



# Data Quality Best Practices

Donna Burbank and Nigel Turner  
Global Data Strategy, Ltd.

August 22<sup>nd</sup>, 2024



The Business of Data

# Donna Burbank



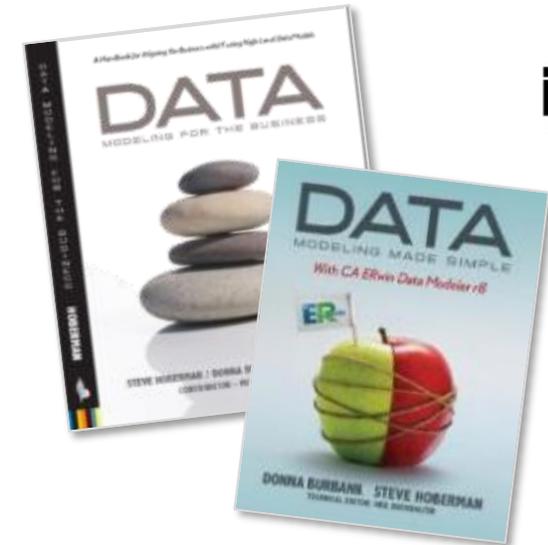
Donna is a recognized industry expert in data management with over 25 years of experience in data strategy, data governance, data modeling, metadata management, and enterprise architecture. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, and business leadership.

She is currently the Managing Director at Global Data Strategy, Ltd., an international data management consulting company that specializes in the alignment of business drivers with data-centric technology.

In past roles, she has served in key brand strategy and product management roles for several of the leading data management products in the market.

As an active contributor to the data management community, she is a long time DAMA International member, contributor to the DMBOK 2.0, Past President and Advisor to the DAMA Rocky Mountain chapter, and was awarded the Excellence in Data Management Award from DAMA International.

She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored several books and is a regular contributor to industry publications. She can be reached at [donna.burbank@globaldatastrategy.com](mailto:donna.burbank@globaldatastrategy.com)  
Donna is based in Boulder, Colorado, US.





## Nigel Turner

- Worked in Information Management (IM) and related areas for over 30 years. Experience has embraced Data Governance, Information Strategy, Data Quality, Master Data Management & Business Intelligence.
- Principal Consultant, EMEA, for Global Data Strategy, Ltd.

- Spent much of his career in British Telecommunications Group (BT) where he led a series of enterprise-wide IM & data governance initiatives.
- Also been VP of Information Management Strategy at Harte Hanks Trillium Software, and Principal Consultant at FromHereOn and IPL.



- Nigel is very active in professional Data Management organizations and is an elected Data Management Association (DAMA) UK Committee member.
- He was the joint winner of DAMA International's 2015 Community Award for the work he initiated and led in setting up a mentoring scheme in the UK where experienced DAMA professionals coach and support newer data management professionals.
- Nigel is based in Cardiff, Wales, UK and can be reached at [nigel.turner@globaldatastrategy.com](mailto:nigel.turner@globaldatastrategy.com)

# DATAVERSITY Data Architecture Strategies

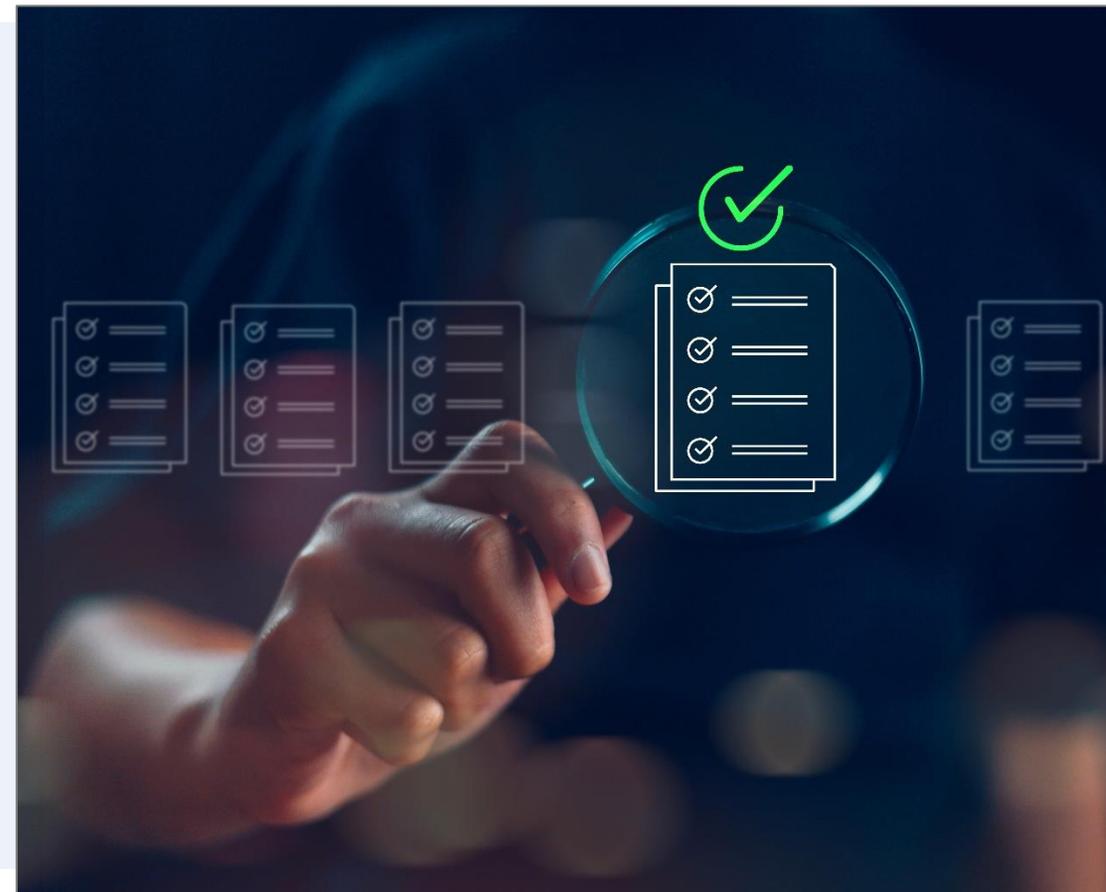
## This Year's Lineup

- **January** Emerging Trends in Data Architecture – What's the Next Big Thing?
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Master Data Management - Aligning Data, Process, and Governance
- **April** How do Data Governance & Data Architecture Support Each Other?
- **May** The Role of the Chief Data Officer (CDO) in Business Transformation
- **June** What Does It Mean to be a Data-Driven Organization?
- **July** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Enterprise Architecture vs. Data Architecture
- **December** The Business Benefits of Data Modeling



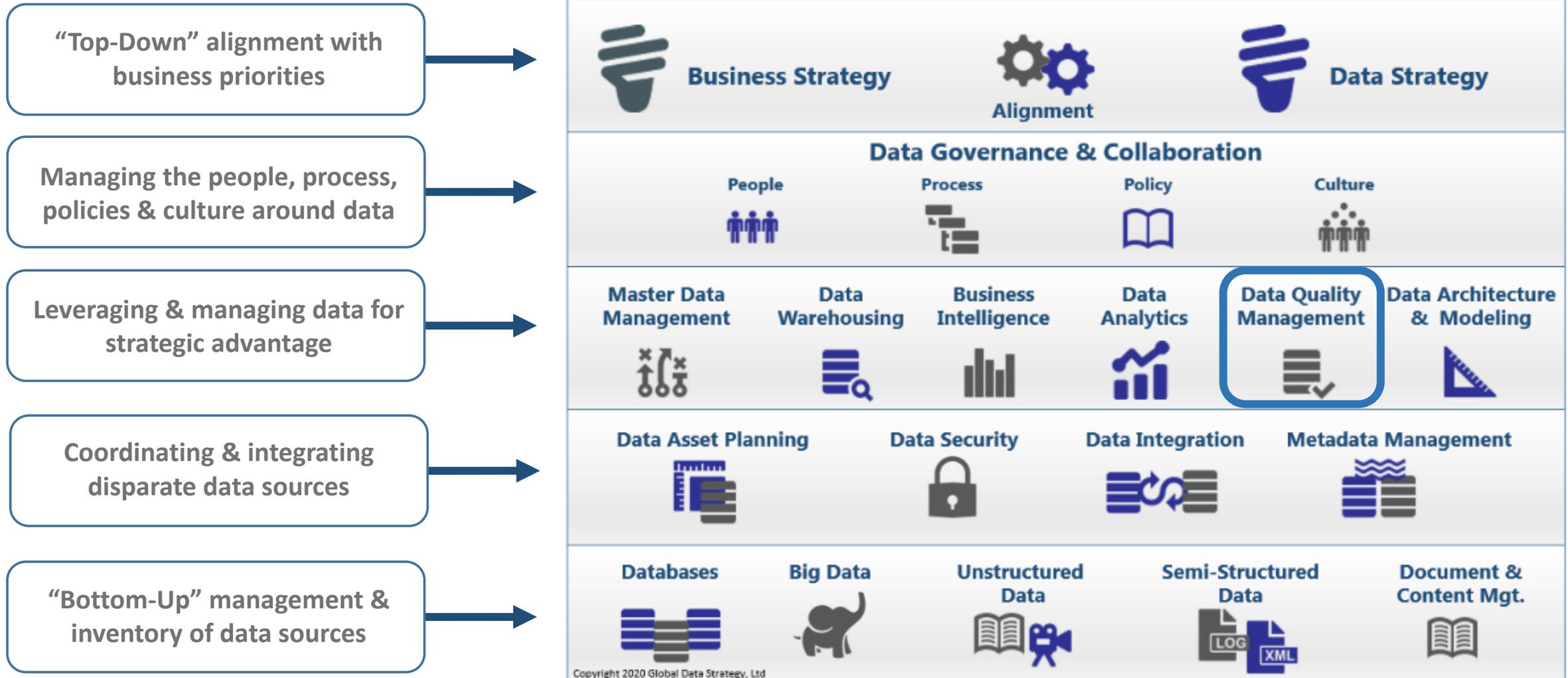
# What We'll Cover Today

- Data quality remains a major challenge for organizations.
- Tackling Data Quality problems requires more than a series of tactical, one-off improvement projects.
- By their nature, many Data Quality problems extend across and often beyond an organization. Addressing these issues **requires a holistic architectural approach combining people, process, and technology.**
- This webinar aims to provide practical ways to control Data Quality issues in your organization.

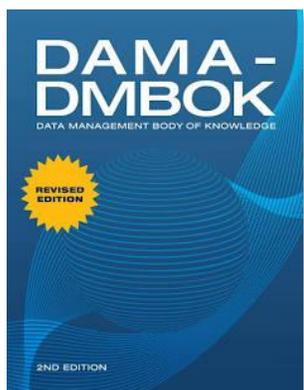
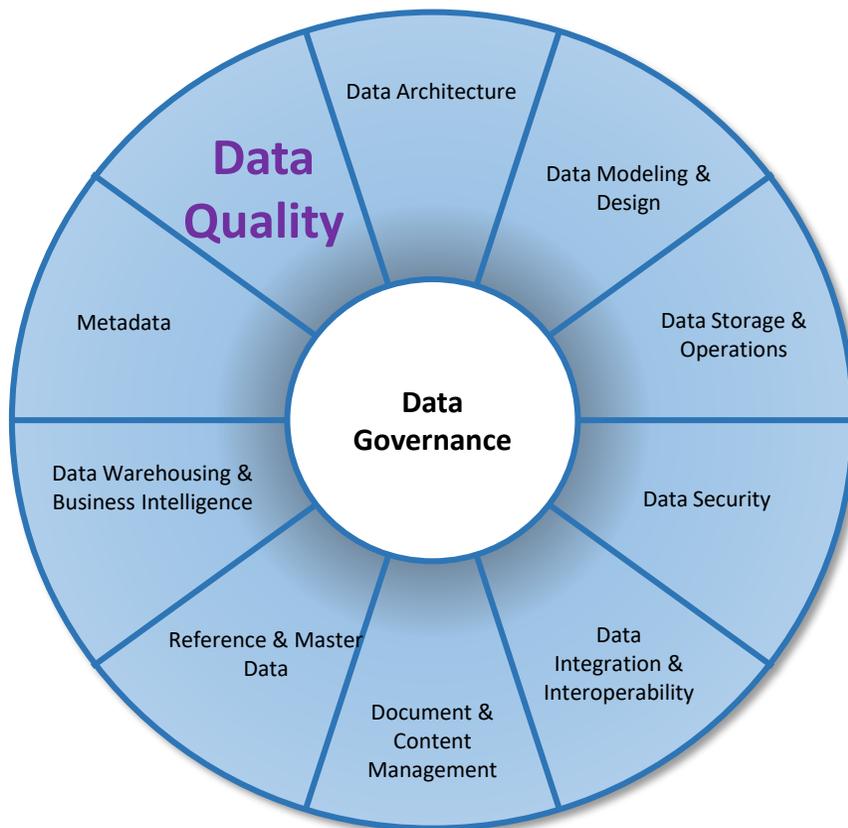


# Data Quality is Part of a Wider Data Strategy

A Successful Data Strategy links Business Goals with Technology Solutions



# Data Quality: a Foundational Data Discipline



DISCIPLINE	EXAMPLE RELATIONSHIPS
<b>Data Governance</b>	DQ requires DG to drive & sustain improvement
<b>Data Architecture</b>	Designs the structural framework for the management of Data Quality
<b>Data Modeling &amp; Design</b>	Identifies business definitions, entities & attributes to focus DQ improvements
<b>Data Storage &amp; Operations</b>	Poor Data Quality impacts efficiency & reliability of Data Storage & Operations
<b>Data Security</b>	Poor Data Quality makes data less secure & more open to fraud
<b>Data Integration &amp; Interoperability</b>	Depends on defined & consistent data formats & content
<b>Documents &amp; Content</b>	Good DQ practices support D&C, e.g. version control, tagging, taxonomies et al
<b>Reference &amp; Master Data</b>	Manages widely shared, business critical data, ensuring single truth, high quality data
<b>Data Warehousing &amp; Business Intelligence</b>	DQ is the foundation of effective DW&BI (e.g. business definitions for KPIs etc.). Also garbage in, garbage out is as true as ever.
<b>Metadata</b>	Provides context & meaning to data and so enhances Data Quality

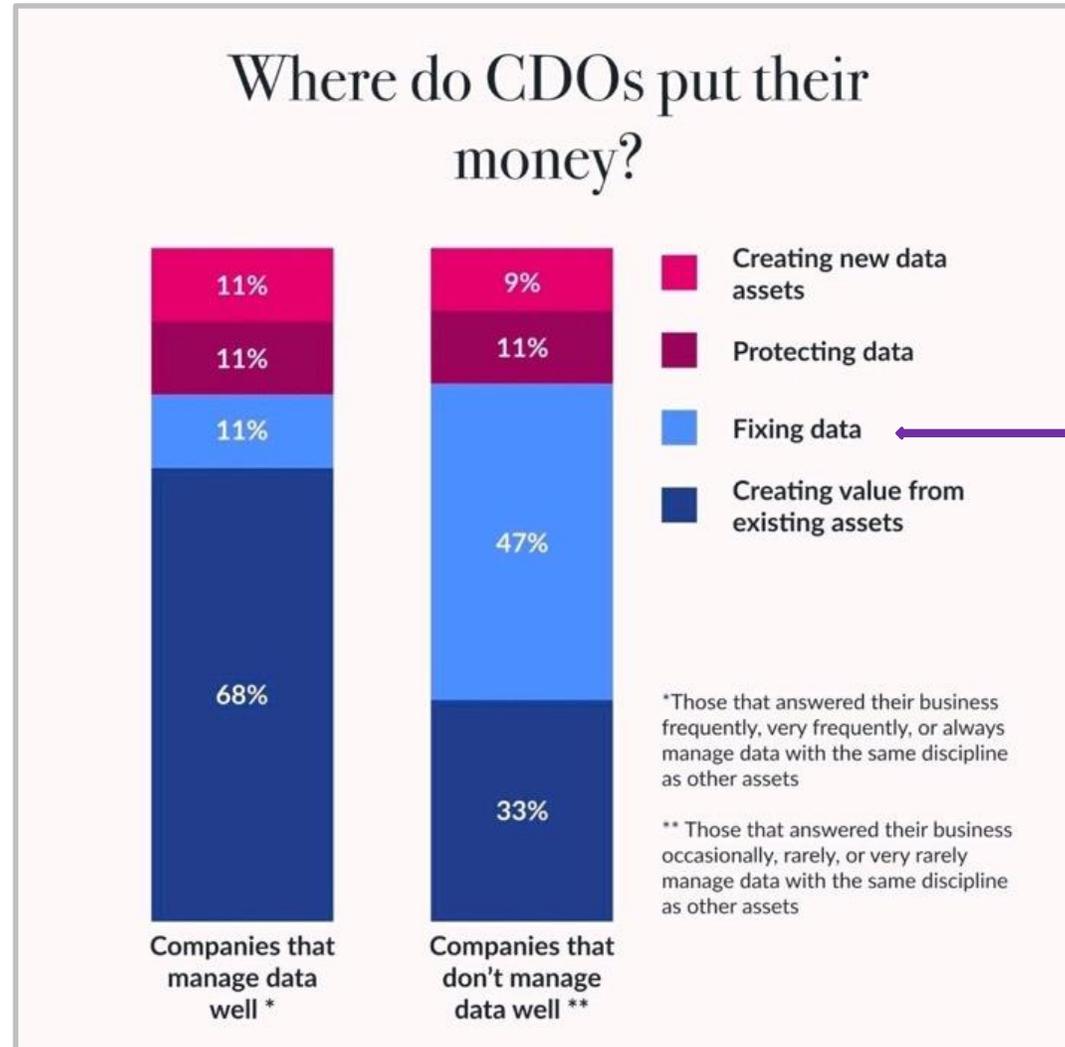
# The Continued Importance of Data Quality

- 2024 survey of 104 CDOs and other key data professionals in the UK
- **Improving data quality was seen as a “critical focus area” by 90.7%** of respondents

“CDOs are increasingly recognizing that data quality is foundational to the success of any data-driven initiative” (Report p8)



# The Value Impact of Managing Data Well

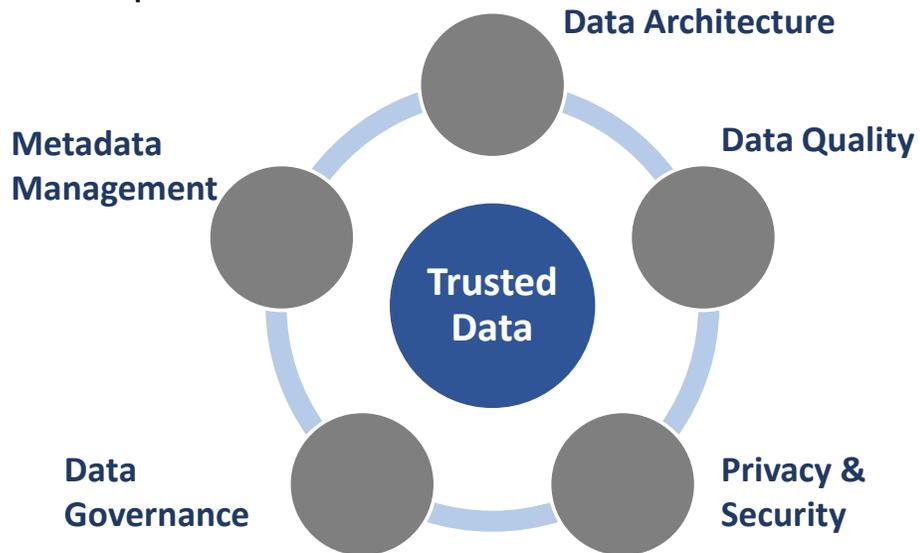


Fixing data is a major cost of failure in virtually all organizations, impacting the bottom line and reducing the value and promised benefits of other data management initiatives

# Building Trusted Data Sets

## What does it take to build Trusted Data Sets?

- Successfully building trusted data sets requires a full range of data management disciplines



# Data Quality – A Simple Definition

Data that is **demonstrably** fit for purpose

**Demonstrably:** Implies that data quality & improvement can be measured, and business impact demonstrated

**Fit for Purpose:** Data quality must meet the needs of the organization and its stakeholders

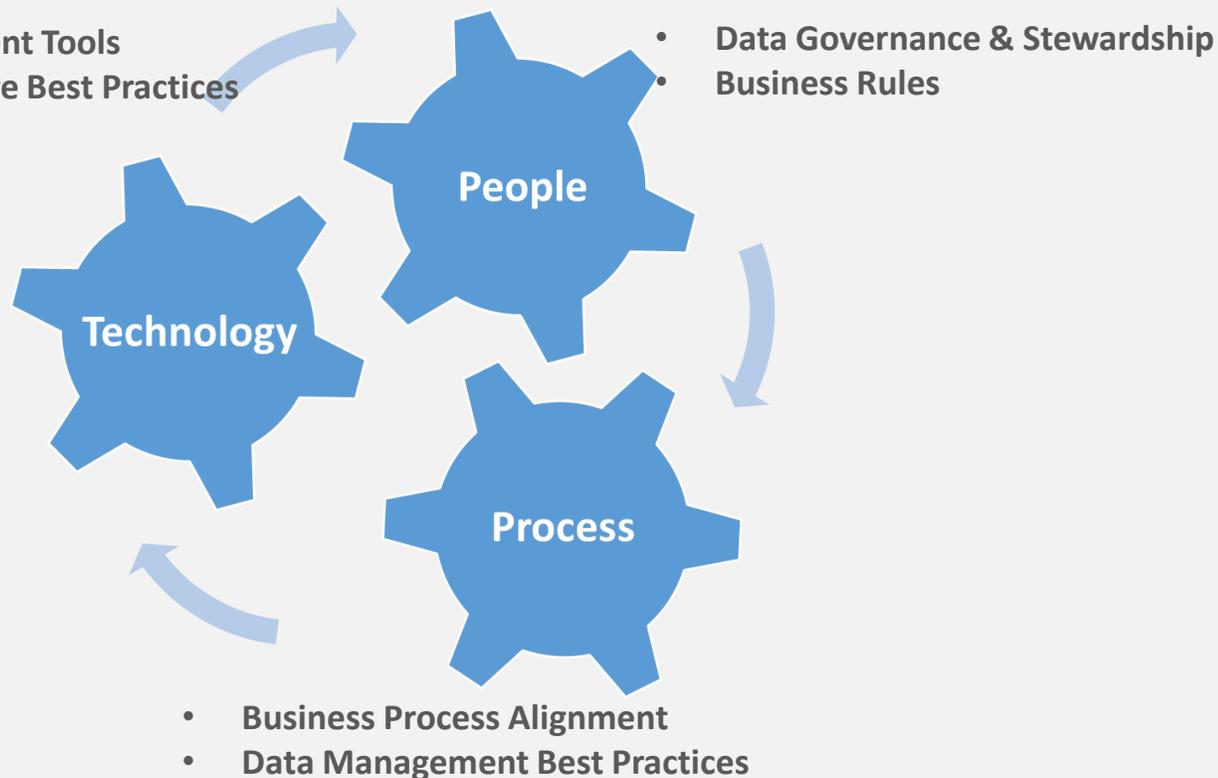
# Fit for Purpose Data

- “Fit for purpose” implies that 100% data quality is not always required or necessary
- What is “fit for purpose” depends on the specific needs of the organization
- It is important to understand the specific business context and use of the data
- When data is used for several different purposes, the definition of fit for purpose will vary from user to user
  - Fit for purpose data ideally needs to meet the needs of all its users
  - Where not possible, a compromise should be sought



# Addressing Data Quality: the Need for a Holistic Approach

Improving Data Quality requires a combination of People, Process, and Technology

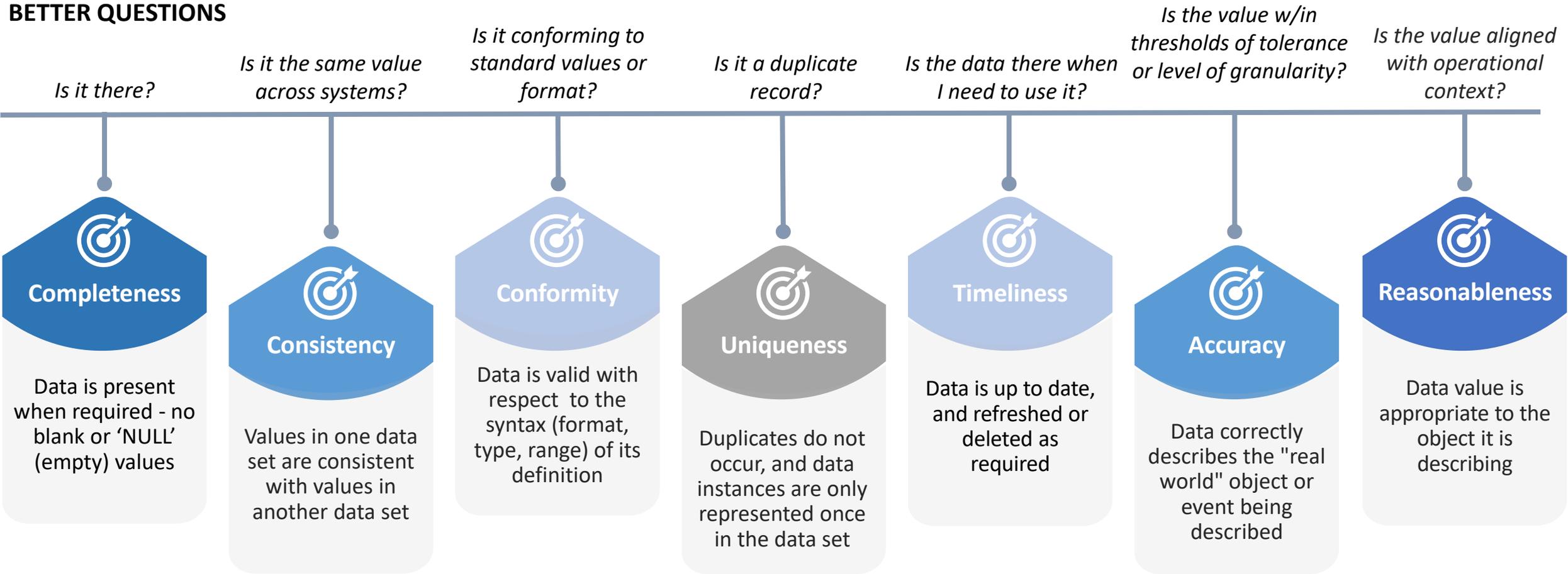


## Data quality is NOT:

- **A synonym for data cleansing.** Data cleansing is a repeated cost of failure and usually does not remove the root causes of poor data quality
- **A one-time effort.** Data is volatile and so data quality has to be a perpetual business as usual activity enabled through data governance

# Data Quality – It's More Than “Is It Right?”

## BETTER QUESTIONS



## DIMENSIONS OF DATA QUALITY

# Applying Data Quality Dimensions

John Smith  
PO Box 9997  
Billings, MT USA

<- We only accept physical addresses, not PO Boxes

## Conformity

That is the address from when John was in school – He doesn't live there anymore.

## Timeliness & Currency

j.smith@aol.com

<- Address means physical address, not email address

## Definition & Purpose

Is this address for the same person?

John Smith 101 Main ST Billings, MT USA	J Smith 101 Main Street Billings, MT US	Jack Smith 1101 Main Billings, MT
---	---	---

## Uniqueness

John Smith  
101 Main ST  
Billings, MT USA

<- The country should come from the Country reference code of 'US'

## Linkage & Integrity

The marketing & billing applications have different addresses for John.

John Smith 101 Main ST Billings, MT USA	John Smith 117 Poncey ST Lincoln, MT USA
---	--

## Consistency

John Smith  
99999 NoName ST  
Billings, MT USA

<- The format is right, but is there really a NoName ST? Can there be a street number of 99999?

## Reasonableness

John Smith  
117 Poncey ST  
Lincoln, MT USA

<- That's not John's address

## Accuracy

John Smith  
101 Main  
Billings

<- That's not a full address.

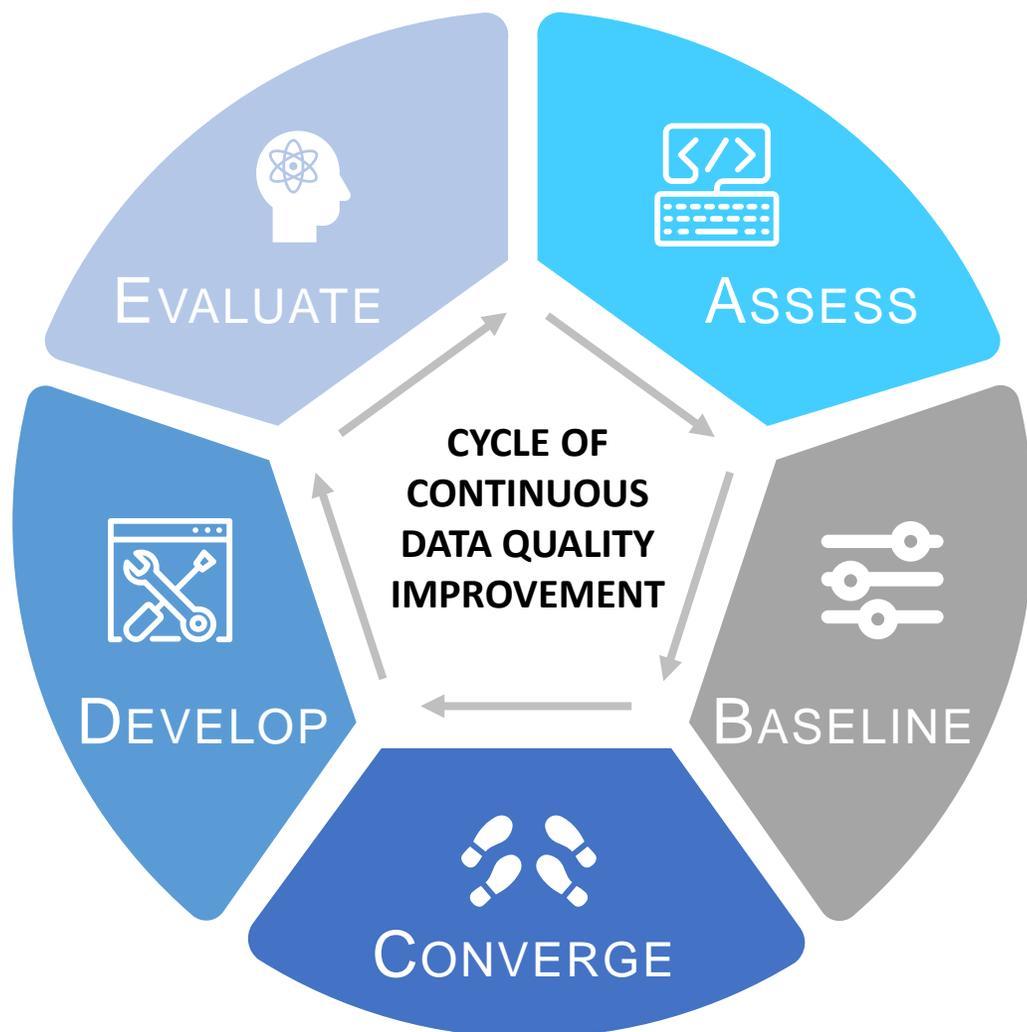
## Completeness

## Accuracy



# Tackling Data Quality Requires a Consistent Method:

## The A2E Approach



### Assess Business Usage

Understand what data exists and how it is used within the organization

### Baseline Data Sources

Baseline the current quality of the data and assess how well it is meeting business needs

### Converge on Business-Critical Areas

Focus priorities to optimize early business benefits and set 'fit for purpose' quality targets to guide improvement activities

### Develop Improvements

Design & deploy improvement initiatives (encompassing people, process, and technology) and measure the impact against targets

### Evaluate Benefits and ROI

Regularly measure the data and continue to improve it so that it continues to meet current and future business needs



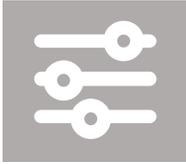
## A2E Step 1: Assess

### ASSESS THE BUSINESS LANDSCAPE

- Understand the business & its primary objectives
  - Data Governance Prioritization Log & Use Cases
- Identify the primary data stakeholders (Business, IT, External parties (e.g. customers, suppliers, partners))
- Evaluate scale and impact of current data quality and establish:
  - Where / how it is captured, stored and processed ?
  - What's working well ?
  - What needs to be improved ?
  - The potential business benefits of better data quality
- Identify potential ROI benefits if Data Quality issues were addressed

### POTENTIAL OUTPUTS & TOOLS

- **Highlight:**
  - Most important business critical data domains
  - Business impact
  - Main data creators and consumers
  - Accountability for the data
  - Current problems and issues with the data
  - Opportunities & potential benefits
- **Outputs may include:**
  - RACI Stakeholder Matrix
  - Rich Picture highlighting real-world issues
  - Data Quality Issues (& Opportunities) Log
  - Business Data Model / Conceptual Data Model
  - Business Process Models
  - High Level Business Case



# A2E Step 2: Baseline

## The Value of Data Profiling

### Quantifying Data Problems

- Data profiling assesses the quality of data sources
- The benefits of data profiling include:
  - Checks conformance of the dataset with business rules
  - Enables fact-based discussion of the causes and impacts of data problems
  - Excellent starting point in data quality workshops

Results Browser

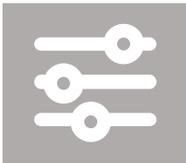
Job: US Customer Data Profiling

Input Field	Total Number	Minimum Length	Maximum Length	Minimum Value	Maximum Value
ID	5438	9	9	AAC434152	ZZZ642455
Name	5438	11	39	Anne Mullen	de Chana, Sergio Marques
Street	5438	2	41	# 3 Riverdrive Rd. East	Wilson & Kirk Road
City	5438	3	20	ABERDEEN	waterloo
State	5438	2	2	AB	WY
ZIP	5438	4	10	01801-6202	n2j4a9
Country	5438	1	13		United States
Phone	5438	1	25	(113) 072 3578	x
Cell	5438	4	14	(113) 575 3765	9978 158
Work	5438	4	28	(113) 007 6029	x7562
eMail	5438	16	35	Aaron.A.Koontz@thu.com	zoi.gibso@snomail.com
DoB	5438	19	19	Jan 1, 1900 12:00:00 AM	Mar 29, 2007 12:00:00 AM
Gender	5438	1	1	F	U
Active	5438	1	1	0	Y
CreditLimit	5438	1	5	0	32800
StartDate	5438	19	19	Apr 1, 2006 12:00:00 AM	Apr 1, 2009 12:00:00 AM
EndDate	5438	19	19	Apr 1, 2008 12:00:00 AM	Apr 1, 2014 12:00:00 AM

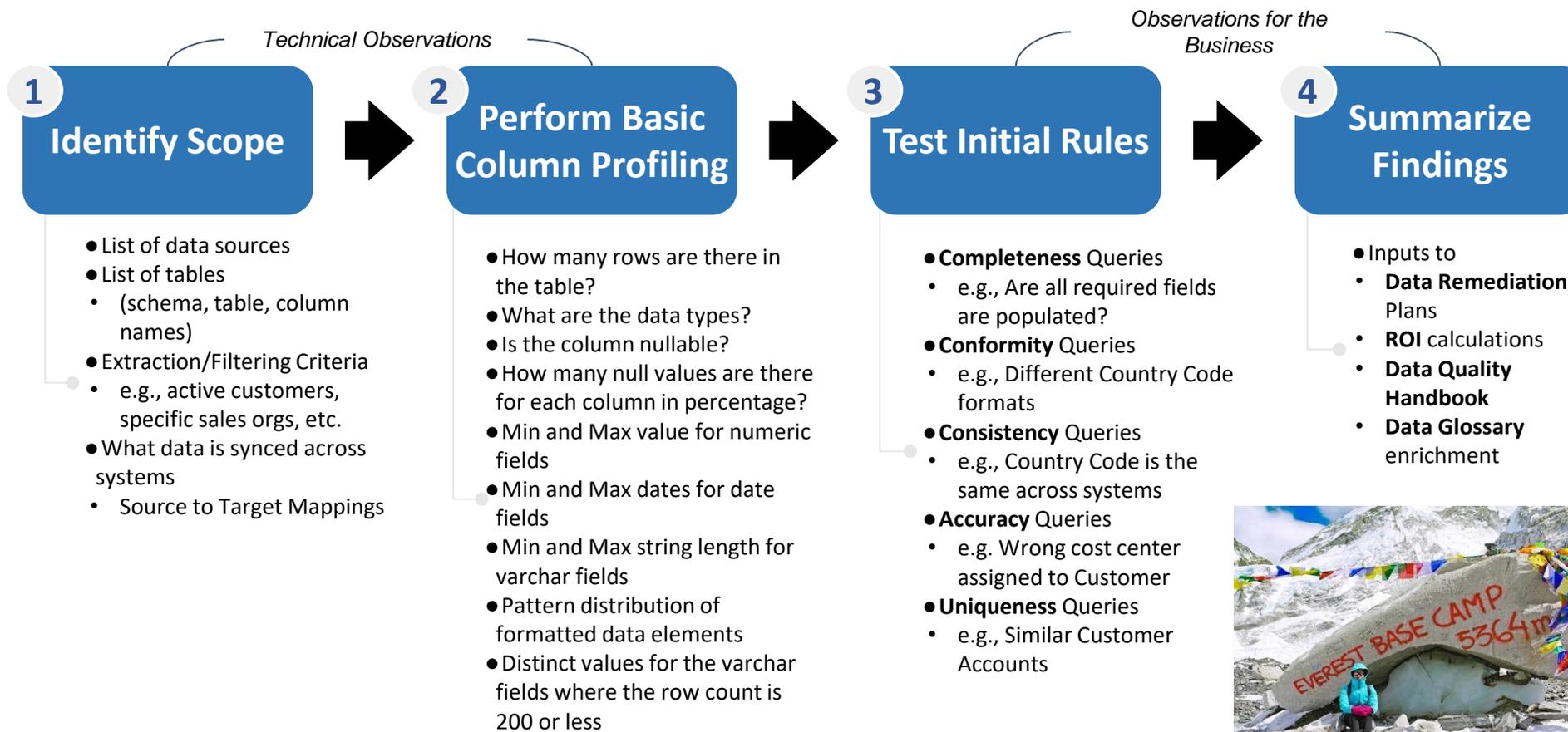
Min and Max Profile | Data

Example partial Data Profiling report

www.globaldatastrategy.com



# A2E Step 2: Data Profiling How-To

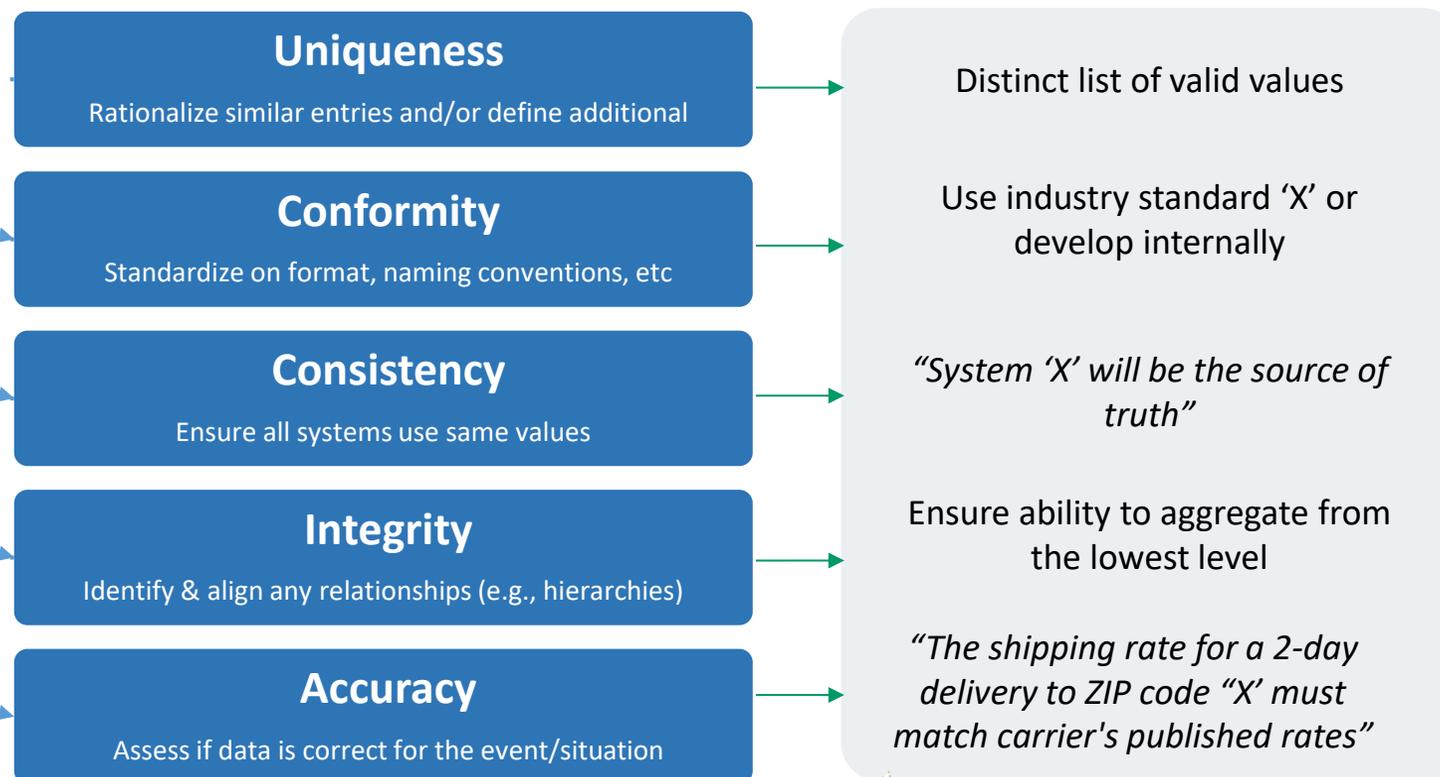




## A2E Step 3: Converge on Business-Critical Areas

- Data profiling findings should be reviewed by appropriate business & IT stakeholders
  - If formal Data Governance in place, this should ideally be led by the Data Stewards responsible for the specific data areas
- Conduct working sessions with the business to assess data profile outputs to create rules so that data quality can be tested and measured
- Focus on the data that provides high value to the business (e.g., Data quality rules for Fax number format is likely less relevant than the accuracy of a location).

### Examples



Ways to consider what types of rules would apply to a data element

Outputs can affect rules, processes, & data flows (integrations)



# Step 3: Identifying Critical Data Elements (CDEs)

## What is a Critical Data Element (CDE)?

A Critical Data Element (CDE) refers to data that is deemed essential for an organization's operations, decision-making processes, regulatory compliance, and overall strategic objectives.

## Why identify CDEs?

- Not all data is equally important so it's important to focus and prioritize on key data that matters most
- May be master or reference data that is widely used and shared across the organization
- Enables Data Steward to focus on a limited subset of data to maximize the business benefits of improvement
- Important first step in data quality monitoring and dashboarding – CDEs often have higher data quality thresholds than non-critical data

## How do you identify a CDE?

- Criticality to an operational or business process
- Importance in delivering a key business change driver
- Sensitivity due to PII or other privacy / security concerns
- Legal or regulatory requirement

EMPLOYEE ATTRIBUTES	CRITICALITY
Employee ID	CDE
First Name	
Middle Name	
Second Name	CDE
Work Address	
Home Address	
Social Insurance Number	CDE
Role Grade	
Salary Band	CDE
Current Monthly Salary	
Work email address	CDE
Personal email address	
Emergency contact name & number	CDE

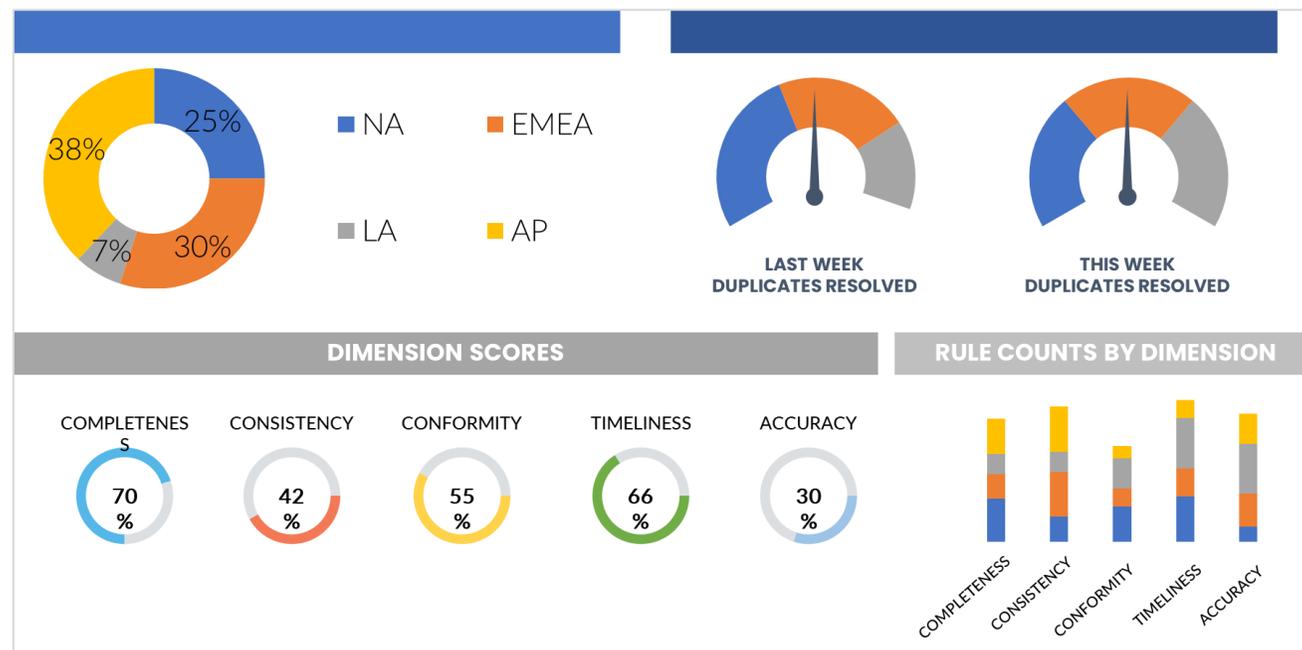


## A2E Step 4: Develop & Implement Improvements

### Design and implement improvements

- Identify related operational KPIs that would be impacted with improvements in data quality
- Establish KPI Thresholds and Targets
- Design Data Quality Dashboard(s)
- Perform Data Quality Dashboard Build
- Create processes that enable the business to monitor and remediate data quality issues more proactively

### Example Foundational Data Quality Dashboard





"A Business Rule is a criterion used to guide day-to-day business activity, shape operational business judgments, or make operational business decisions."

Ronald Ross, quoted in  
[architectureandgovernance.com](http://architectureandgovernance.com)

- In a data context, business rules are used to define and enforce the standards that data must conform to
  - Have a key role in assessing, baselining and improving data quality
  - Are used to specify data design, e.g. drop down lists, data input validation etc.
- Business rules can be discovered or derived from:
  - Data models
  - Business and IT documentation
  - Documented metadata
  - Data profiling activities
  - Talking to subject matter experts
- A simple typology of Business Rules as applied to data is:
  - Format business rules – specify the format standards data should comply with
  - Content business rules – specify the allowable content of records or fields

# Deploying Business Rules: Approaches



Data Entry  
Guidelines,  
Business Glossary /  
Catalog & Training



Master & Reference  
Data Management



Application Code  
(e.g. data input  
validation)



Data Quality Tool:  
DQ Business Rules  
Engine



## A2E Step 5: Evaluate

### EVALUATE & SUSTAIN GAINS

- Embed Data Quality improvement as a business as usual activity
- Evolve Data Quality improvement teams into wider Data Governance structure
- Track Data Quality improvements via Data Quality Dashboards
- Monitor financial and business benefits over time
- Evangelizing benefits – part of your job is marketing!

### POTENTIAL OUTPUTS & TOOLS

- Evolving & incremental Data Improvement Plans
- Formal Data Governance Roles in place for targeted data areas
- Regular Data Quality Dashboard updates and analysis
- Business Process Change
- Continued ROI and financial benefits
- Communication Plan and Organizational Change Efforts



# Monitor & Report Business Rule Adherence

- When Business Rules are implemented they can be used to:
  - Check continued adherence of existing data
  - Enforce the rules on new data to prevent new problems
- Monitor via Data Quality Dashboards
  - Regular reports on adherence of data to Business Rules
  - KPIs drive continuous data improvement
  - Data quality trends
  - Highlight areas where corrective action required
  - Indicate where Business Rules may need to be amended to meet changing business needs
- Relate Data Quality to Business Outcomes
  - Address the “so what” objection
  - Puts a financial or other benefit on continued data quality improvement



Data Quality Dashboards

# The Future?: The Potential Synergy of Data Quality and AI/ML

## Data Capture & Validation

- AI can help to automate data capture and so avoid human error
- Where human data entry is necessary AI can validate data input based on defined business rules and prevent the input of bad data

## Automated Data Cleansing

- AI can automate the process of data cleansing, including identifying and removing duplicates, filling in missing values and correcting inconsistent or inaccurate data
- Machine learning algorithms can be trained to recognise errors that are often hard to spot with manual review

## Predictive Analysis

- AI can be used to predict missing or unknown values in a dataset based on patterns in the data
- This can improve the completeness and accuracy of the data



## Anomaly Detection

- Machine Learning Models (ML) can be used to identify data that could indicate anomalies
- For instance, a ML model can be trained to identify transactions that fall outside of the normal range for a business, which could indicate data quality errors or fraudulent activity

## Data Standardisation

- AI can automate the process of transforming data into a standard format
- This can help to ensure consistency and improve the accuracy of data

## Proactive DQ Management

- Proactively monitoring data quality metrics and applying predictive analytics can detect potential data quality issues before the business / its customers are impacted
- ML algorithms can analyse large scale, historical data to detect patterns, issues and provide recommendations to enable Data Stewards to implement preventative measures

# In Conclusion: How do you fix Data Quality & Keep It Fixed?

Key message: it's not about IF you do it but about HOW you do it

## OPTION 1: FIGHT FIRES (REACTIVE)

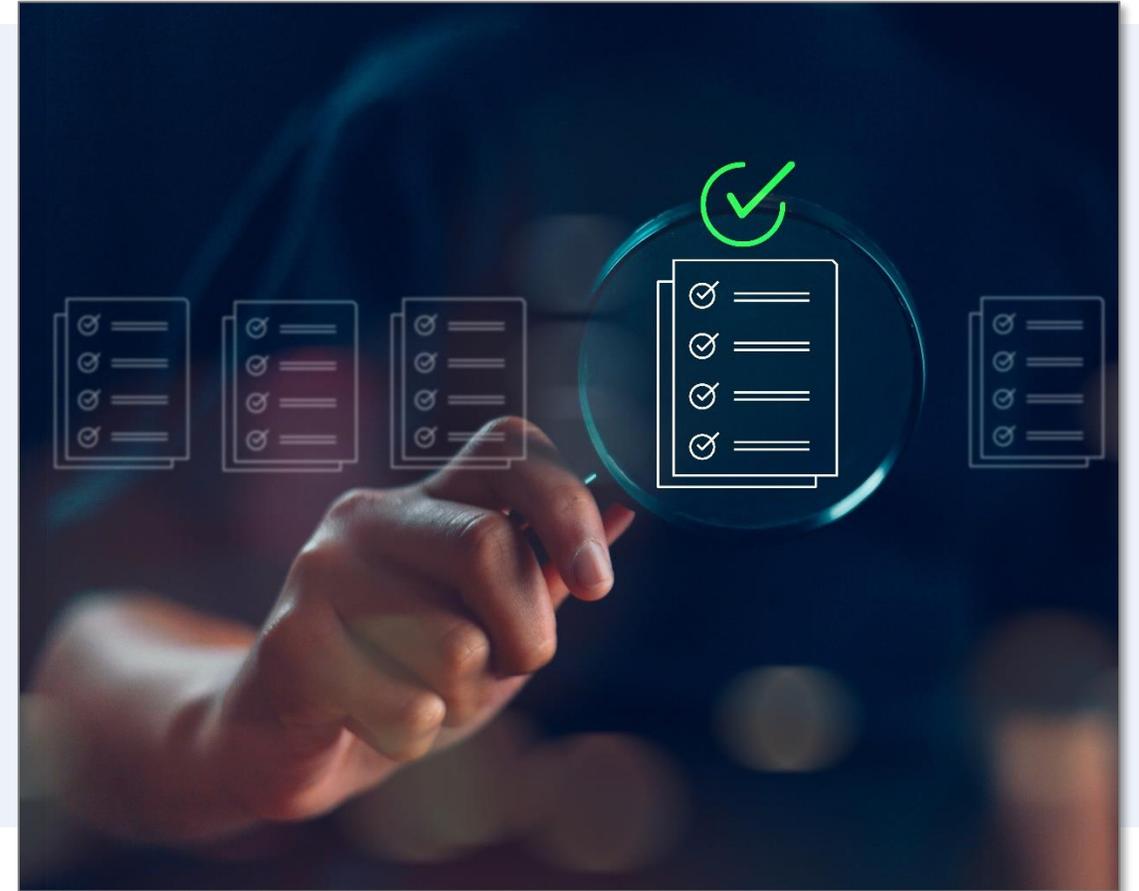


## OPTION 2: PREVENT FIRES (PROACTIVE)



# Summary

- **Data quality management remains a key challenge** for most organizations
- **Build a solid data quality foundation** as a prerequisite of success for other data management capabilities
- **To fix data quality requires a holistic approach** involving People, Process, Technology & Data
- **Use a repeatable and common methodology such as A2E** to increase the chances of delivering and sustaining a high data quality culture



# DATAVERSITY Data Architecture Strategies

## This Year's Lineup

- **January** Emerging Trends in Data Architecture – What's the Next Big Thing?
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Master Data Management - Aligning Data, Process, and Governance
- **April** How do Data Governance & Data Architecture Support Each Other?
- **May** The Role of the Chief Data Officer (CDO) in Business Transformation
- **June** What Does It Mean to be a Data-Driven Organization?
- **July** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Enterprise Architecture vs. Data Architecture
- **December** The Business Benefits of Data Modeling



# Who We Are: Business-Focused Data Strategy

Maximize the Organizational Value of Your Data Investment



In today's business environment, showing **rapid time to value** for any technical investment is critical.

But technology and data can be complex. At Global Data Strategy, **we help demystify technical complexity** to help you:

- Demonstrate the ROI and **business value of data**
- Build a data strategy **at your pace to match your unique culture** and organizational style.
- Create an **actionable roadmap for “quick wins”**, which building towards a long-term scalable architecture.

Global Data Strategy shares experience from some of the largest international organizations scaled to the pace of your unique team.

Global Data Strategy has worked with organizations globally in the following industries:

Finance · Retail · Social Services · Health Care · Education · Manufacturing  
· Government · Public Utilities · Construction · Media & Entertainment ·  
Insurance .... and more



Thoughts? Ideas?  
**Questions?**