DATA QUALITY

DATA QUALITY FRAMEWORK PART 1

DATA QUALITY IS A PROCESS

There are three focus areas in addressing data quality:

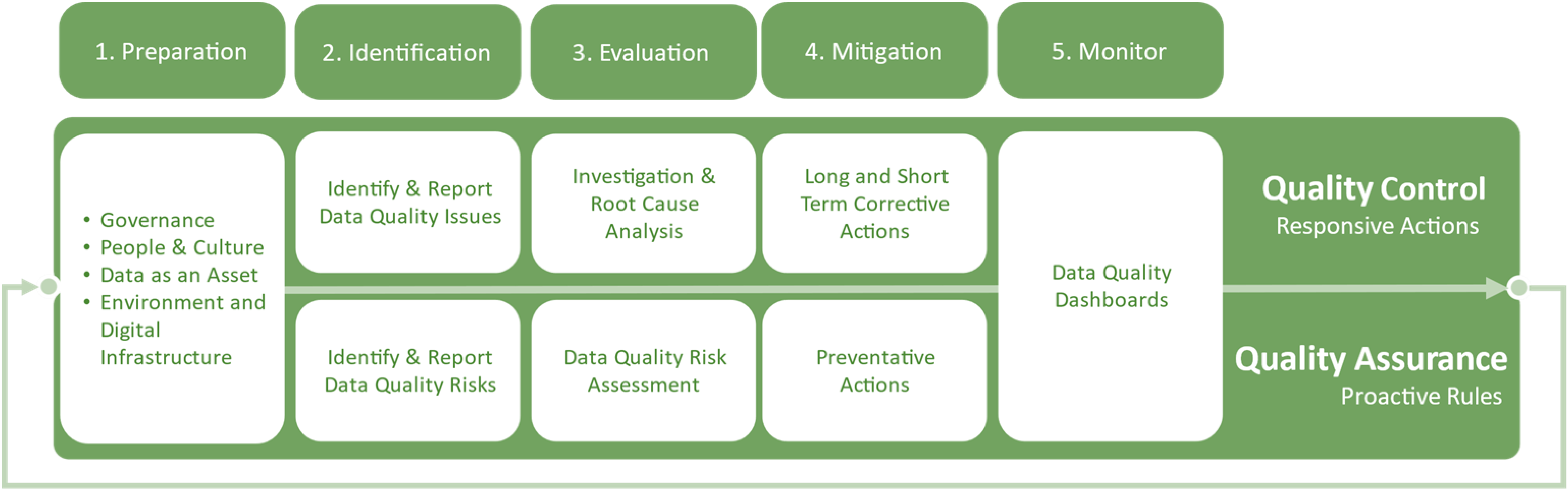
1. identify and mitigate **existing** data quality issues through quality control (e.g., testing a database to identify incorrect values then replacing them);
2. identify sources of high risk that could **create** quality issues and mitigate those risks through quality assurance (e.g., replacing a free form text field on a new software app with a dropdown list), and
3. track and **monitor** all known data quality issues and report them on a regular basis through quality monitoring (e.g., creating a list of all known issues and monitoring when they get fixed).

DATA QUALITY IS A PROCESS

We cannot implement a **data quality program** all at once, so we typically break it down to 5 stages:

1. preparation
2. issue and risk identification
3. issue and risk evaluation
4. issue and risk mitigation
5. ongoing monitoring

DATA QUALITY IS A PROCESS



1. Preparation

- Governance
- People & Culture
- Data as an Asset
- Environment and Digital Infrastructure

2. Identification

Identify & Report Data Quality Issues

Identify & Report Data Quality Risks

3. Evaluation

Investigation & Root Cause Analysis

Data Quality Risk Assessment

4. Mitigation

Long and Short Term Corrective Actions

Preventative Actions

5. Monitor

Data Quality Dashboards

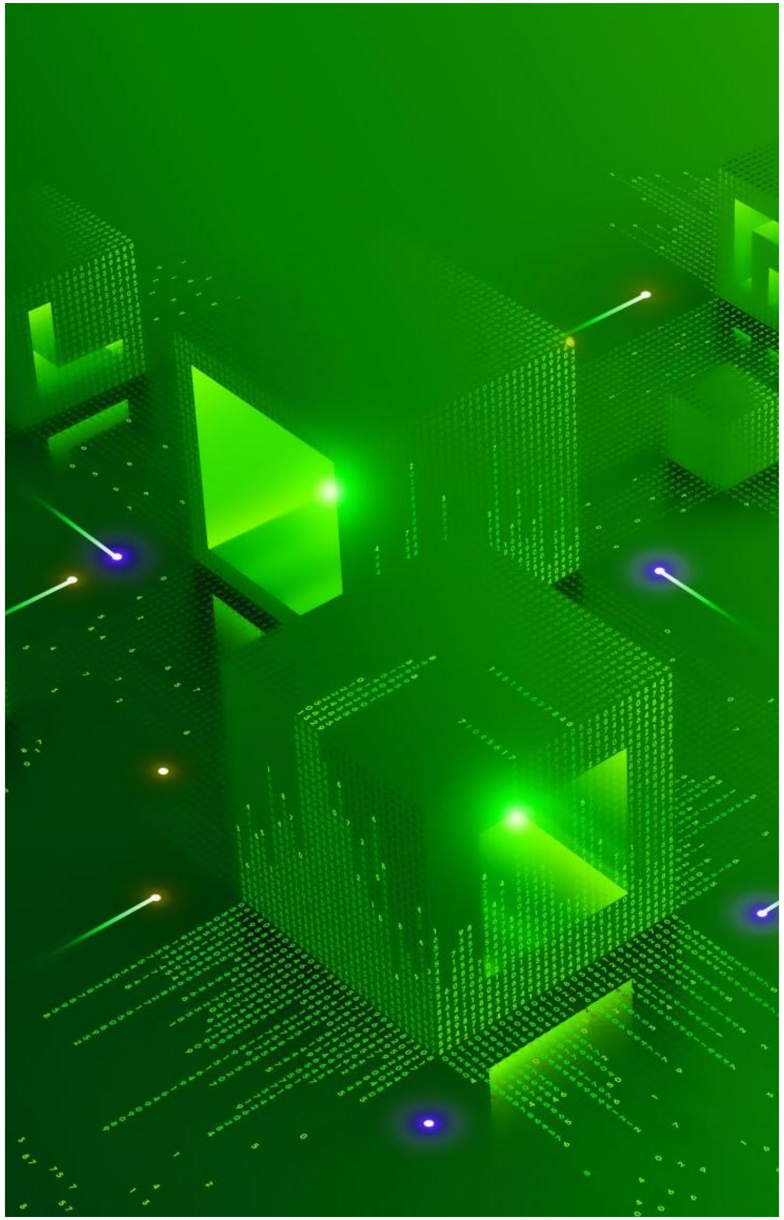
Quality Control

Responsive Actions

Quality Assurance

Proactive Rules

data-action-lab.com



STAGE 1: PREPARATION

We can improve data quality by implementing programs such as:

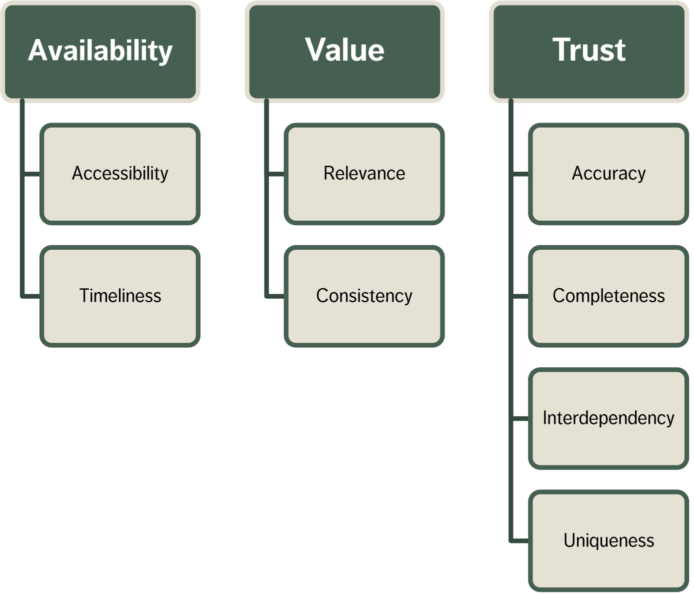
- **People & Culture** (data literacy, culture, and communication)
- **Environment & Digital Infrastructure** (tools, data asset catalogue)
- **Data Management** (metadata, reference/master data, dimensions & rules)
- **Governance** (roles & responsibilities, DQ planning, process definition)

Although it isn't critical to have all the above activities in place before starting on Data Quality, they do make a significant impact on the effectiveness of all DQ activities.

STEP 2: IDENTIFICATION

The second step in the process is to identify **data quality issues**. Currently-existing issues are called **data quality non-conformances**; issues that are yet to appear are known as **data quality risks**.

- DQ dimensions identify data attributes that can be used to measure data quality (often defined in an organization's **Data Quality Framework**).
- Business rules define the **business requirements** for data and how **data quality tests** are performed.
- Data quality metrics track the results of these tests over time.


Availability
Accessibility
Timeliness
Value
Relevance
Consistency
Trust
Accuracy
Completeness
Interdependency
Uniqueness

STAGE 2: IDENTIFICATION METHODS

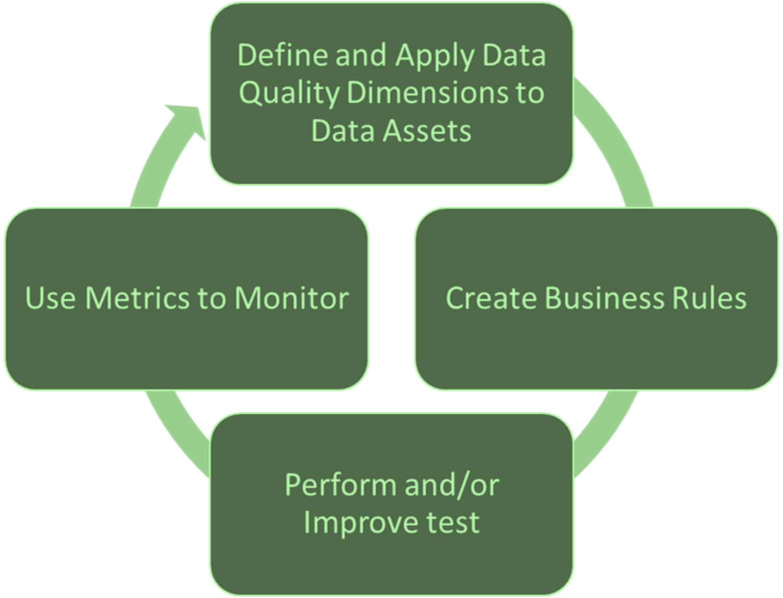
Methods for identifying **DQ non-conformances** and **DQ risks** include:

Quality Control

- DQ testing using software
- systems, process, and procedure auditing
- data consumer feedback

Quality Assurance

- creation of risk register



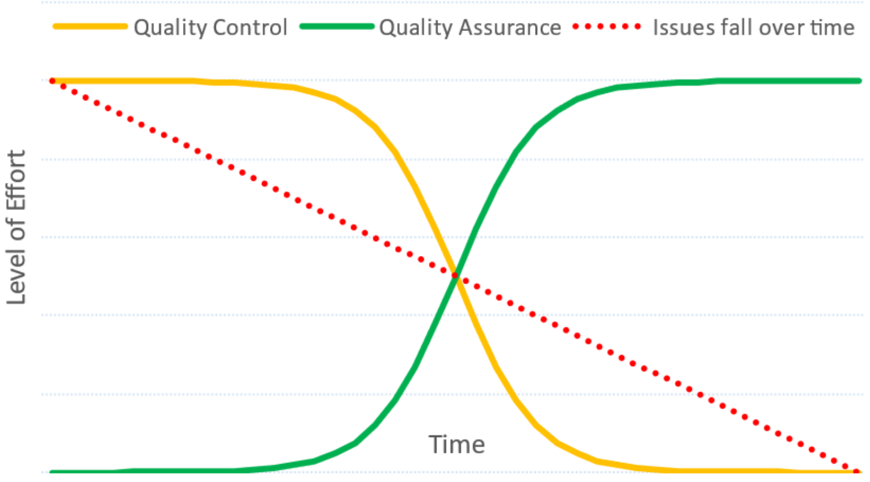
Define and Apply Data Quality Dimensions to Data Assets

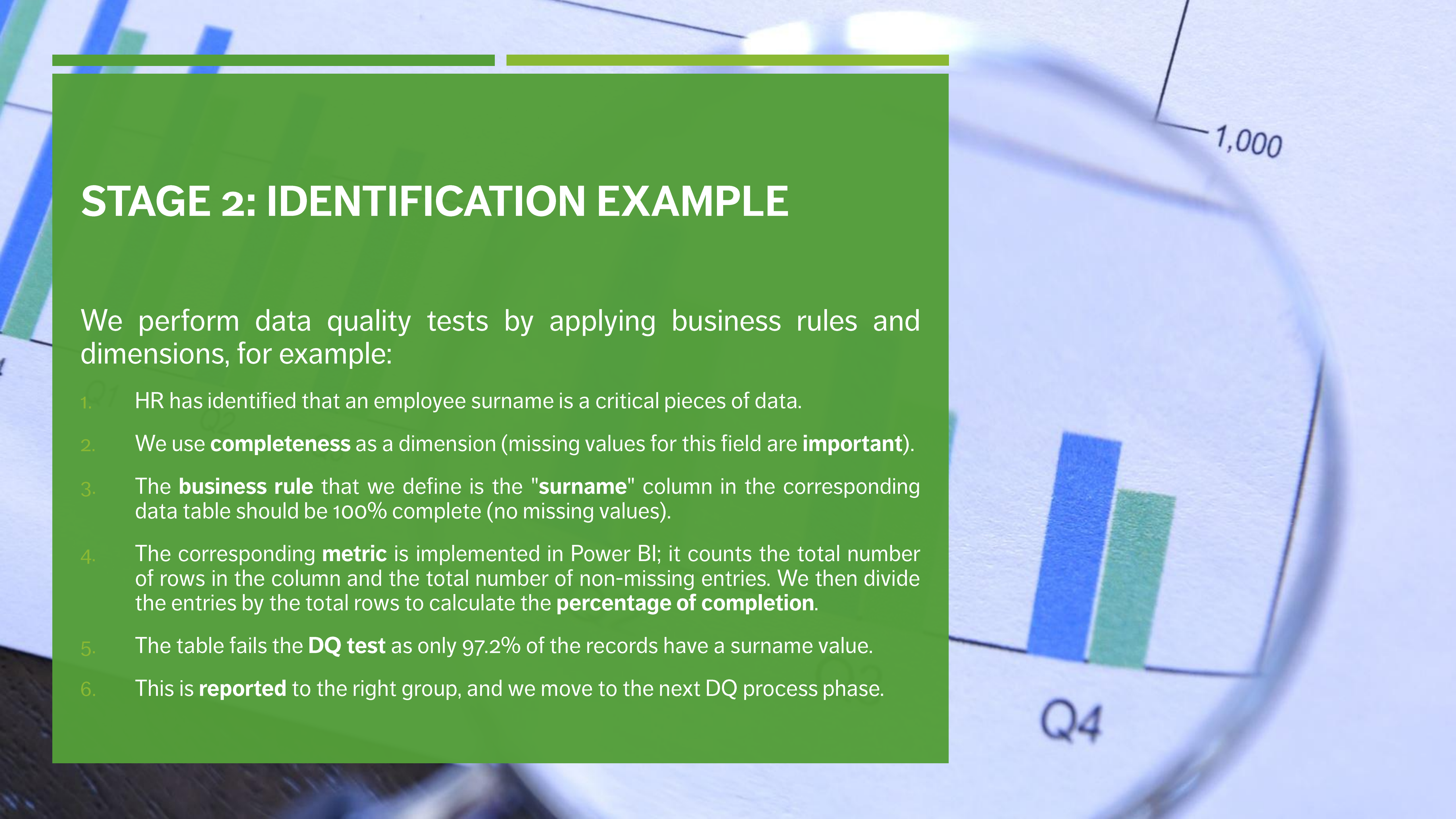
Create Business Rules

Perform and/or Improve test

Use Metrics to Monitor

Quality Assurance vs Quality Control over time





STAGE 2: IDENTIFICATION EXAMPLE

We perform data quality tests by applying business rules and dimensions, for example:

1. HR has identified that an employee surname is a critical pieces of data.
2. We use **completeness** as a dimension (missing values for this field are **important**).
3. The **business rule** that we define is the "**surname**" column in the corresponding data table should be 100% complete (no missing values).
4. The corresponding **metric** is implemented in Power BI; it counts the total number of rows in the column and the total number of non-missing entries. We then divide the entries by the total rows to calculate the **percentage of completion**.
5. The table fails the **DQ test** as only 97.2% of the records have a surname value.
6. This is **reported** to the right group, and we move to the next DQ process phase.

DATA QUALITY

DATA QUALITY FRAMEWORK PART 1