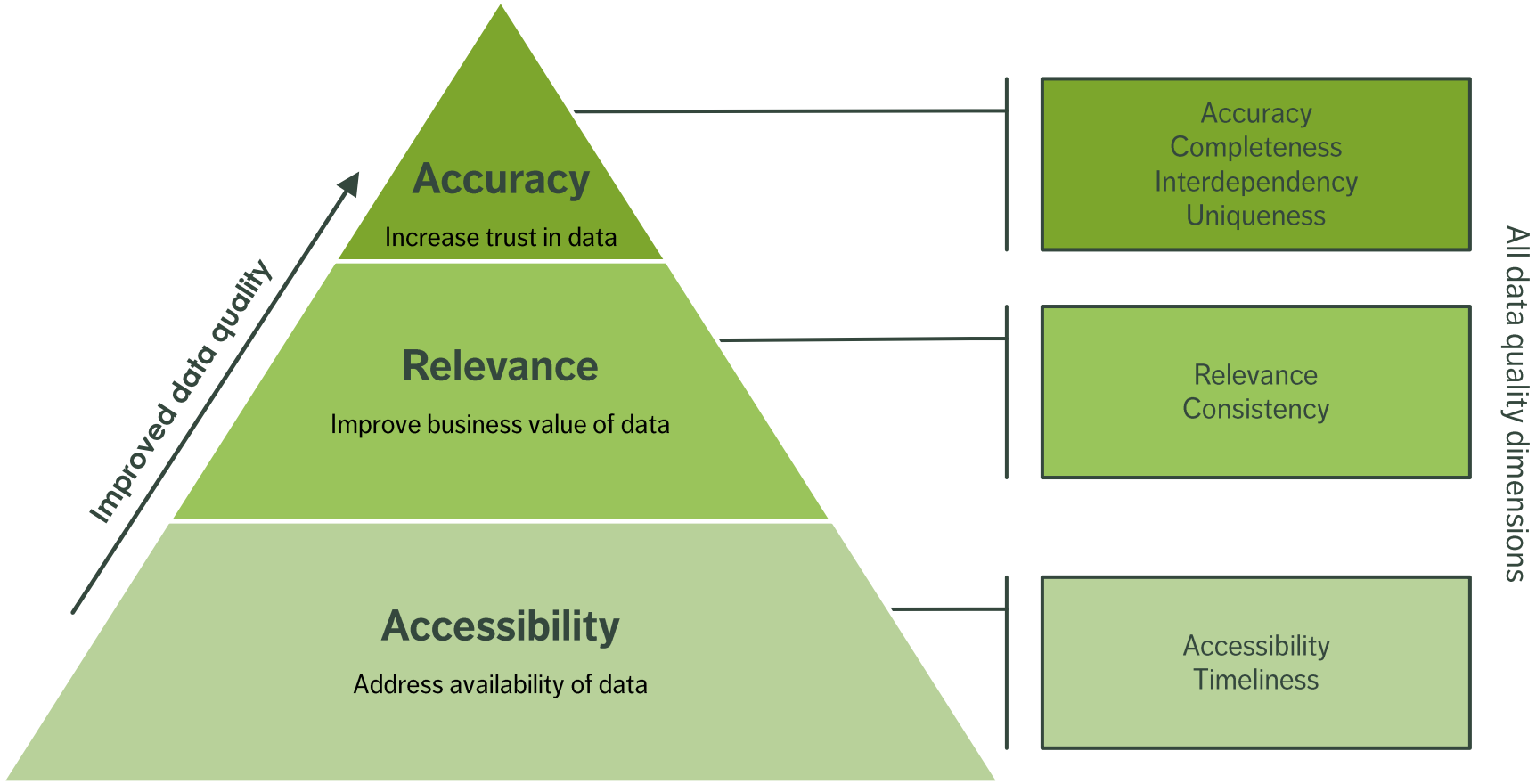
- perform **trend analyses**;
- identify underlying, **systemic issues**;
- set-up **data quality testing programs**, etc.

DATA QUALITY DIMENSIONS

Typical set of data quality dimensions

Accessibility	Business processes and consumers can access and use the data asset
Timeliness	Data values are sufficiently up-to-date for business processes and consumers needs
Relevance	Data asset is of value to and used by business processes and data consumers
Consistency	Data representations are the same within and across data assets and repositories
Accuracy	Data accurately represents a real-world entity, object, concept, etc.
Completeness	Data asset has no missing data values
Interdependency	Relationships between data elements are preserved within or across data assets
Uniqueness	Data representations are not duplicated within or across data assets

DATA QUALITY DIMENSIONS



DATA QUALITY

DATA QUALITY FOUNDATIONS