



Data Quality Best Practices

Donna Burbank and Nigel Turner
Global Data Strategy, Ltd.

August 24th, 2023





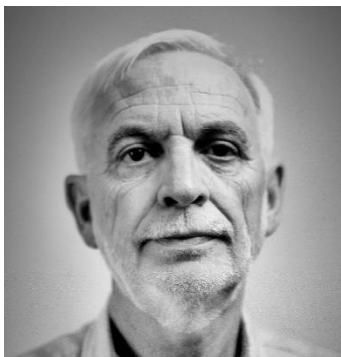
Donna Burbank

- Recognized industry expert in information management with over 25 years of experience in data strategy, information management, data modeling, metadata management, and enterprise architecture
- Managing Director at Global Data Strategy, Ltd., an international information management consulting company that specializes in the alignment of business drivers with data-centric technology
- Worked with dozens of Fortune 500 companies worldwide in the Americas,

Europe, Asia, and Africa and speaks regularly at industry conferences

- Excellence in Data Management Award from DAMA International
- Past President and Advisor to the DAMA Rocky Mountain chapter
- Co-author of several books on data management
- She can be reached at 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, Data Governance, 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

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** Data Mesh or Data Mess? Separating the Reality from the Hype
- **April** Master Data Management - Aligning Data, Process, and Governance
- **May** How do Data Governance & Data Architecture Support Each Other?
- **June** Why You Need Data Management – Getting Executive Buy-In
- **July** Artificial Intelligence and Machine Learning – Building the Right Architectural Foundation
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Designing Data for Business Intelligence & Analytics – Where the Star Schema Fits in a Modern Data Architecture
- **December** Enterprise Architecture vs. Data Architecture



What We'll Cover Today

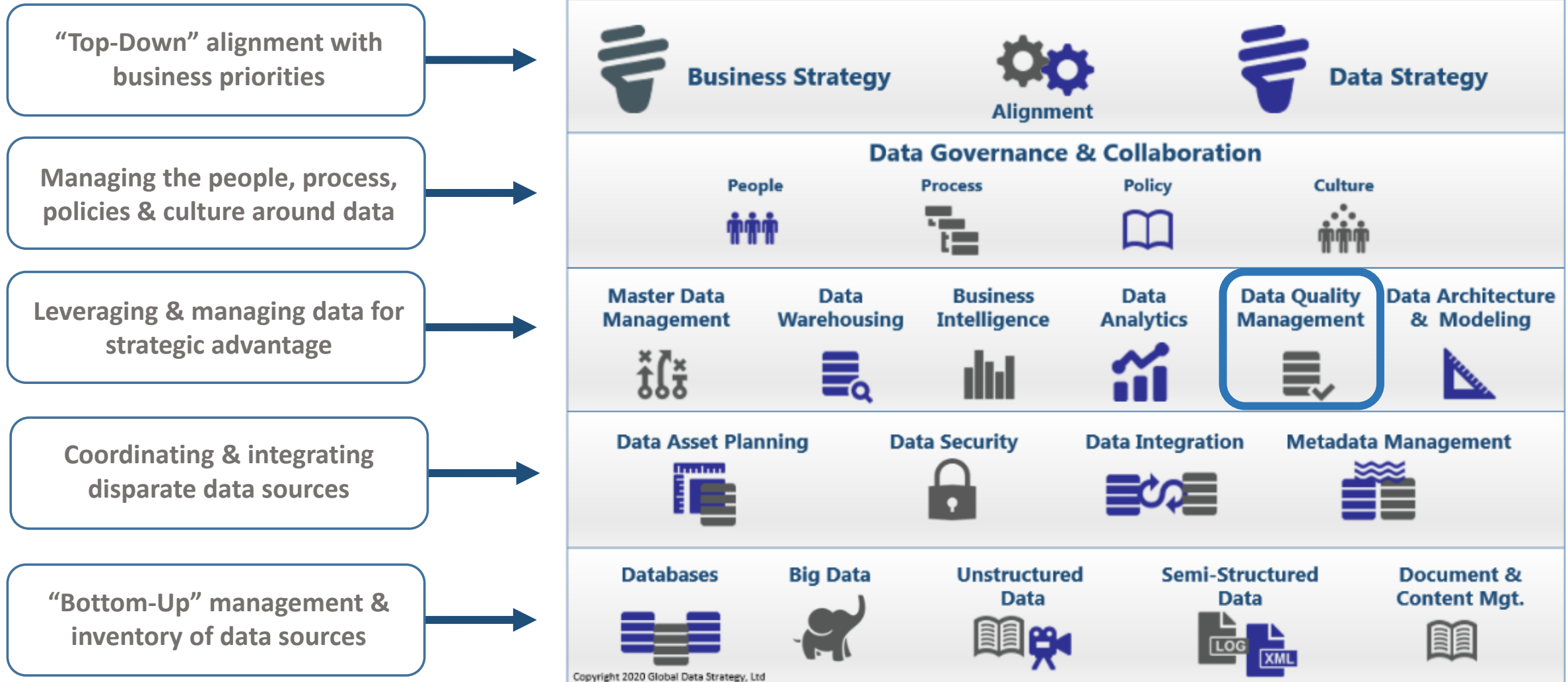
- Tackling Data Quality problems requires more than a series of tactical, one-off improvement projects.
- By their nature, many Data Quality problems extend across and often beyond an organization. Addressing these issues requires a holistic architectural approach combining people, process, and technology.
- Join Nigel Turner and Donna Burbank as they provide practical ways to control Data Quality issues in your organization.



Data Quality

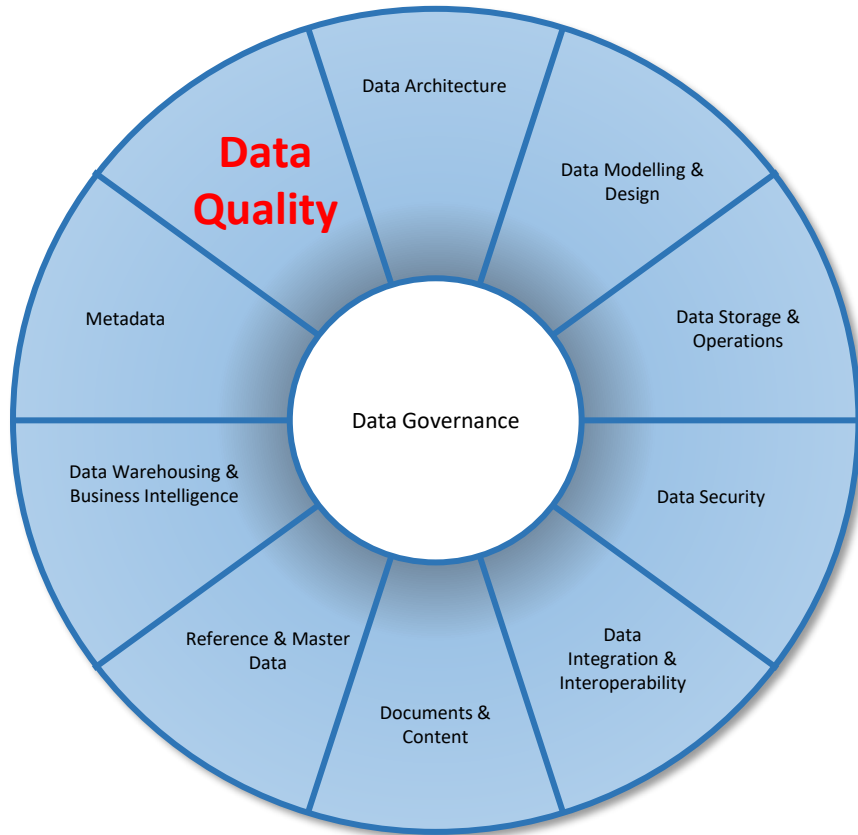
Data Quality is Part of a Wider Data Strategy

A Successful Data Strategy links Business Goals with Technology Solutions



Data Quality – Relationships with Other Data Disciplines

11 DATA DISCIPLINES



DAMA DMBOK – Version 2 2017

DISCIPLINE	EXAMPLE RELATIONSHIPS
Data Governance	DQ requires DG to drive & sustain improvement
Data Architecture	DA designs the structural framework for the management of DQ
Data Modelling & Design	DA&M identifies business definitions, entities & attributes to focus DQ improvements
Data Storage & Operations	Poor DQ impacts DS&O efficiency & reliability
Data Security	Poor DQ makes data less secure & more open to fraud
Data Integration & Interoperability	DI&O depends on defined & consistent data formats & content
Documents & Content	Good DQ practices support D&C, e.g. version control, tagging, taxonomies et al
Reference & Master Data	R&MD manages widely shared, business critical data, ensuring single truth, high quality data
Data Warehousing & Business Intelligence	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.
Metadata	MD provides context & meaning to data and so enhances DQ

Data Quality Problems Continue to Hit the Headlines

Typo leaks millions of US military emails to Mali

FINANCIAL TIMES HEADLINE
17 July 2023

UK Ministry of Defence accidentally sends classified emails meant for US to Russian ally

UK GUARDIAN HEADLINE
16 August 2023

PROBLEM

- The US and UK military have sent millions of emails meant for the US military to Mali. Details included documents, passwords and travel plans of military personnel

CAUSE

- The emails should have been sent to the domain name of .MIL (US Military).
- Instead typos meant many were sent to .ML – the government of Mali

IMPACT

- These emails released key military information to a country which is closely allied to Russia
- Both the US and UK have been warned about these issues for 10 years, but have never addressed the problem

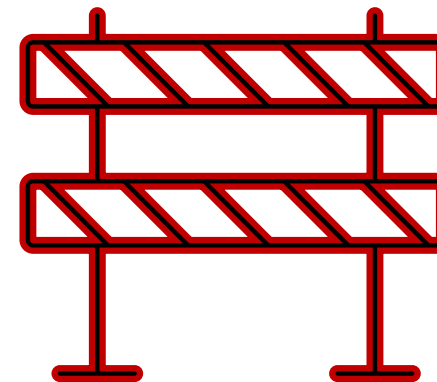


Key Lesson:
Simple data quality errors can cause major problems

Poor Data Quality is an Impediment to Analytics

Correcting poor data quality is a Data Scientist's least favorite task, consuming on average 80% of their working day

Source: Forbes 2016



Poor Data Quality is a Barrier to getting the most out of an organization's data

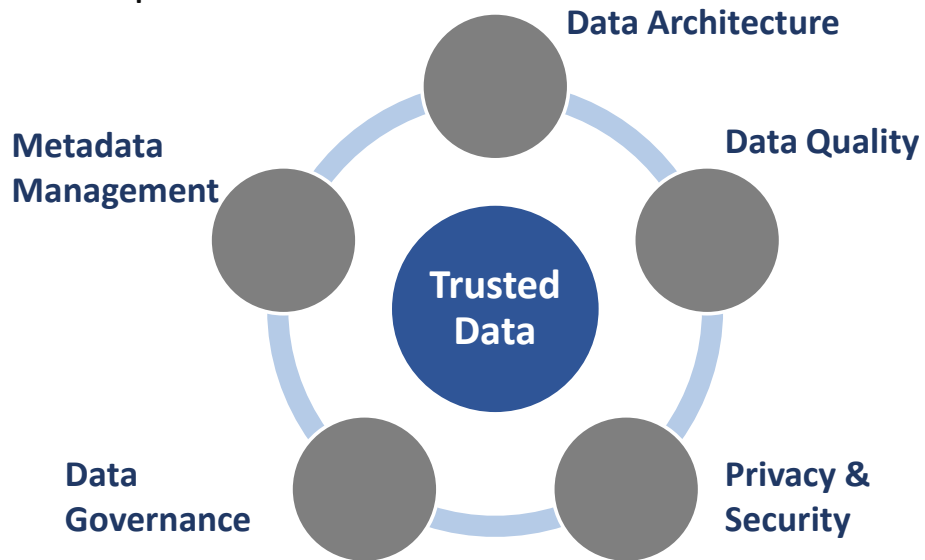
Raw data used in Self-Service Analytics and BI environments is often so poor that **many data scientists and BI professionals spend an estimated 50 – 90% of their time cleaning and reformatting data to make it fit for purpose.**⁽⁴⁾

Source: DataCenterJournal.com

Building Trusted Data Sets

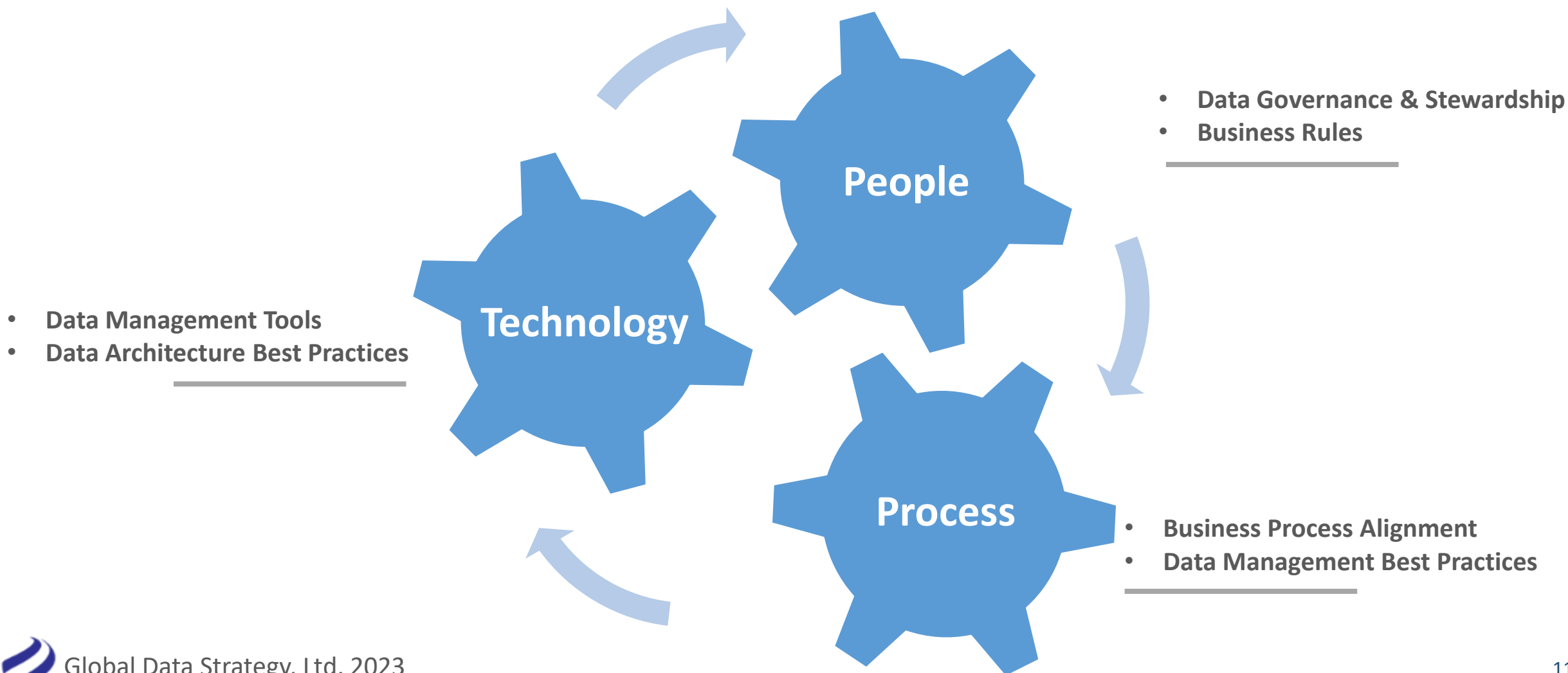
What does it take to build Trusted Data Sets?

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

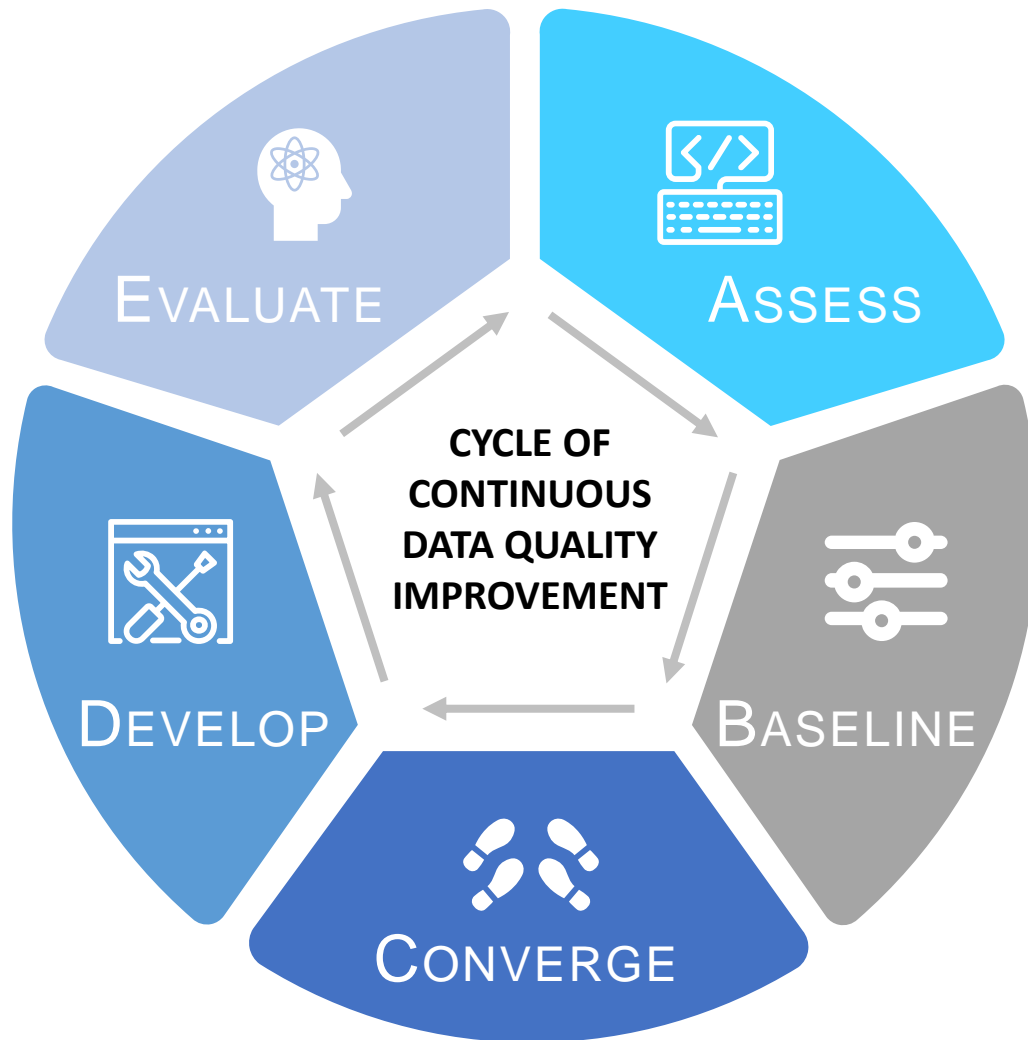


Addressing Data Quality – a Holistic Approach

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



Tackling Data Quality: 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

ASSESS THE BUSINESS LANDSCAPE

- Understand the business and its primary goals & objectives
- Analyze what data the business:
 - Relies on today
 - Will need to support its future aspirations
- Identify the primary data stakeholders:
 - Business
 - IT
 - External parties (e.g. customers, suppliers, partners)
- Work with them to 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
- Create a Data Quality Issues (& Opportunities) Log

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

Data Issues & Opportunities Log: Suggested Template

ID	Short Name	Brief Description	Impact of the Problem / Potential Opportunity (Business & IT)	Raised By
1	Customer Data Duplication	Both in the CRM platform and the Customer Data Warehouse there are known customer record duplications, mainly caused by marketing and salespeople not being able or willing to search for an existing customer record. One estimate is that up to 25% of CRM customer records are duplicates. Data Warehouse duplication unknown.	<ul style="list-style-type: none"> Multiple marketing communications sent to the same customer, causing brand damage Inability to evaluate total customer lifetime revenue value Impossible to derive a single view of a customer Risks contravening GDPR if customer submits Data Subject Access Request (DSAR) and all data not returned Resolving this problem would accelerate MDM ambitions and enable better targeted 1-1 customer marketing 	<ul style="list-style-type: none"> Bob Mills (Marketing Manager) Anna Ford (CRM Technical Architect) George May (Senior Sales Rep)
2	Product Data Inconsistencies	Product data is held and processed in several different platforms and systems. Each system has its own set of product reference codes. Many inconsistencies have been caused by the company making several acquisitions over recent years. No attempt has yet been made to standardize codes despite wide awareness that this causes problems for the business.	<ul style="list-style-type: none"> Product codes need to be manually amended and rekeyed when input into other systems. At present it is estimated that around 3 person years are spent each year on this task. The above process is subject to error and the incorrect codes are often input to systems. This can lead to the wrong product being dispatched, leading to customer complaints and rework. 	<ul style="list-style-type: none"> Rachel Smith (Product Manager) Akhtar Abdul (IT Support) Betty Willis (Dispatch Coordinator) Arya Patel (Finance)
3	Poor Data Training of Data Entry People	Data entry people are distributed throughout the company in many different siloed parts of the organization. This has led to many data entry people being unaware of where data they enter is used across the business, other than within their own immediate functional areas.	<ul style="list-style-type: none"> Often data is not fit for purpose when processed by downstream business units, especially dispatch and finance functions. This requires these units to contact data entry teams to try to resolve data errors, and this impacts the productivity of both data entry teams and downstream functions. In general, awareness of the importance of accurate and complete data entry is low. 	<ul style="list-style-type: none"> Sara Braun (HR Director) Fred Sarat (Sales Manager) Rachel Smith (Product Manager) Arya Patel (Finance)

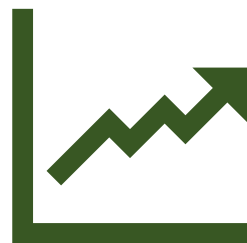
Making the Business Case

While Business Cases and ROI Calculations can be complex, they generally fall into 4 categories:



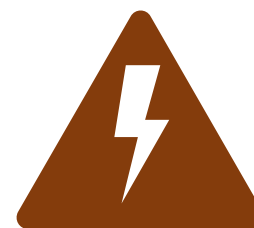
Decreasing Costs

- **Wasted labor costs due to manual efforts**
(Data cleansing, finding data, manual integration, etc.)
- **Inefficient business processes for data management**
(Product Master Data process)
- **Data quality cost avoidance**
(Wasted mailings sent to wrong addresses)



Increasing Revenue

- **Price optimization through Analytics**
- **Improved marketing campaigns through quality customer data**
- **Data-driven recommendation engines to enhance the sales cycle**
- **Better grant applications through data-driven analysis**



Reducing Risk

- **Industry regulations**
(GDPR, HIPAA, BCBS 239, Spice, etc.)
- **Product traceability**
(Food lineage from farm/catch)
- **Litigation due to data breaches**
- **Health and safety audits**



Protecting Reputation

- **Customer satisfaction**
- **Brand trust**
- **Social media voice of consumer**
- **Loyalty & 'stickiness'**

Include the Risk of Doing Nothing

- There is significant cost and risk in the status quo
- Doing nothing often has a higher cost than investing in data management
- Make sure to include the “do nothing” option in your analysis



BASELINE CRITICAL DATA SOURCES

- Establish a quantitative view of key data quality problems
 - Measure the baseline quality of key data sources to quantify the issues
- To do this:
 - Select the key data sources and data domains identified in the Step 1 Assess
 - Profile the data and focus on key objects and attributes
 - Assess the data according to the 7 Dimensions of Data Quality
 - Present the results to relevant stakeholders - gain consensus on the business impact of the problems found
 - Expand and refine the Data Quality issues log

POTENTIAL OUTPUTS & TOOLS

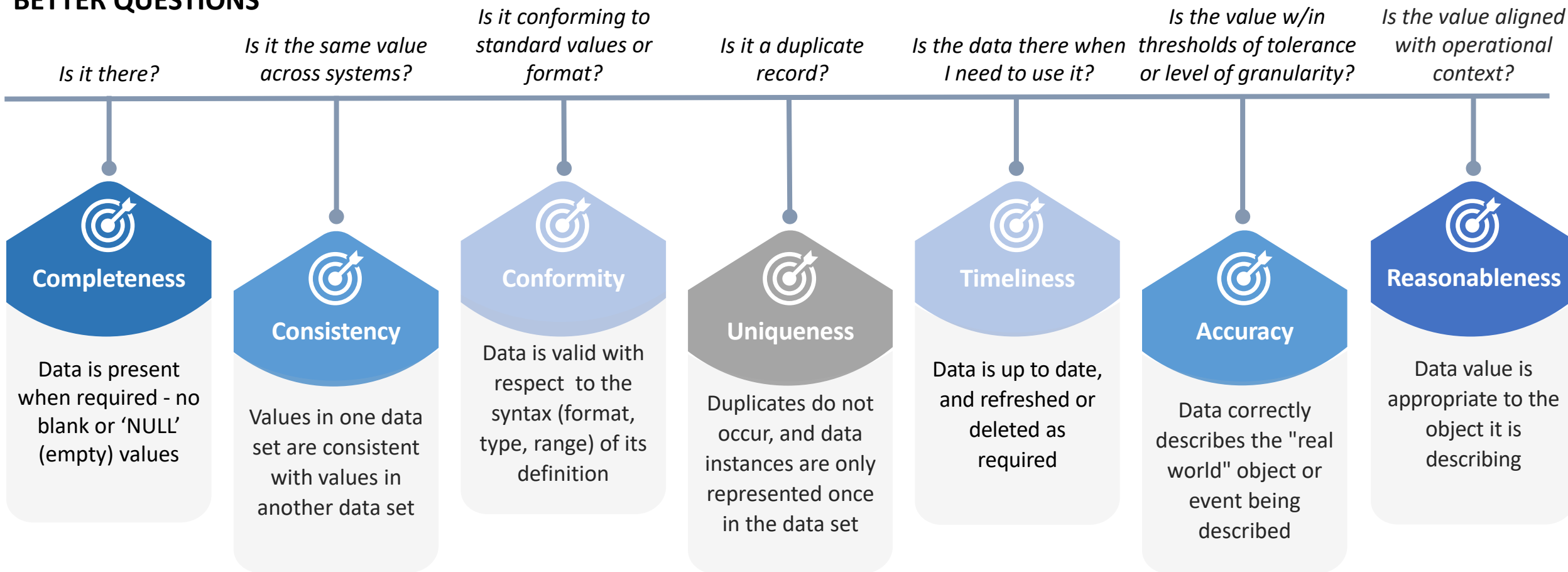
- Data Quality Report(s)
- Data Profiling outputs – derived metadata
- Updated Issues & Opportunities Log, with quantification of financial costs and other business impacts



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

Using specific language when assessing data quality reduces ambiguity and makes it measurable and actionable

BETTER QUESTIONS



DIMENSIONS OF DATA QUALITY

TARGET EXPECTATIONS

Applying Data Quality Dimensions

How would the Data Quality Dimensions Apply to Address?

John Smith
PO Box 9997
Billings, MT USA

<- We only accept physical addresses, not PO Boxes

Conformity

j.smith@aol.com

<- Address means physical address, not email address

Definition & Purpose

John Smith
101 Main ST
Billings, MT USA

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

Linkage & Integrity

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

That is the address from when I was in school – I don't live there anymore.

Timeliness & Currency

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

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
101 Main
Billings

<- That's not a full address.

Completeness

Address

The Importance of Business Review & Validation

- 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
- Aim to reach consensus on what the business impact is
- Ways of doing this:
 - Workshops and / or meetings (virtual or F2F)
 - By workflows, seeking views on the potential problem areas
- For priority areas, agree Business Rules which should be in place to baseline current data quality and measure data quality improvement



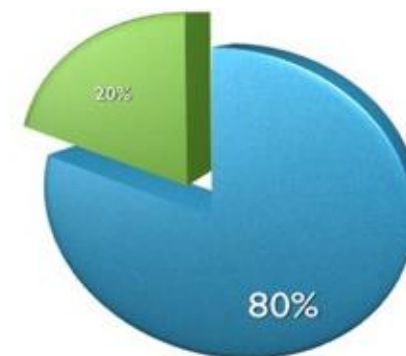
A2E Step 3: Converge

PRIORITIZE & FOCUS ON SPECIFIC ISSUES & OPPORTUNITIES

- Determine initial data quality improvement projects; focus in on two things:
 - Potential pilot / proof of concept data quality improvement project(s)
 - Data quality improvement projects with the largest net benefits
- Note: these are often NOT the same thing; in the early stages of a DQ initiative it's important to establish credibility and prove the potential benefits of wider adoption via a PoC
- Work with stakeholders to identify priorities from the Data Quality Issues log
- Prioritize projects (e.g. Priority Grid – see later)
- Run pilots / proofs of concept
- Identify and run initial DQ improvement projects

POTENTIAL OUTPUTS & TOOLS

- Prioritized Data Quality Issues Log
- Priority Grid
- Agreed pilot project(s)
- Agreed potential DQ projects
- Business cases



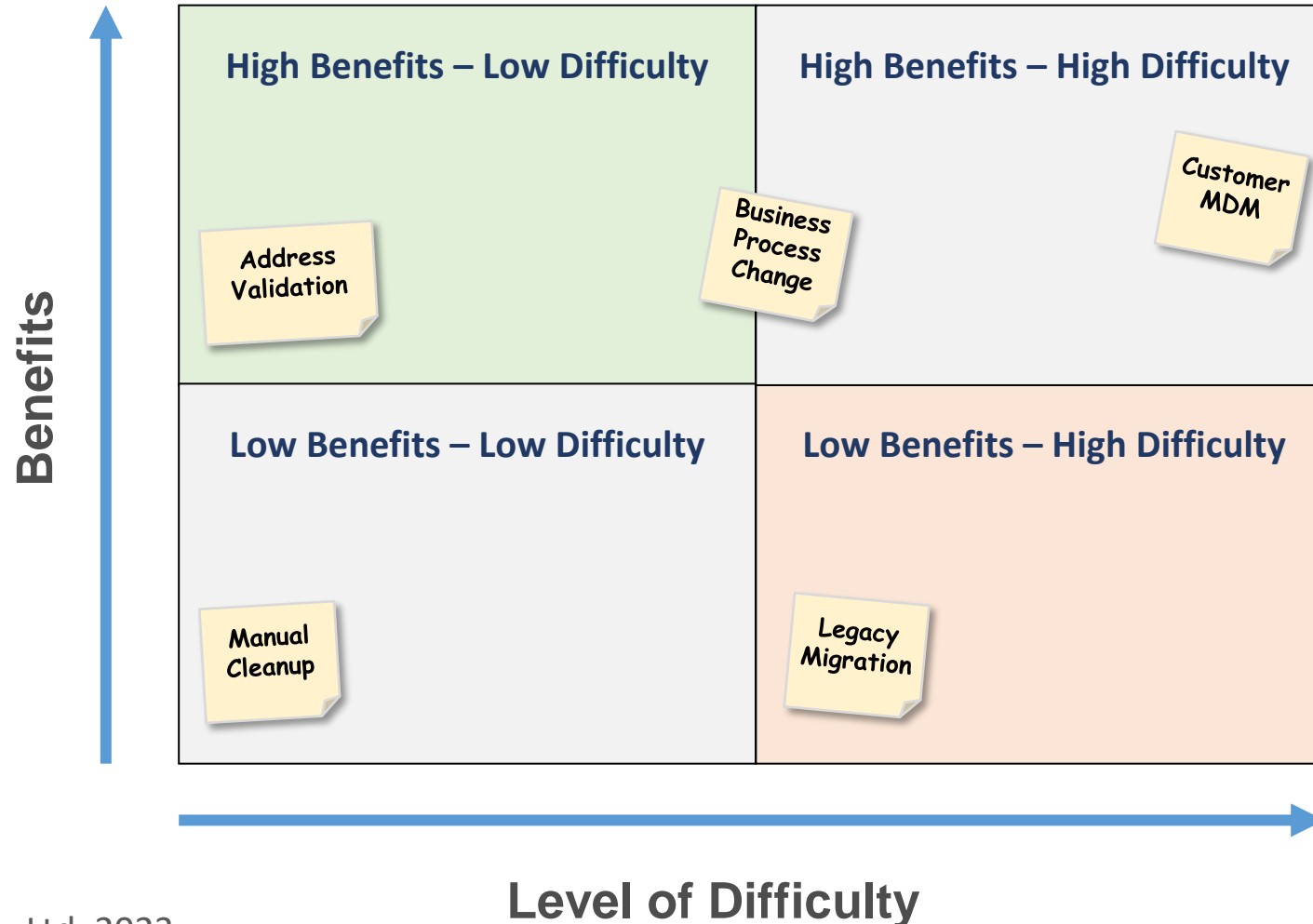
KEY MESSAGE:

Focus & Purpose: the Pareto Principle

80% of business benefit can often be delivered through improving the quality of **20%** of the data – concentrate on the **20%** that really matters (good candidates are often shared master data, reference data etc.)

Setting Priorities: Priority Grid

Priorities based on Benefits vs. Level of Difficulty can often be easily determined via a workshop activity using a Priority Grid



Data Issues & Opportunities Log: Updating Priority

ID	Short Name	Brief Description	Priority Score 1 – High Benefits / Low Difficulty 2 – High Benefits / High Difficulty 3 – Low Benefits / Low Difficulty 4 – Low Benefits / High Difficulty
1	Customer Data Duplication	Both in the CRM platform and the Customer Data Warehouse there are known customer record duplications, mainly caused by marketing and salespeople not being able or willing to search for an existing customer record. One estimate is that up to 25% of CRM customer records are duplicates. Data Warehouse duplication unknown.	<p style="text-align: center; font-size: 2em; color: blue;">2</p> <p style="text-align: center; color: blue;">Further analysis needed to understand the root causes of data duplication. Conducting an initial data cleanse of the CRM platform may be a potential quick win and a starting point for a more strategic solution</p>
2	Product Data Inconsistencies	Product data is held and processed in several different platforms and systems. Each system has its own set of product reference codes. Many inconsistencies have been caused by the company making several acquisitions over recent years. No attempt has yet been made to standardize codes despite wide awareness that this causes problems for the business.	<p style="text-align: center; font-size: 2em; color: blue;">2</p> <p style="text-align: center; color: blue;">There is a widely accepted need for a single, consistent reference source for all product data. The first step is to get data owners to agree definitive product codes and then to design a solution embracing people, process and technology to implement it</p>
3	Poor Data Training of Data Entry People	Data entry people are distributed throughout the company in many different siloed parts of the organization. This has led to many data entry people being unaware of where data they enter is used across the business, other than within their own immediate functional areas.	<p style="text-align: center; font-size: 2em; color: blue;">1</p> <p style="text-align: center; color: blue;">Consistent data literacy and data entry awareness training could be introduced for all existing data entry people and new starters. This could be done quite quickly to include around 30 data entry people.</p>

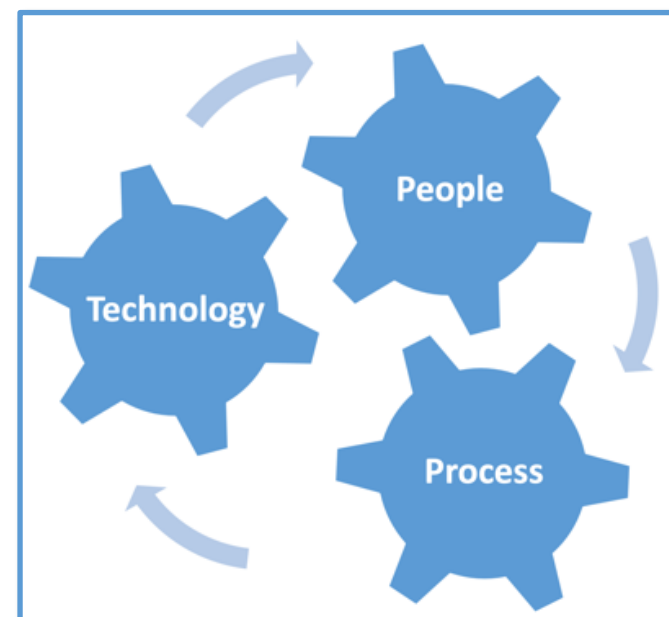
A2E Step 4: Develop

DESIGN & IMPLEMENT IMPROVEMENTS

- Create data quality improvement team to include:
 - Business stakeholders (Data producers, consumers and others, e.g. process owners)
 - IT stakeholders – SMEs, DBAs etc.
 - Other specialists as required (e.g. Data Protection Officer if Personal Data involved)
 - Note: It is important to align with Data Governance Initiatives & Roles (e.g. Data Owners, Data Stewards)
- Re-analyze current problems
 - Perform root cause analysis
- Design and implement improvements
 - Design and implement changes
 - Identify Business Rules & set data quality KPIs
 - Measure improvements against KPIs
 - Revisit the business case to log benefits
 - Identify future improvements
 - Produce case study

POTENTIAL OUTPUTS & TOOLS

- Root Cause Analysis diagrams
- Updated business cases & case study
- Data Quality KPIs and thresholds based on the 7 Data Quality Dimensions and identified Business Rules
- Data Improvement Plans



Data Improvement Plan

A **Data Improvement Plan** is a formal plan to specify and manage improvements to a specified data domain and / or data problem area

The benefits of a Data Improvement Plan are that it:

- Sets out goals and expectations for data improvement
- Acts as a focal point for all data improvement activities
- Prioritizes improvement activities
- Can be used to track improvements and communicate successes
- Can evolve to align with the changing needs of the business

Data domain improvement plans can be rolled up to form the core of a company wide Data Quality Improvement Program

ANONCO DATA IMPROVEMENT PLAN			
DATA AREA / ELEMENT	PRODUCT		
DATA STEWARD	Anne Wilson		
CONTENTS			
Context of Plan			Page 2
Data Area / Element Analysis			Page 4
Key Issues & Problems			Page 7
Improvement Actions			Page 9
Key Success Measures & Targets			Page 14
Plan Monitoring & Corrective Actions			Page 15
Version Control			
Version No.	Date	Comment	Changes marked
0.2 (Draft)	31/05/2020	Updated after DQ Steering Group Review	YES

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



Measuring Data Improvements

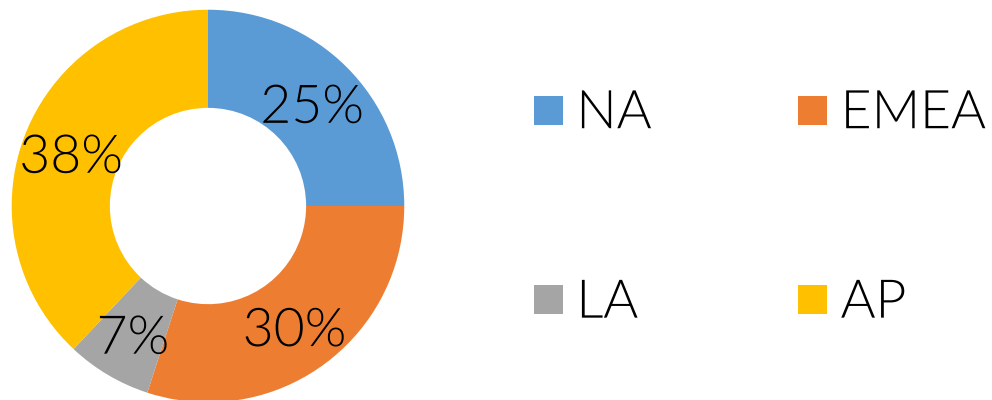
Align Data Quality Metrics to Business Improvements

Business Driver: Improving Customer Data for Marketing Launch Campaign

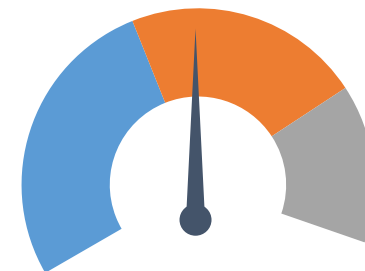
KPI	Current	Target	Status	Business Benefits	Type
Number of duplicate customer records	2,000,000	1,000	Red	<ul style="list-style-type: none"> • Correct # of customers for sales estimations • Better single view of customer for integrated social media campaign • Reduce cost of physical mailing by \$20K 	<ul style="list-style-type: none"> • Cost savings • Brand Reputation • Marketing Innovation
Incorrect Salutation (Mr, Ms, etc.)	5,000	1,000	Yellow	<ul style="list-style-type: none"> • Customer satisfaction & Brand reputation harmed by incorrect salutation. • Targeted marketing campaigns by gender. 	<ul style="list-style-type: none"> • Brand Reputation • Campaign Effectiveness
Incorrect address/location	10,000	500	Yellow	<ul style="list-style-type: none"> • Lower return rate on physical mailings • Better targeted marketing by region. 	<ul style="list-style-type: none"> • Cost Savings • Campaign Effectiveness
Missing Sales Rep Assigned	500	100	Green	<ul style="list-style-type: none"> • Ability for Sales to execute on customer leads • Revenue growth 	<ul style="list-style-type: none"> • Sales Effectiveness

Example Data Quality Dashboard

DUPLICATE ACCOUNTS BY REGION

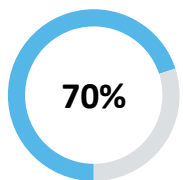


WEEKLY COMPARISON

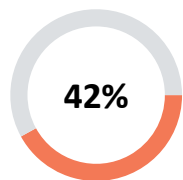


DIMENSION SCORES

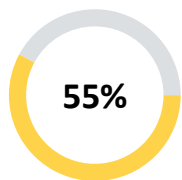
COMPLETENESS



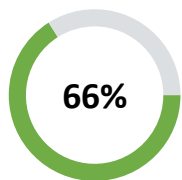
CONSISTENCY



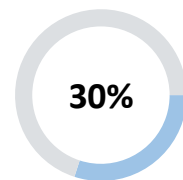
CONFORMITY



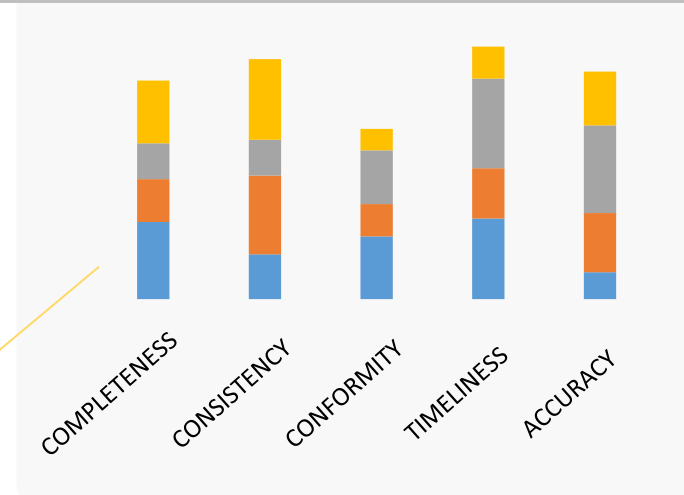
TIMELINESS



ACCURACY



RULE COUNTS BY DIMENSION



Drill down to result set that Data Steward can action

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** Data Mesh or Data Mess? Separating the Reality from the Hype
- **April** Master Data Management - Aligning Data, Process, and Governance
- **May** How do Data Governance & Data Architecture Support Each Other?
- **June** Why You Need Data Management – Getting Executive Buy-In
- **July** Artificial Intelligence and Machine Learning – Building the Right Architectural Foundation
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Designing Data for Business Intelligence & Analytics – Where the Star Schema Fits in a Modern Data Architecture
- **December** Enterprise Architecture vs. Data Architecture



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** to your management
- 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's 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?