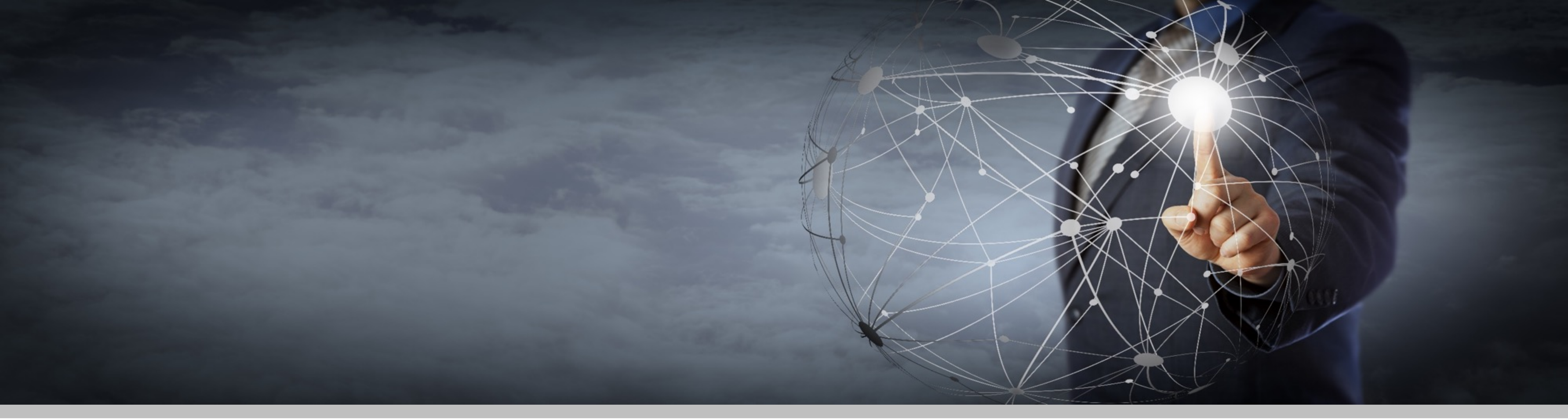


Tackling Data Quality Problems: As Simple As A2E

Adept Events Training

Nigel Turner

Principal Consultant EMEA, Global Data Strategy

Tuesday 10 May 2022

AdeptEvents

Global Data Strategy

The Business of Data

Scene Setting & Introductions

SELF INTRODUCTIONS

Can you introduce yourself by sharing:

- Your name and job role
- Your experience of data management / data quality
- Primary expectation of the course:

WHAT DO YOU WANT / EXPECT TO GET OUT OF IT?

NIGEL TURNER - Role & Credentials

- 35 years experience in IT & Business Strategy; 27 years in Data Management and Data Quality; previously a college and University lecturer in UK and Canada
- Initiated and coordinated BT's enterprise wide data governance and information quality improvement programme
- Subsequently ran a 200 strong Information Management & CRM practice serving BT's global business customers
- Later VP of Strategic IM at Trillium Software, Principal Business Consultant at IPL & Principal IM Consultant at FromHereOn / Enterprise Architects
- Now Principal IM Consultant EMEA at Global Data Strategy and Committee Member of DAMA UK
- Live in Cardiff, Wales & researching Data Governance as an MPhil student at Cardiff University School of Computer Science & Informatics

EXAMPLE COMPANIES WORKED WITH

BT

HSBC

Foreign & Commonwealth Office

Intellectual Property Office

British Gas

First

DNV

UNIVERSITY of GUELPH

Valeo

UISCE ÉIREANN : IRISH WATER

Course & Learning Objectives

- Understand what ‘fit for purpose’ data is, and is not
- Describe the dimensions of data quality
- Know the main causes of poor data quality
- Highlight the impact of poor data quality on individuals and organisations
- Understand the relationship between data quality and other data management disciplines, with particular emphasis on data governance
- Highlight the shortcomings of traditional ways of tackling poor data quality and the importance of a holistic approach, involving people, process and technology
- Learn the five steps of the A2E methodology and how to apply it to identify, prioritise and address data quality problems
- Specify and apply the main activities and deliverables of each of the five steps
- Be able to understand and develop business rules to baseline data quality and to set improvement thresholds
- Be aware of software tools that can help to support and automate the A2E approach

Course Agenda (Indicative)

SESSION	TITLE	START TIME	END TIME
1	Scene Setting & Introductions	9:00	9:20
2	Data Quality: Myths & Realities	9:20	10:00
	Break	*10:00*	*10:10*
3	Holistic Approaches to Data Quality Improvement	10:10	10:20
4	The A2E Approach: **A**SSESS	10:20	11:00
	Break	*11:00*	*11:10*
5	The A2E Approach: **B**ASELINE	11:10	11:40
6	The A2E Approach: **C**ONVERGE	11:40	12:00
	Break	*12:00*	*12:10*
7	The A2E Approach: **D**EVELOP	12:10	12:35
8	The A2E Approach: **E**VALUATE	12:35	12:45
9	The Future of Data Quality	12:45	12:50
10	Summary & Conclusions	12:50	13:00

Data Quality: Myths & Realities

QUESTION

What is Data Quality?

Suggested Definitions?

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 organisation and its stakeholders

ACTIVITY

Data Management & Measures: Defining 'Fit for Purpose' data?

- Work through the following examples
- What do you think are desirable and achievable levels of data quality 'fitness for purpose' in each?
- Answer as a percentage from 0% to 100%
- Prepare a one sentence justification for your answer

Use Cases 1 & 2: ‘Fit for Purpose’

USE CASE 1

You are a marketeer within a consumer products business. Your company is about to launch a new product. You choose initially to target this product at your existing customer base.

To do this you decide to send an email to every customer on your marketing database. However, some customers may have moved without informing you, and you know you have duplicate records on the database. What should your data quality target be and why?

ANSWER	**%**	**REASON:**

USE CASE 2

You are a senior radiographer in a major hospital. You are responsible for ensuring that patients receive the correct dose of radiation at each treatment session. Before each treatment you check previous doses given against the proposed current dose to check consistency and query any variation with the patient’s consultant before administering it.

ANSWER	**%**	**REASON:**

Use Cases 3 & 4: ‘Fit for Purpose’

USE CASE 3

You are a student records manager in a College of Further Education. Your job is to ensure that marks obtained by students on courses run by your college are recorded in a central database. You strive to enter records accurately but know that as you put marks manually into the database occasionally you enter a wrong score. In any case the lecturers who provide you with the marks themselves make mistakes. Eventually you know that someone will notice any errors (often the student) and the marks can be corrected.

ANSWER	**%**	**REASON:**

USE CASE 4

You are a sales manager in a pharmaceutical manufacturing company who sells a variety of products to health-related organisations. At the end of every month, you need to forecast sales volumes and revenues for the following month. This forecast is used by Finance to predict monthly revenues and cashflow and Operations to set their manufacturing targets. You know that despite your frequent reminders, sales people do not always record their sales promptly, and sometimes enter sales figures incorrectly, only correcting them when they submit their sales into the quarterly bonus calculation exercise.

ANSWER	**%**	**REASON:**

‘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 organisation
- When assessing what ‘fit for purpose’ means in any particular use case, 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
- Where this happens, ‘fit for purpose’ data ideally needs to meet the needs of all its users; where not possible or practicable a compromise should be sought

FIT FOR PURPOSE

ACTIVITY

What can go wrong when data is not ‘fit for purpose’?

Data Quality Horror Stories Quiz

Data Quality Horrors Quiz – Question 1

In April 2021 a report highlighted a that “a serious incident” had occurred at Birmingham airport, UK in July 2020. Three planes took off that day, but the total weight of each of these planes was underestimated by an average of 1,200kg per flight. As weight is key in determining the take off speed this error could have caused a potential catastrophe.

This problem was the result of:

a) ***Pilot error as the pilots misread the weight calculations provided***
b) ***The system which calculated the weights had inbuilt business rules that assumed every passenger with the prefix ‘Miss’ was a child***
c) ***The weight information provided was wrongly labelled in Pounds (lb) rather than Kilograms (kg)***
d) ***The weight of the passenger luggage in the hold had been omitted from the calculations***

THE CORRECT ANSWER IS

b)

Horror Story 1: Plane Stupid

- UK Air Accidents Investigation Branch (AAIB) report (April 2021) declared a ‘Serious Incident’ at Birmingham airport, UK
- Report highlighted that 3 Tui flights to Europe in July 2020 had taken off with the weight of the plane load underestimated by an average 1,200kg
- This miscalculation could have caused a ‘serious incident’ on take off as it determines take off speed, thrust etc.
- Problem happened because all passengers with the title ‘Miss’ were automatically assumed to be children and not adults
- A child’s standard estimated weight is 35kg; an adult 69kg
- Tui described it as ‘ a simple flaw in its IT system’
- In reality, there was a serious problem with its data quality business rules!
- Tui have now introduced manual validation of all passengers at check in to ensure adults titled ‘Miss’ are changed to ‘Ms’ on the passenger roster

Data Quality Horrors Quiz – Question 2

In February 2021 Liam Thorp, a 32 year old living in Liverpool, was invited by letter to receive a priority Covid vaccination as he was 'morbidly obese'. Liam was 6 feet 2 inches tall and weighed around 12 stone.

This error occurred because:

a) ***The Health Board had confused him with another person in Liverpool also called Liam Thorp who was 'morbidly obese'***

b) ***Liam Thorp had deliberately exaggerated his weight to jump the vaccine queue***

c) ***The Health Board had recorded his height as 6.2 centimetres (cm) rather than his actual height of 6 feet 2 inches***

d) ***In estimating Liam Thorp's Body Mass Index (BMI) an internal calculation error in a system had resulted in him getting a recorded BMI of 28,000; the BMI needed to be classed as 'morbidly obese' is 40 and above.***

THE CORRECT ANSWER IS

c)

Horror Story 2: Covid data quality 'short'comings

Beatles statue
City of Liverpool

- Liam Thorp made headline news in the UK in Feb 2021
- Received a priority invite for a Covid-19 vaccination because he was medically classed as 'morbidly obese'. He clearly isn't!
- The reason – his local health board had recorded his height as 6.2 centimetres and not his real height of 6 feet 2 inches
- This made his Body Mass Index (BMI) 28,000, calculated by his weight / height
- A BMI of 40 and above is classed as 'morbidly obese'
- This made headline news across the UK in February 2021
- Now corrected, and he's back in his rightful place in the vaccine queue!

Liam Thorp
32 years old
Liverpool
resident

> "I can see the funny side of this story but also recognise there is an important issue for us to address'
>
> Chair of the Liverpool Clinical Commissioning Group (leading the city's vaccine roll out)

Data Quality Horrors Quiz – Question 3

In April 2018, the UK bank TSB migrated all its customers to a new customer management platform. The migration was completed in one weekend.

Within days it became clear that the migration had been a disaster because:

a) ***Many business and personal customers could no longer access their bank accounts online***

b) ***Many customers could access the accounts of other customers***

c) ***Many customers who could access their accounts found that money had been withdrawn from their accounts***

d) ***All of the above***

e) ***None of the above***

THE CORRECT ANSWER IS

d)

Horror Story 3: Banking on Failure

- TSB undertook a major data migration of 5.4 million customers and their 1.3 billion records to a new platform on the weekend of 21/22 April 2018
- At first, the migration was hailed as a great success…
- But emerged that:
    - 2 million TSB customers could not access their accounts
    - Many customers who could access account information were presented with information belonging to other customers
    - Customers found money had been wrongly transferred into or out of their accounts
    - Recovery activities still ongoing
    - TSB expected the disaster to cost ‘tens of millions of pounds’ and had already experienced significant loss of customers
- Key cause: Data was not profiled, analysed and fully standardised before migration began so data quality was unknown

“The conversion of the systems – the data and the interface accessing the data, clearly had not been well-tested before it went online,” ***Industry Expert***

Horror Story 3: AND it got even worse…

- TSB Board commissioned an independent report by Slaughter & May (Commercial law firm)
- Published report in November 2019
- Main outcomes:
  - Eventually lost 80,000 customers
  - £130 million paid to customers in compensation
  - Total cost to TSB estimated at £366 million
  - CEO forced to resign
  - IT department closed down and IBM brought in to run IT services

Data Quality Horrors Quiz – Question 4

In a wide ranging survey in 2017, The Harvard Business Review asked 75 executives across 75 different organisations to extract 100 records from core systems in their companies to analyse if the records contained data quality errors.

The results of the survey showed that the percentage of the 100 records which contained one or more critical data quality errors was:

a) 17%

b) 47%

c) 77%

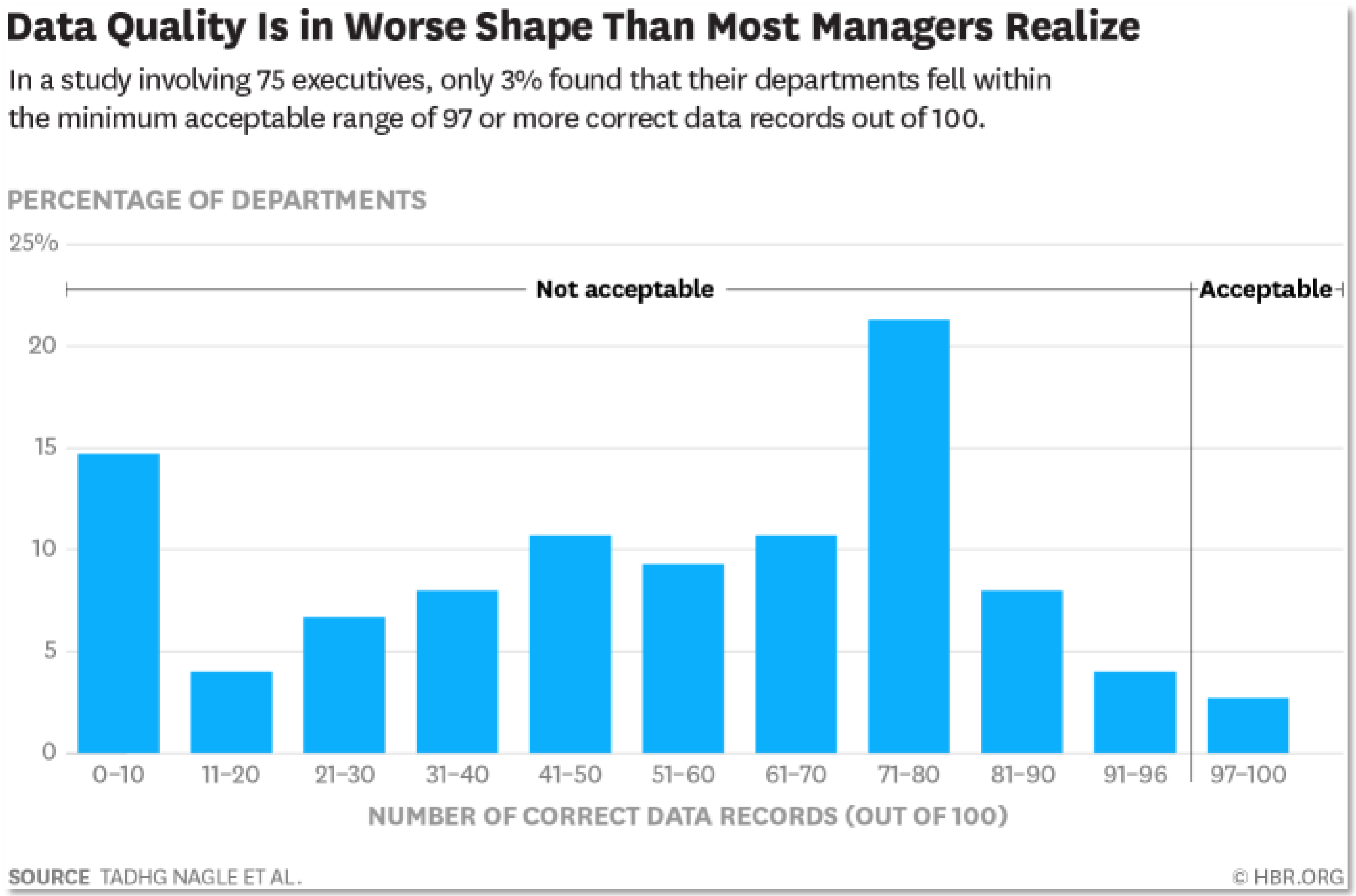
d) 97%

Harvard Business Review

THE CORRECT ANSWER IS

d)

Horror Story 4: Evidence of Failure



Data Quality Is in Worse Shape Than Most Managers Realize

In a study involving 75 executives, only 3% found that their departments fell within the minimum acceptable range of 97 or more correct data records out of 100.

SOURCE TADHG NAGLE ET AL.

© HBR.ORG

Source:
Only 3% of Companies' Data Meets Basic Quality Standards

Tadhg Nagle, Thomas C. Redman & David Sammon

Harvard Business Review
September 11 2017

Data Problems – Real Stakeholder Feedback 2020 / 2021

“We should be checking the data before we enter it on the system but we don’t have the time”

“There is a lack of appreciation of what happens to data from the front end to the back end”

“We’ve got lots of data, but it’s hard to connect it”

“We need to be Partner-ready, but every partner meeting discusses the challenges with data”

“Our customers say ‘you’re embarrassing yourselves – the maths are wrong!’”

“If we could just know that we can trust what we are looking at”

“There is no accountability for bad quality data”

“We cannot achieve the growth we need if we cannot sort out the data.”

“A lot of time (spent) fixing things & not enough time to be creative about the future”

“We only know if we have a data problem if a customer contacts us…… [too much marketing activity is done] in blind faith… I press ‘send’ and hope for the best”

“Our systems don’t talk to each other”

The Industry Impact of Poor Data – The Evidence

On average, half of all organisations believe at least 26% of their data is inaccurate

(Source: BARC 2019)

On average, poor data quality costs companies between 15-25% of revenue

(Source: MIT Sloan 2017)

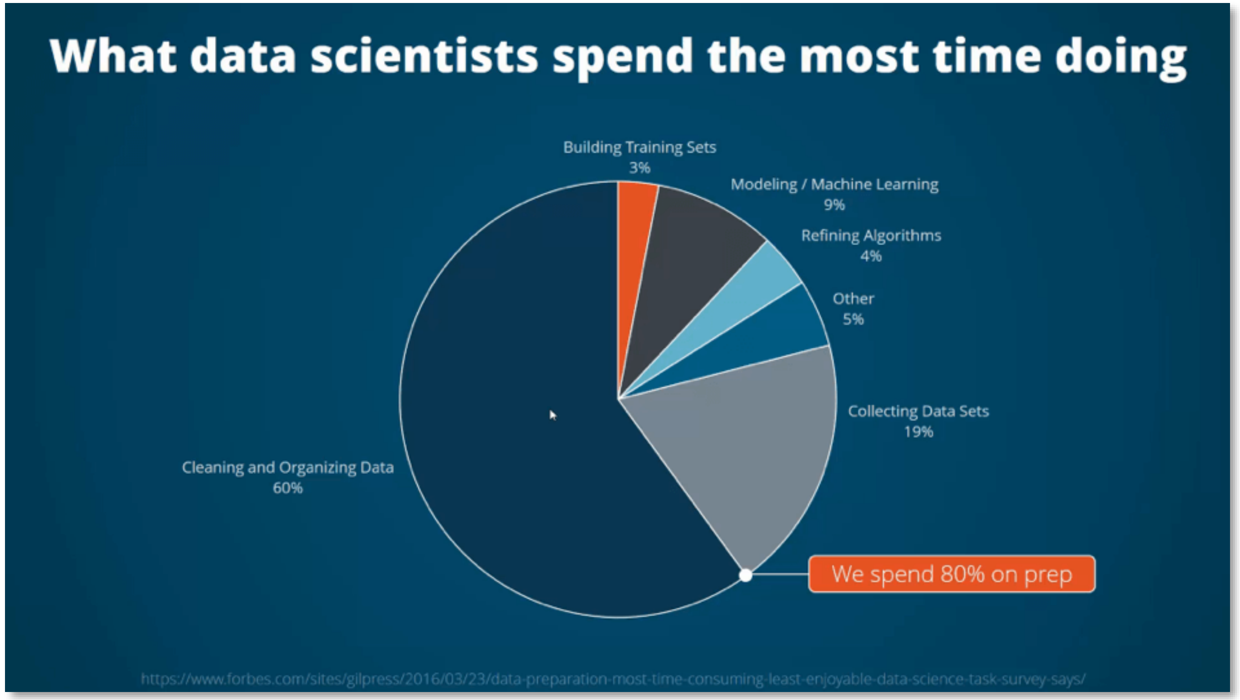
Most organisations believe at least 30% of their customer data is inaccurate; 49% do not trust the data in their ERP or CRM systems

(Source: Experian 2019)

Poor quality customer data is costing UK companies an average of 6% of their annual revenues

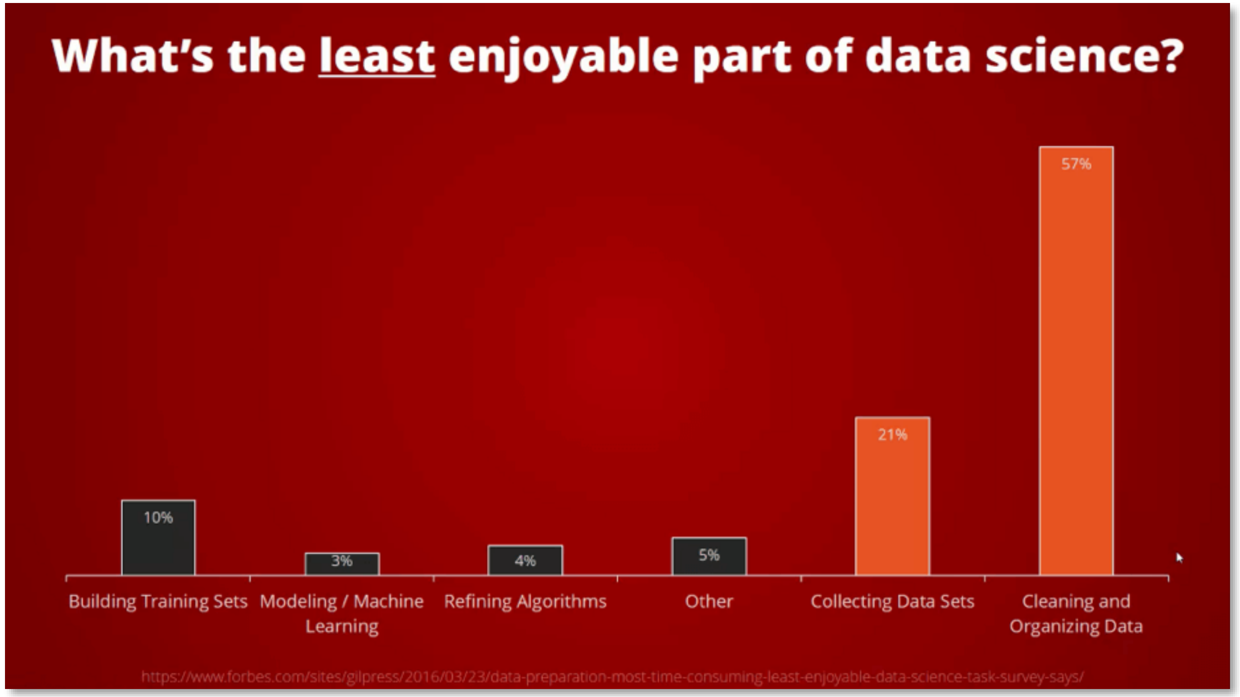
(Source: Royal Mail Data Services 2017)

Data Science: the Data Quality Dimension


What data scientists spend the most time doing

We spend 80% on prep

https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/


What's the least enjoyable part of data science?

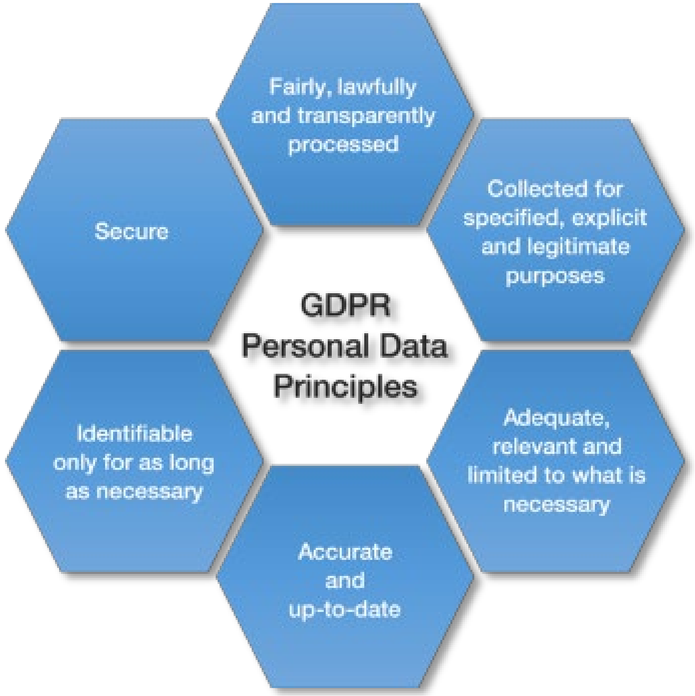
https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

Law & Regulation: a reminder

GDPR 2018: the Six Key Principles

Data should be:

1. Processed lawfully, fairly and in a transparent manner…
2. Collected for specified, explicit and legitimate purposes and not processed further in a manner that is incompatible with those purposes…
3. Adequate, relevant and limited to what is necessary…
4. **Accurate, and where necessary, kept up to date…personal data that are inaccurate… are erased and rectified without delay**
5. Kept in a form which permits identification of data subjects for no longer than is necessary…
6. Processed in a manner that ensures appropriate security…


GDPR Personal Data Principles
Fairly, lawfully and transparently processed
Collected for specified, explicit and legitimate purposes
Adequate, relevant and limited to what is necessary
Accurate and up-to-date
Identifiable only for as long as necessary
Secure

Source: Quant Marketing

Data Quality – why it matters

- Data quality is a foundational data discipline
- Without 'fit for purpose' data quality all other data management disciplines (e.g. BI, MDM, Analytics etc.) can never deliver their promised benefits
- Poor data quality in organisations:
    - Increases costs
    - Leads to operational inefficiencies and lower productivity
    - Reduces revenues and profits
    - Increases risk – poor decision making, breaching legislative / regulatory requirements etc.
    - Damages organisational brand and reputation
    - Alienates customers and deters potential customers
- GARBAGE IN, GARBAGE OUT ……………………………………………………………

The Causes of Poor Data Quality

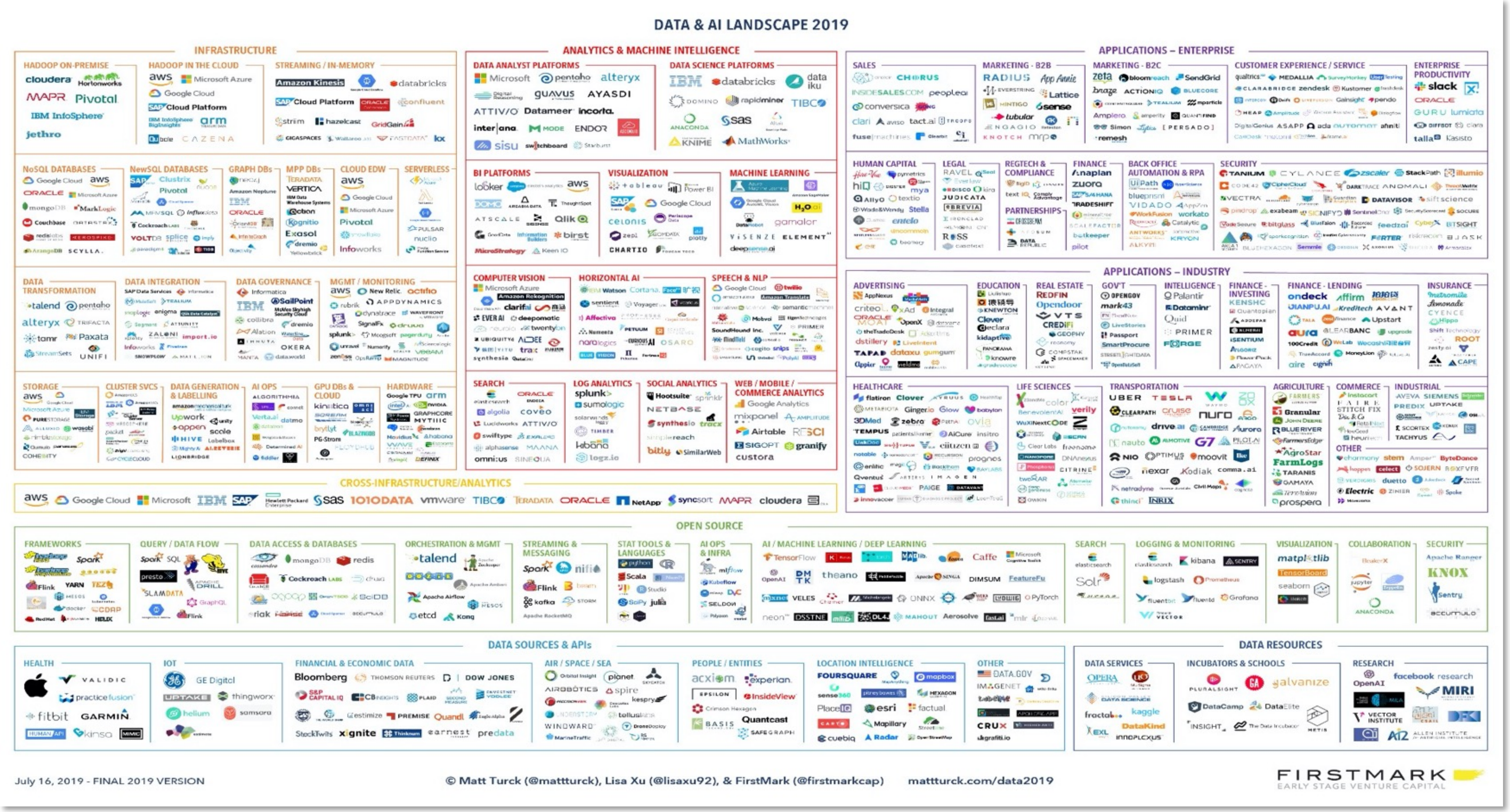
ACTIVITY

So what causes poor data quality?

Why Does Poor Data Persist? 7 Reasons…

1. The data world has become more complex and diffuse
2. The world changes, and data models the world
3. Not recognising that poor data is a business problem, not an IT problem
4. People will make mistakes with data
5. Conflict or absence of common data definitions & metadata context
6. The data Newton's Cradle
7. Lack of accountability for improving data

1. The Data World Becoming More Complex….



July 16, 2019 - FINAL 2019 VERSION

© Matt Turck (@mattturck), Lisa Xu (@lisaxu92), & FirstMark (@firstmarkcap) mattturck.com/data2019

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

2. B2C Data Volatility – UK Facts (1)

67.2 million people in UK (2020)

Live in **27.8** million households (2020)

2. B2C Data Volatility – UK Facts (2)

370,000 people move house each year (2019)

613,936 babies were born (2020)

608,002 people died (2020)

107,599 couples got divorced (2019)

715,000 immigrants come to UK (Yr 2019/20)

403,000 emigrants left the UK (Yr 2019/20)

2. B2B Data Volatility – UK Facts

There are **6.0 million** businesses in the UK (2019)

30% of people change email addresses each year (2018)

726,000 new companies start up every year (2020)

Average decay of UK B2B contact database is **70% per annum** (2020)

3. Too Much Technology Focus - what does IT stand for?

©Cartoonbank.com

"We have lots of information technology. We just don't have any information."

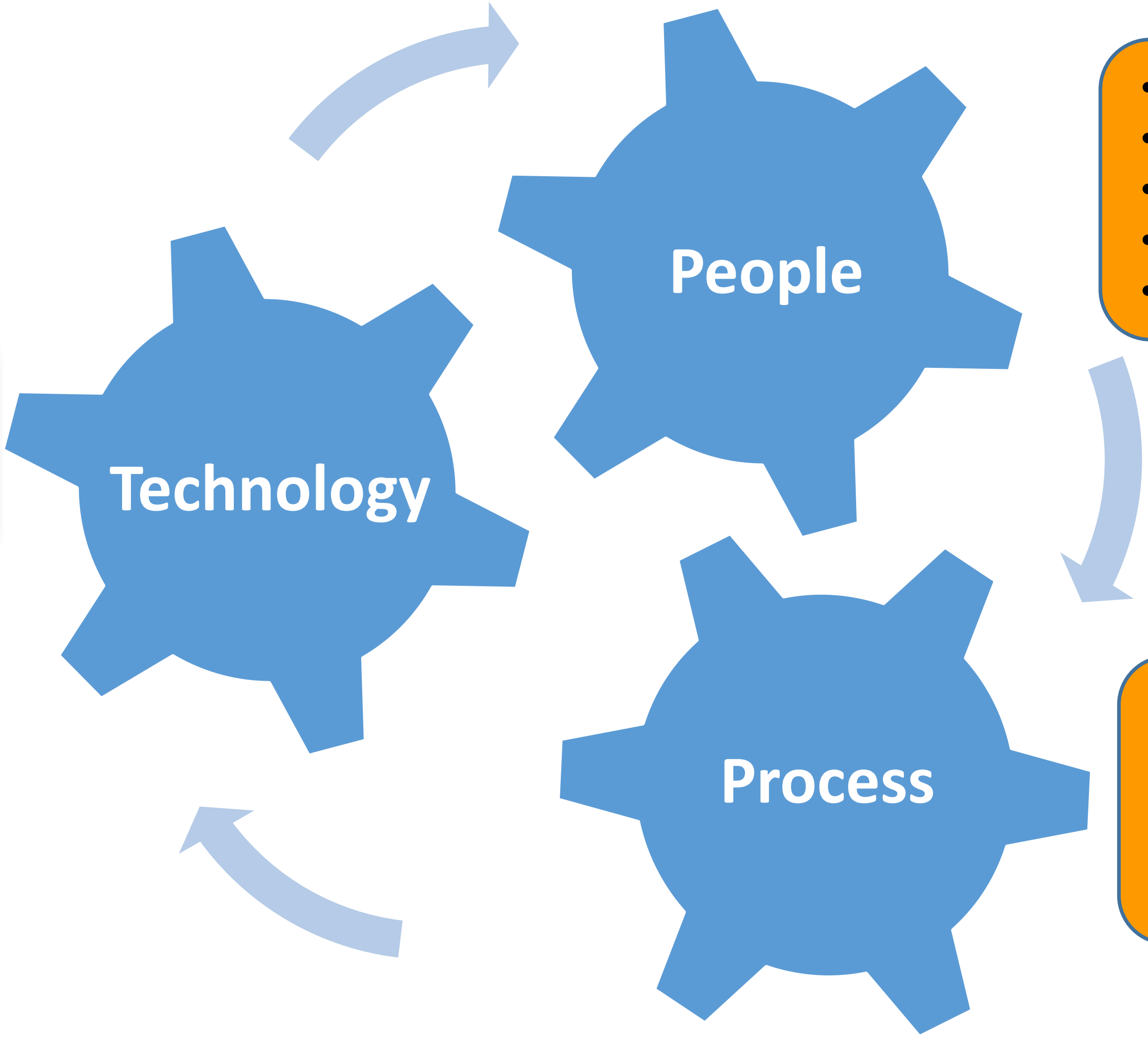
"Most companies have an IT organisation, but they haven't thought of the possibilities of decoupling the 'I' from the 'T' and managing information and technology as separate assets"

Doug Laney, author of 'Infonomics'

3. It’s NOT an IT problem…it’s a business problem

Data Quality problems result from failures in People, Process, and Technology

Fixing them therefore requires addressing People, Process, and Technology



People

- **No data accountability**
- **Human error**
- **Poor training**
- **Internal politics**
- **Denial of failure**

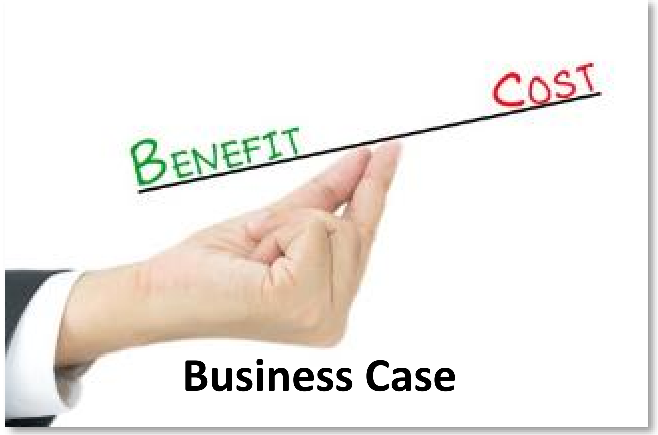
Technology

- **Data capture & U/D failures**
- **Multiple data silos**
- **Interface errors**

Process

- **Poor process design**
- **Process failures**
- **Flawed goal setting**
- **No agreed data standards**

4. People Make Mistakes… a real life example

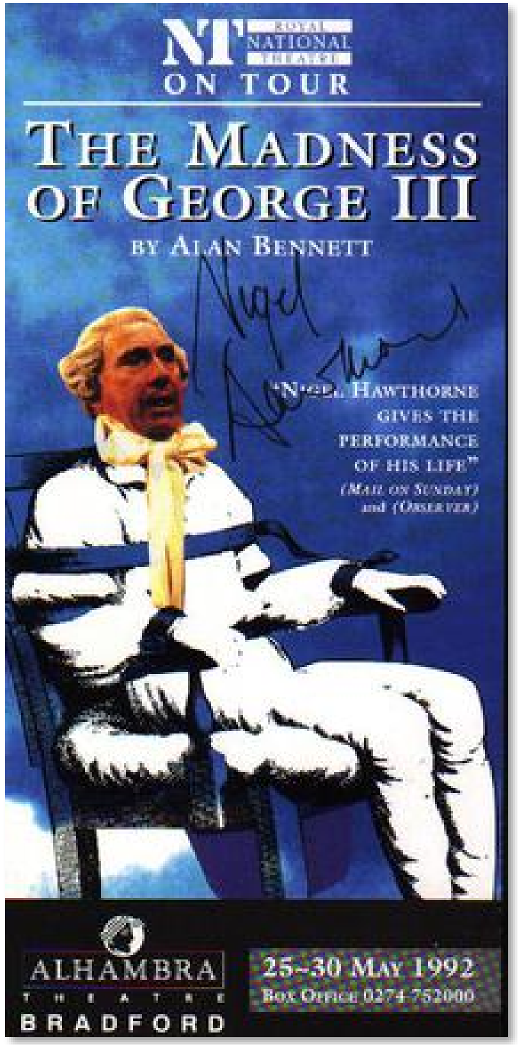


Business Case

Developing a Business Case

Before I make my decision, I'd like to see those meaningless statistics again

5. The Perils of Inadequate Context


NT ROYAL NATIONAL THEATRE
ON TOUR
THE MADNESS OF GEORGE III
BY ALAN BENNETT
Nigel Hawthorne
"NIGEL HAWTHORNE GIVES THE PERFORMANCE OF HIS LIFE"
(MAIL ON SUNDAY) and (OBSERVER)
ALHAMBRA THEATRE BRADFORD
25–30 MAY 1992
BOX OFFICE 0274 752000

On release in the UK, the movie was a great success with critics & at the box office

On its trial release in New York City, the movie flopped, with very poor takings

WHY?

NIGEL HAWTHORNE
HELEN MIRREN
IAN HOLM
His Majesty was all powerful and all knowing.
But he wasn't quite all there.
THE MADNESS OF KING GEORGE

5. Example lack of data definitions & standards… What is a customer?

MARKETING	SOMEONE WHO WE TARGET IN OUR CAMPAIGNS…CURRENT OR PROSPECTIVE CUSTOMER
SALES	SOMEONE WHO BUYS A PRODUCT OR SERVICE
FINANCE	SOMEONE WHO PAYS THE BILLS & INVOICES
OPERATIONS	SOMEONE WHO REQUIRES SERVICE & SUPPORT

How many customers do we have?

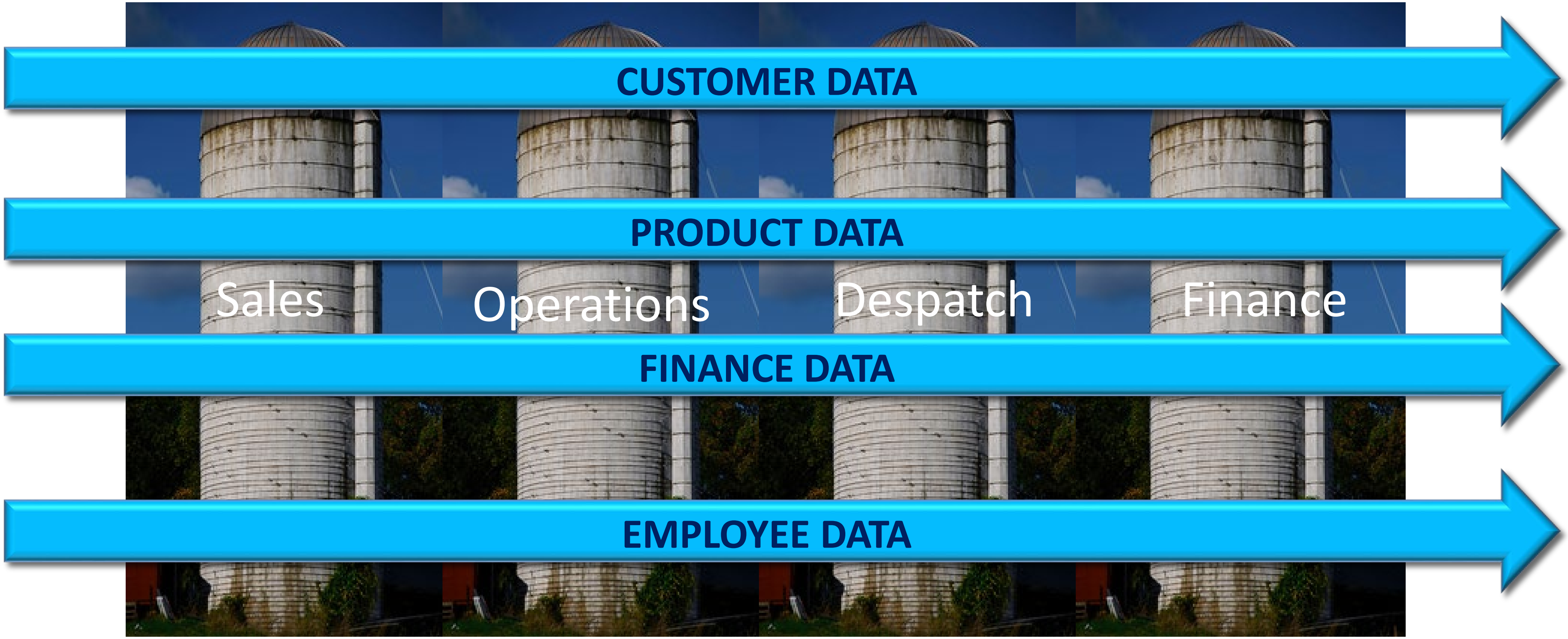
Who are our highest value customers?

How do we create a single customer view?

How should we segment our customer base?

AND SO ON…

6. Why it can be hard - the Horizontal Data Flow

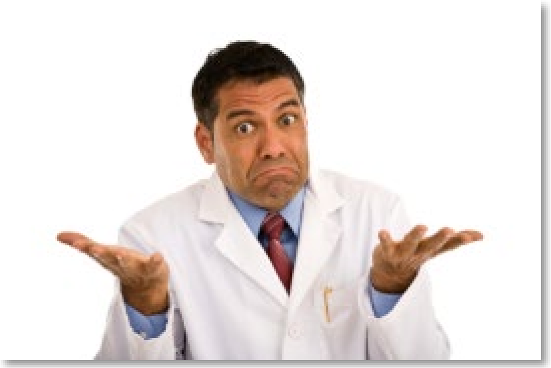


Problems often emerge far away from the cause

7. Lack of Governance & Accountability for data

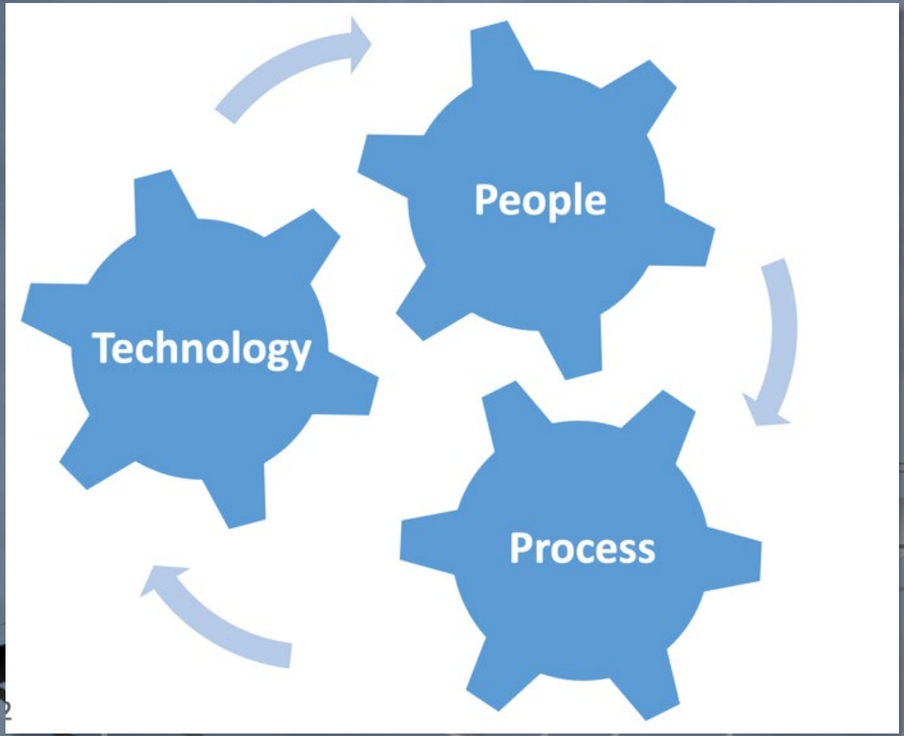
In many organisations, nobody is formally responsible for data and its governance…

so bad data never gets systematically fixed



"If we are all supposed to be responsible, no one is responsible and nothing changes"

(Quote from senior GDS client – Professional Services Organisation 2019)



A Better Way:
Holistic Approaches to Data Quality Improvement

Traditional Approaches to Data Quality

Will continue to have value

- **Inspect data sources** and highlight data deficiencies, omissions and duplication
- **Develop data standards** and common data definitions
- **Build business rules** to enforce and police data standards
- **Automate data cleanse and enhancement projects**, using the business rules defined
- **Embed the standards & rules** into both batch and real-time environments to keep the data clean
- **Produce Data Quality KPIs** and measures to monitor ongoing quality and track trends

Data Quality: Some Common Misconceptions

Data Quality is a stand alone discipline

NOT TRUE – Data Quality is closely interdependent with other disciplines, e.g. Data Governance, MDM, Data Architecture, BI, Analytics, etc.

Data Quality is an IT problem & so IT tools can fix it

NOT TRUE – Data Quality is multi-faceted, caused by process, people and IT issues, so solutions must be holistic and business-driven

Data Quality improvement is a choice

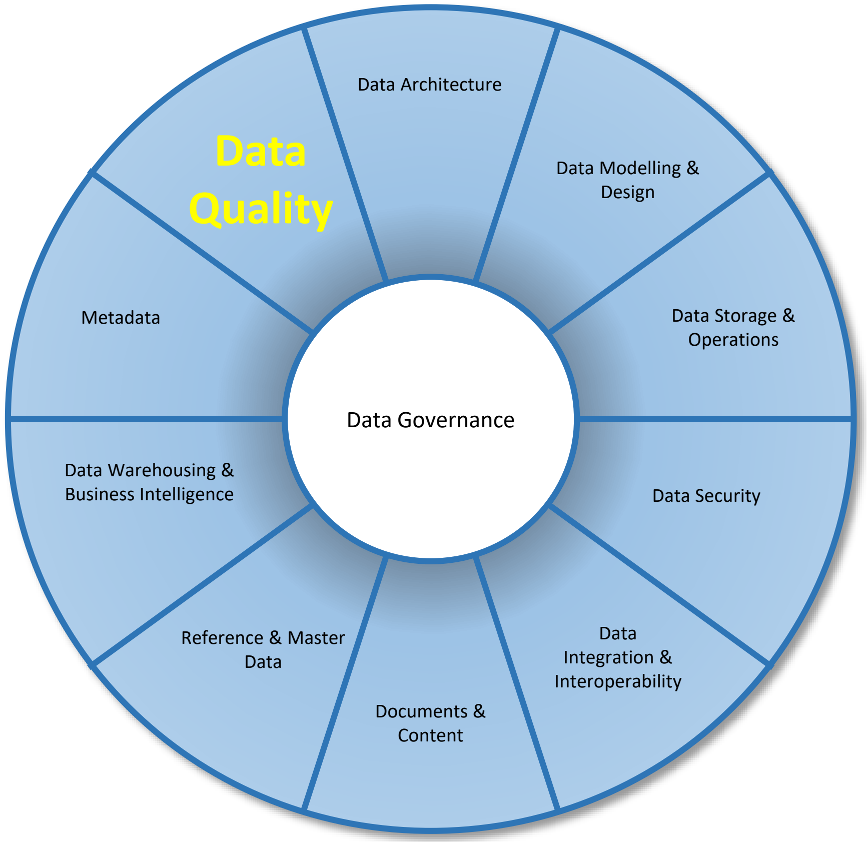
NOT TRUE – all organisations continually do data quality improvement; it's not about **IF** you do it but **HOW** you do it

Data Quality improvement is a project

NOT TRUE – it may start with a project, but it has no end; it must evolve into a Business As Usual (BaU) continuous process of improvement

Data Quality – relationships with other data disciplines

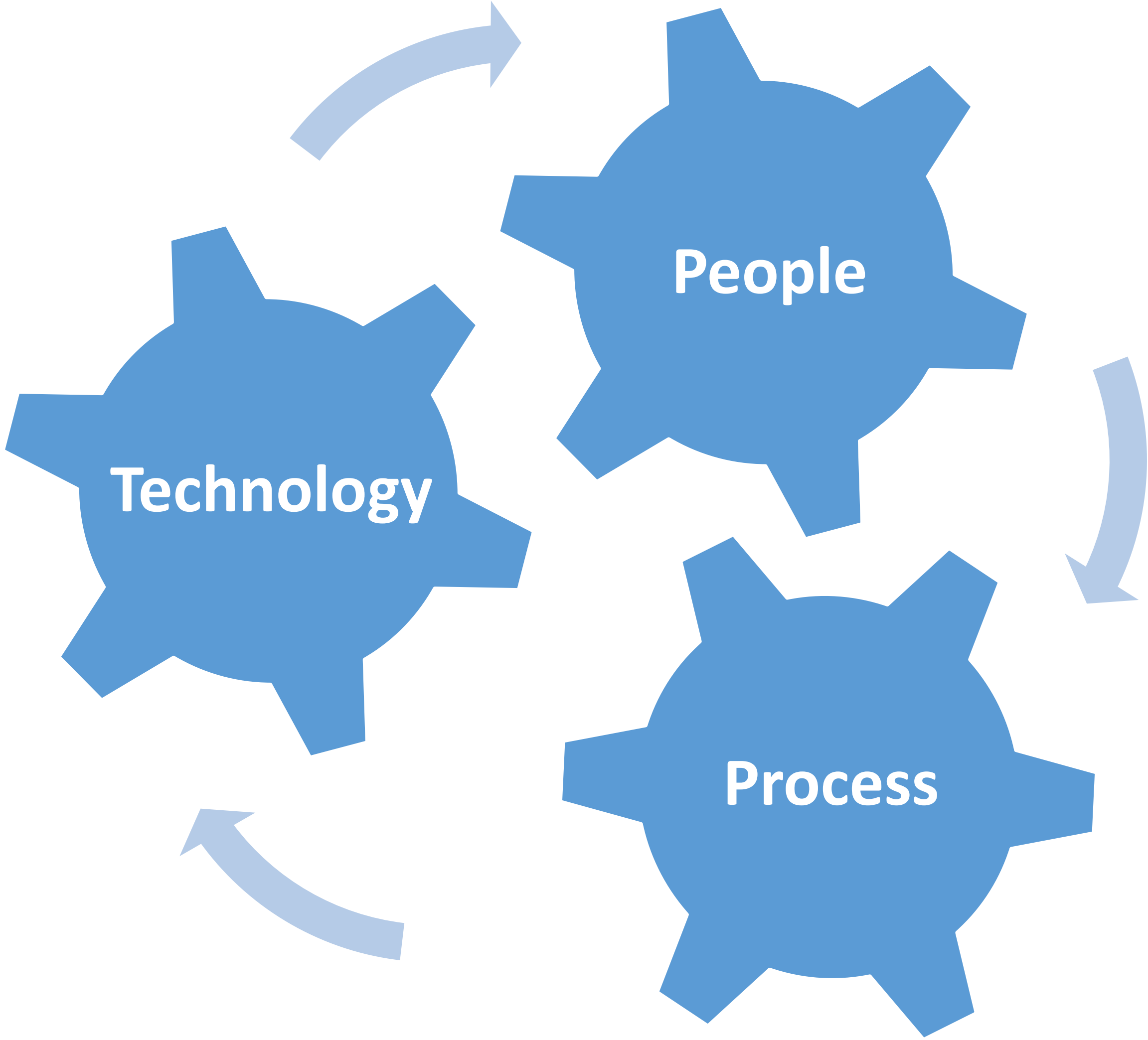
11 DATA DISCIPLINES



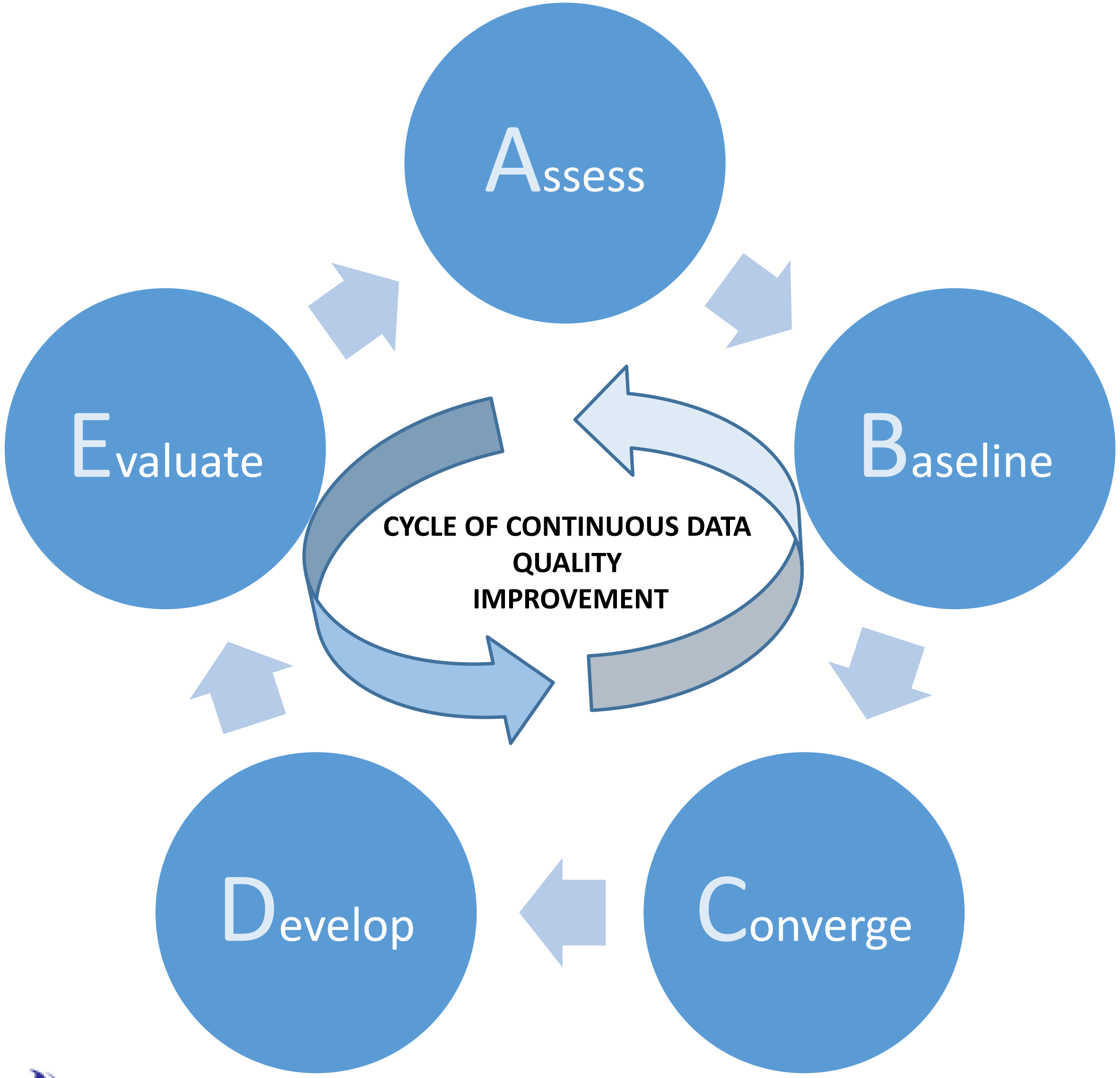
DAMA DMBOK – Version 2 2017

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.
Meta-data	MD provides context & meaning to data and so enhances DQ

The Need to Be Holistic:
Focus on People, Process & Technology



Tackling Data Quality: the Holistic A2E approach

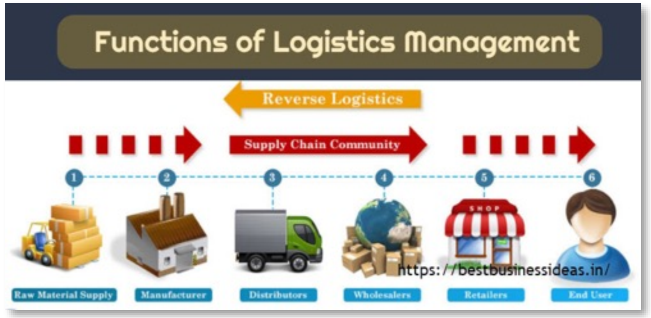


Step	Purpose
Assess Business Usage	Understand what data exists and how it is used within the organisation
Baseline Data Sources	Baseline the current quality of the data and assess how well it is meeting business needs
Converge on Business Critical Areas	Focus priorities to optimise early business benefits 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 & ROI	Regularly measure the data and continue to improve it so that it continues to meet current and future business needs

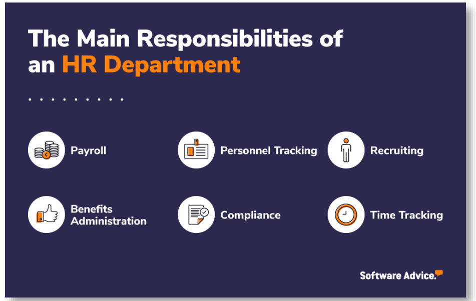
The A2E Approach: Where To Use It

TOP DOWN

ORGANISATION



FUNCTION / PROCESS



The Main Responsibilities of an HR Department

. . . .

- Payroll
- Personnel Tracking
- Recruiting
- Benefits Administration
- Compliance
- Time Tracking

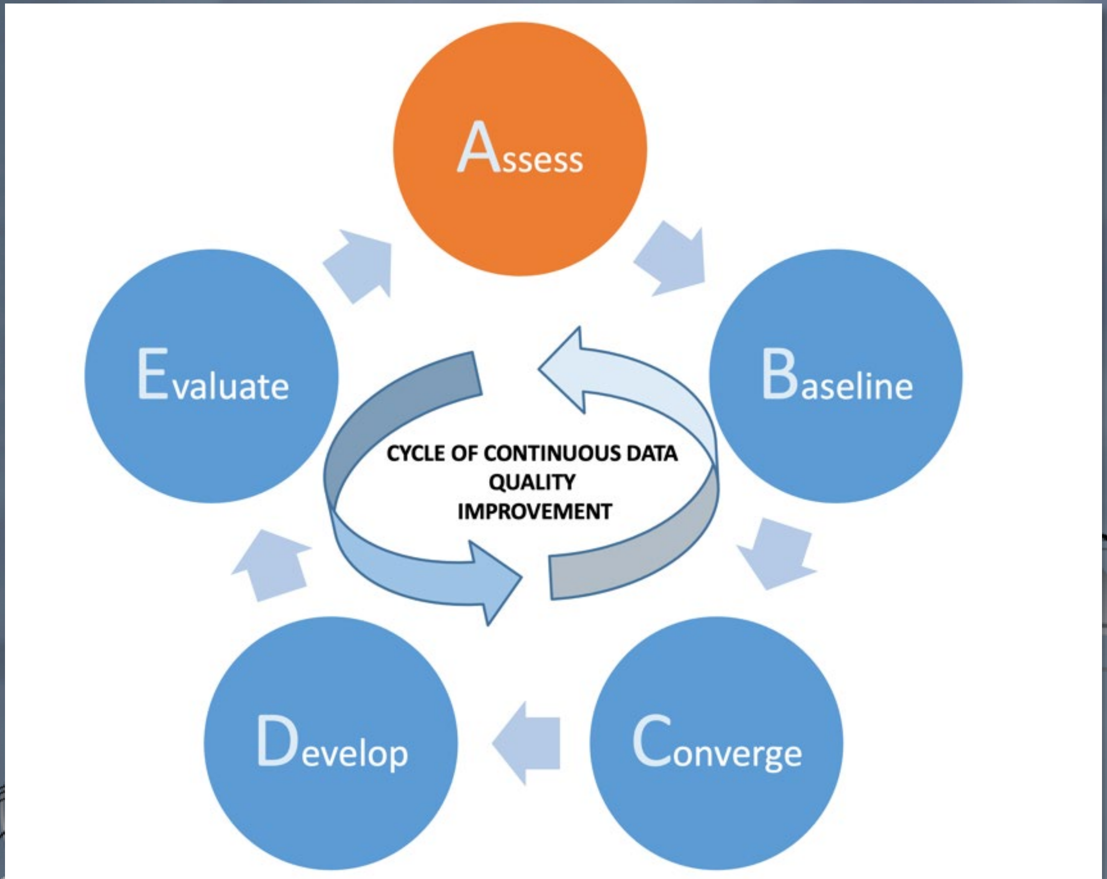
Software Advice.

DEPARTMENT

DATA DOMAIN

SPECIFIC DATA QUALITY ISSUE

BOTTOM UP



The A2E Approach: Step 1 – Assess

A2E Step 1: Assess

ASSESS THE BUSINESS LANDSCAPE

- Understand the business and its primary goals & objectives
- Analyse 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 current data 'fitness for purpose' and establish:
  - Where / how it is captured, stored and processed
  - What's working well
  - What needs to be improved
  - The potential benefits of better data quality
- Create a Data Quality Issues (& Opportunities) Log

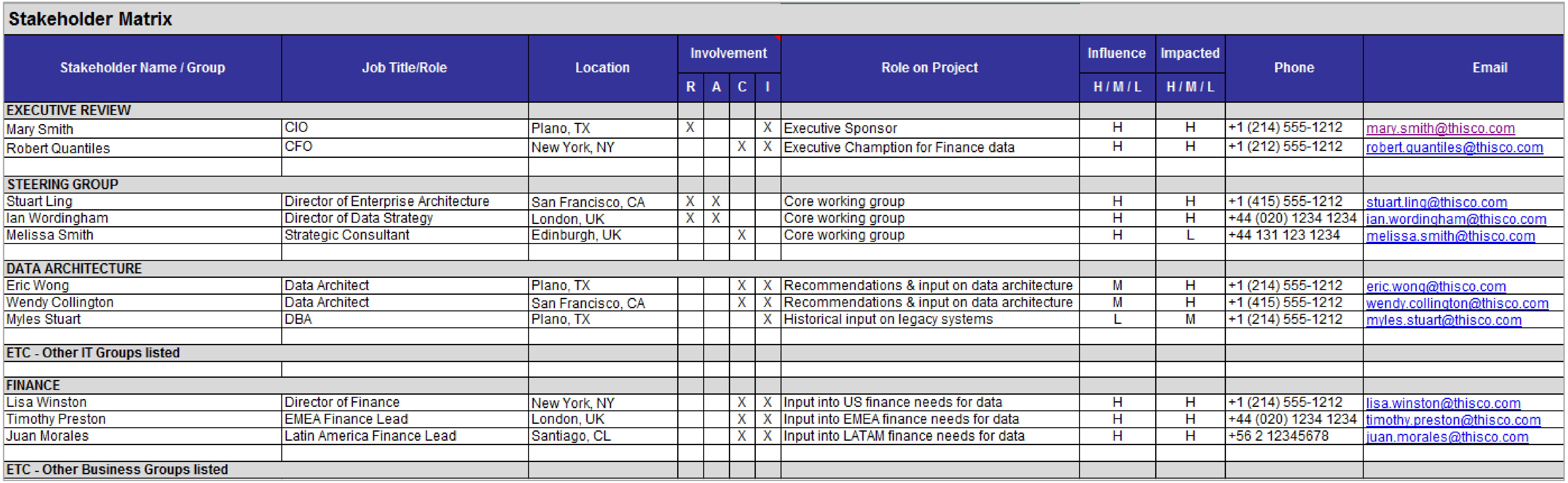
POTENTIAL OUTPUTS & TOOLS

- 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 Log
  - Business Data Model / Conceptual Data Model
  - Business Process Model
  - ROI analysis

How to capture business goals and drivers (1) – Business

- Look at organisation / company websites & documents (external and internal) to highlight:
  - Current Mission & Vision
  - Strategic business aims and goals
  - Current challenges – external and internal
- Consider how these depend on data and its effective management
  - For example, a business goal to **'Increase our revenues from our top 10% revenue generating customers'**
  - Data questions:
    - Can we identify our top 10% revenue generating customers?
    - What data issues may stop us doing that (data quality, data duplication, missing data etc.)
    - What are the business implications if we can't?

Example Stakeholder Matrix



Stakeholder Matrix

Stakeholder Name / Group	Job Title/Role	Location	Involvement				Role on Project	Influence	Impacted	Phone	Email
			R	A	C	I		H / M / L	H / M / L		
EXECUTIVE REVIEW											
Mary Smith	CIO	Plano, TX	X			X	Executive Sponsor	H	H	+1 (214) 555-1212	mary.smith@thisco.com
Robert Quantiles	CFO	New York, NY			X	X	Executive Chamption for Finance data	H	H	+1 (212) 555-1212	robert.quantiles@thisco.com
STEERING GROUP											
Stuart Ling	Director of Enterprise Architecture	San Francisco, CA	X	X			Core working group	H	H	+1 (415) 555-1212	stuart.ling@thisco.com
Ian Wordingham	Director of Data Strategy	London, UK	X	X			Core working group	H	H	+44 (020) 1234 1234	ian.wordingham@thisco.com
Melissa Smith	Strategic Consultant	Edinburgh, UK			X		Core working group	H	L	+44 131 123 1234	melissa.smith@thisco.com
DATA ARCHITECTURE											
Eric Wong	Data Architect	Plano, TX			X	X	Recommendations & input on data architecture	M	H	+1 (214) 555-1212	eric.wong@thisco.com
Wendy Collington	Data Architect	San Francisco, CA			X	X	Recommendations & input on data architecture	M	H	+1 (415) 555-1212	wendy.collington@thisco.com
Myles Stuart	DBA	Plano, TX				X	Historical input on legacy systems	L	M	+1 (214) 555-1212	myles.stuart@thisco.com
ETC - Other IT Groups listed											
FINANCE											
Lisa Winston	Director of Finance	New York, NY			X	X	Input into US finance needs for data	H	H	+1 (214) 555-1212	lisa.winston@thisco.com
Timothy Preston	EMEA Finance Lead	London, UK			X	X	Input into EMEA finance needs for data	H	H	+44 (020) 1234 1234	timothy.preston@thisco.com
Juan Morales	Latin America Finance Lead	Santiago, CL			X	X	Input into LATAM finance needs for data	H	H	+56 2 12345678	juan.morales@thisco.com
ETC - Other Business Groups listed											

RACI *:

R: Responsible

A: Accountable

C: Consulted

I: Informed

How to capture business goals and drivers (2) – Data

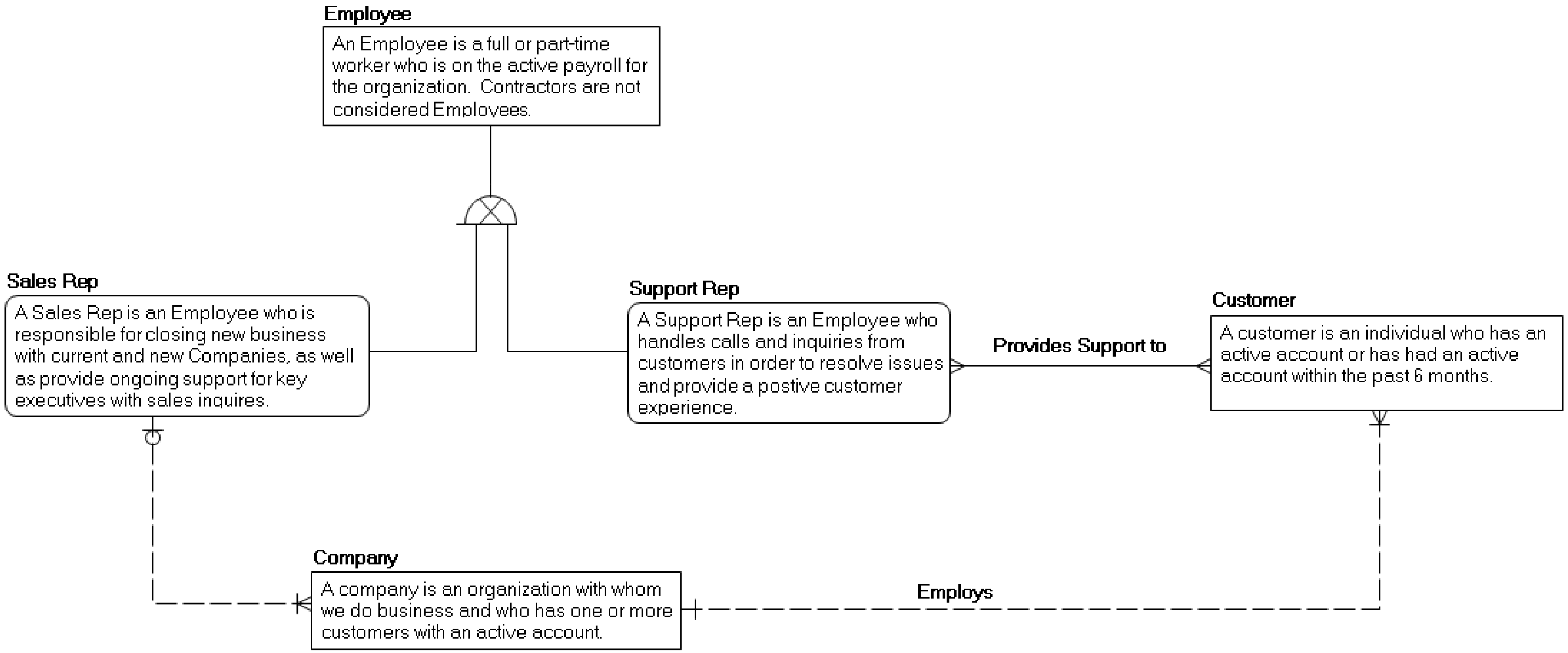
- Identify the primary data stakeholders (Include both Business & IT)
- Set up 1-1 interviews, group interviews or workshops to highlight:
  - How are you currently using or managing data in your role?
  - What's working well?
  - What needs to be improved and why?
  - What future data needs do you have and what opportunities can be taken with better (use of) data?
  - What are the current costs / lost opportunities of current data shortcomings? (Quantify in financial terms if possible)
  - What is your One Wish for data?
- It's important to speak to a wide range of roles across the organisation:
  - Senior Executives
  - Management roles (Business & IT)
  - Front line roles (Business & IT)

Business Data Model (Conceptual)

Communication & definition of core data concepts & their definitions

- A business data model provides core **definitions** of key data objects
- It also shows key **relationships** between data objects
- Even a simple diagram as the one on the right can tell a powerful **"story"**

.... And uncover key **business issues and opportunities**



Employee

An Employee is a full or part-time worker who is on the active payroll for the organization. Contractors are not considered Employees.

Sales Rep

A Sales Rep is an Employee who is responsible for closing new business with current and new Companies, as well as provide ongoing support for key executives with sales inquires.

Support Rep

A Support Rep is an Employee who handles calls and inquiries from customers in order to resolve issues and provide a postive customer experience.

Customer

A customer is an individual who has an active account or has had an active account within the past 6 months.

Company

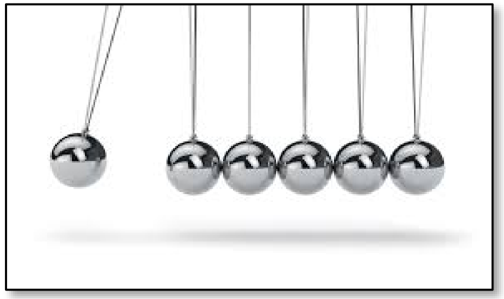
A company is an organization with whom we do business and who has one or more customers with an active account.

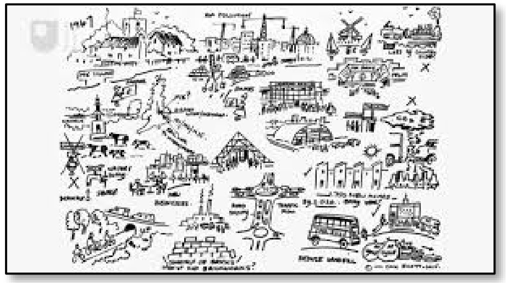
Data Issues Log: Suggested Template from Interviews

Unique ID	Interview No.	Short Name	Description	Impact of the Problem	Raised By

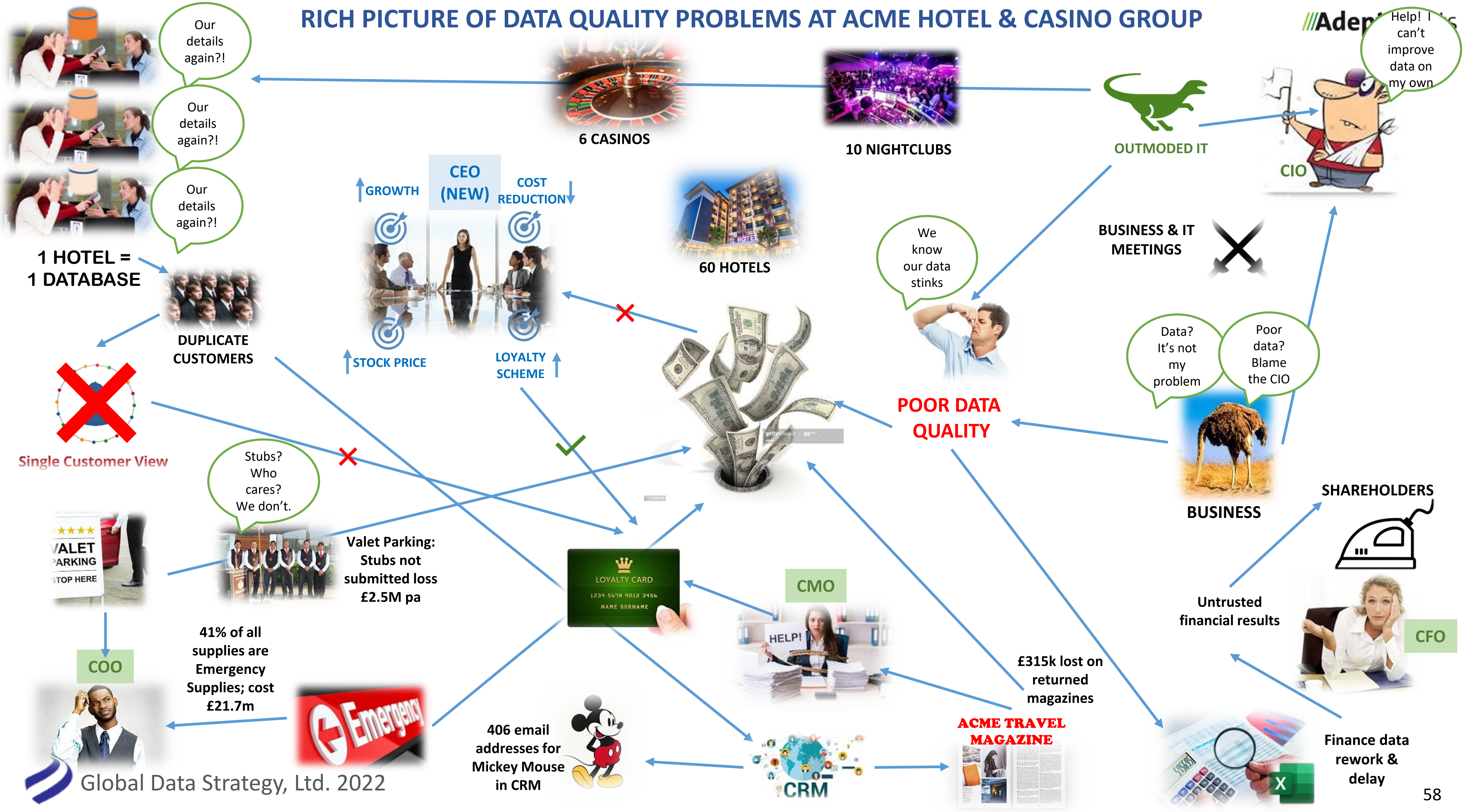
Data Quality Complexity & Value of Rich Pictures

- Data Quality is a ‘messy’ and complex issue:
  - Problems often poorly understood (e.g. data flows and lineage)
  - Lack of information & hard facts (e.g. measures)
  - Large numbers of people involved with differing perspectives (e.g. data producers, data consumers, senior executives, customers, suppliers)
  - Problem ownership unclear (e.g. problem origins and impacts)
- Rich Pictures have great value:
  - Ideal starting point for complex (messy) organisational problems like data quality
  - Holistic, embracing people, process & technology
  - Highlight interconnectedness of problems
- Best initially created in a workshop (whiteboard and coloured pens ideal!) - encourage participants to contribute
- Primary use is to derive ‘problem themes’ to enable focus on key issues

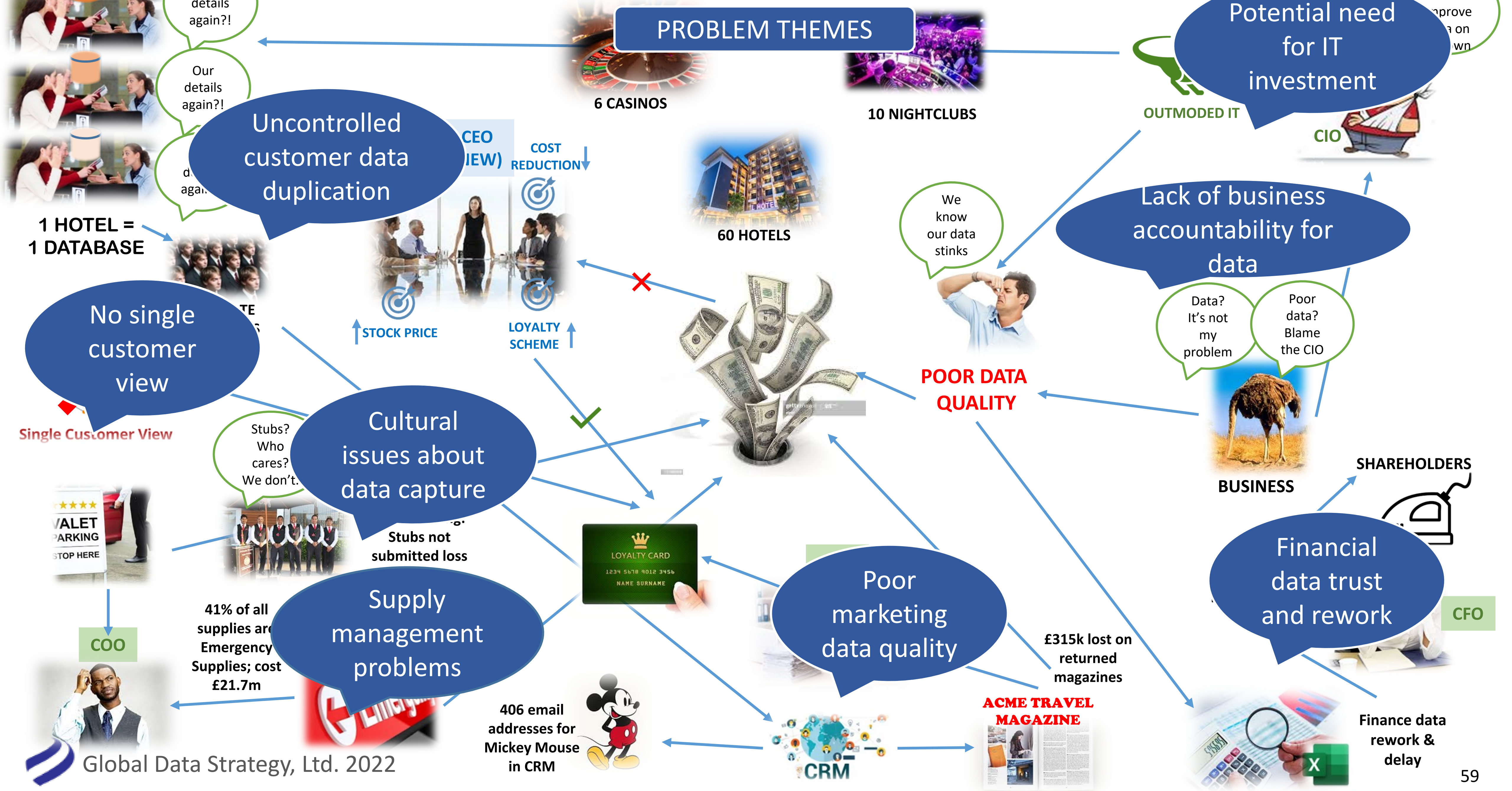




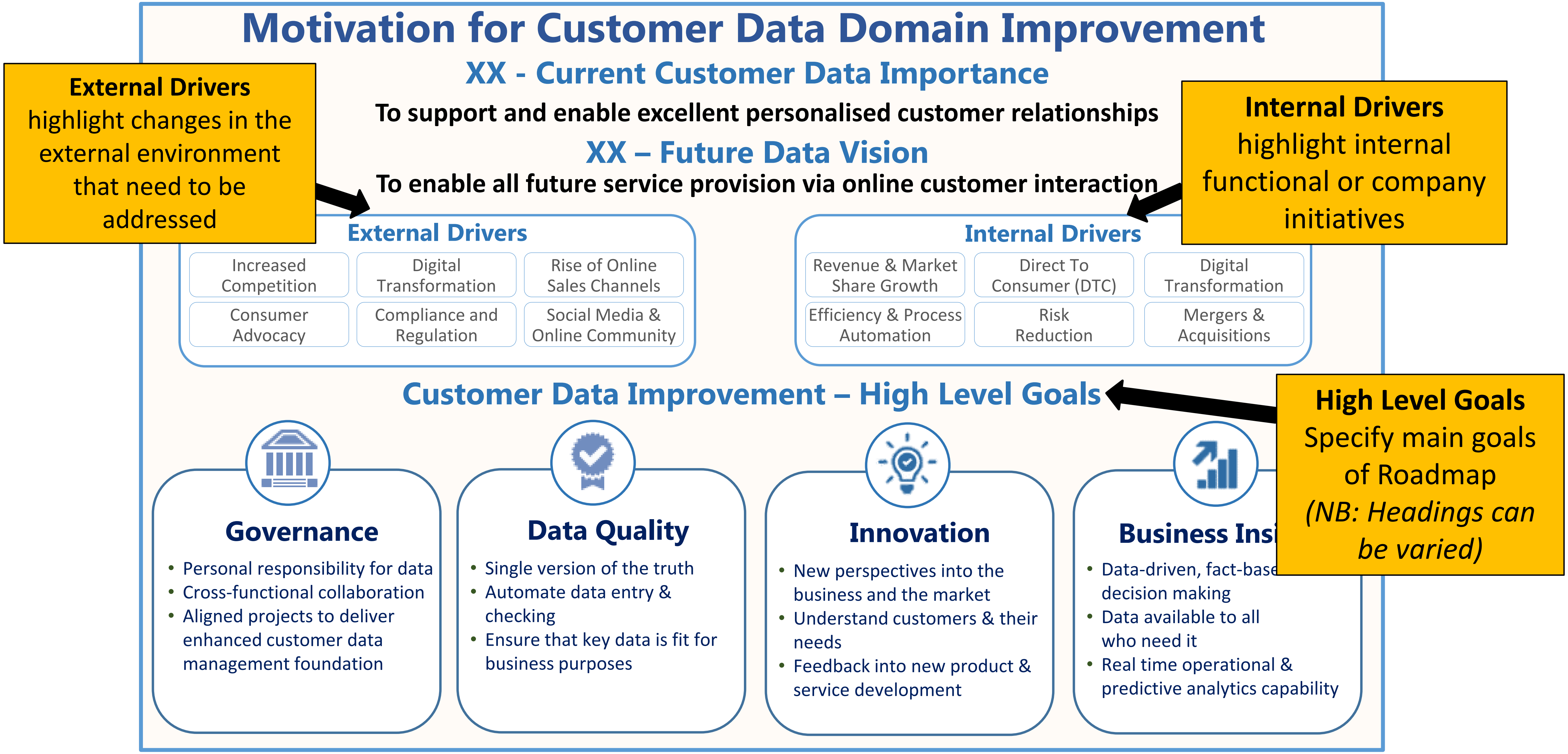
RICH PICTURE OF DATA QUALITY PROBLEMS AT ACME HOTEL & CASINO GROUP



RICH PICTURE OF DATA QUALITY PROBLEMS AT ACME HOTEL & CASINO GROUP


PROBLEM THEMES
Our details again?!
Our details again?!
d… aga…
1 HOTEL = 1 DATABASE
Uncontrolled customer data duplication
CEO
IEW)
COST REDUCTION
STOCK PRICE
LOYALTY SCHEME
6 CASINOS
10 NIGHTCLUBS
60 HOTELS
"Adept…s
Help! I can't …nprove …a on …wn
Potential need for IT investment
OUTMODED IT
CIO
We know our data stinks
Lack of business accountability for data
Data? It's not my problem
Poor data? Blame the CIO
POOR DATA QUALITY
BUSINESS
SHAREHOLDERS
TE
No single customer view
Single Customer View
Stubs? Who cares? We don't.
Cultural issues about data capture
Stubs not submitted loss
★★★★★ VALET PARKING STOP HERE
COO
41% of all supplies are Emergency Supplies; cost £21.7m
Supply management problems
LOYALTY CARD
1234 5678 9012 3456
NAME SURNAME
406 email addresses for Mickey Mouse in CRM
CRM
Poor marketing data quality
£315k lost on returned magazines
ACME TRAVEL MAGAZINE
Financial data trust and rework
CFO
Finance data rework & delay

Data Quality: Real Example Motivation Model



Motivation for Customer Data Domain Improvement

XX - Current Customer Data Importance

To support and enable excellent personalised customer relationships

XX – Future Data Vision

To enable all future service provision via online customer interaction

External Drivers highlight changes in the external environment that need to be addressed

Internal Drivers highlight internal functional or company initiatives

External Drivers

Increased Competition	Digital Transformation	Rise of Online Sales Channels
Consumer Advocacy	Compliance and Regulation	Social Media & Online Community

Internal Drivers

Revenue & Market Share Growth	Direct To Consumer (DTC)	Digital Transformation
Efficiency & Process Automation	Risk Reduction	Mergers & Acquisitions

Customer Data Improvement – High Level Goals

High Level Goals Specify main goals of Roadmap *(NB: Headings can be varied)*

Governance

- Personal responsibility for data
- Cross-functional collaboration
- Aligned projects to deliver enhanced customer data management foundation

Data Quality

- Single version of the truth
- Automate data entry & checking
- Ensure that key data is fit for business purposes

Innovation

- New perspectives into the business and the market
- Understand customers & their needs
- Feedback into new product & service development

Business Insi

- Data-driven, fact-base decision making
- Data available to all who need it
- Real time operational & predictive analytics capability

Making the Business Case

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

Decreasing Costs

- **Wasted labour 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 optimisation through Analytics**
- **Improved marketing campaigns through quality customer data**
- **Data-driven recommendation engines to enhance the sales cycle**
- **Better grant applications through data-driven needs analysis**

Reducing Risk

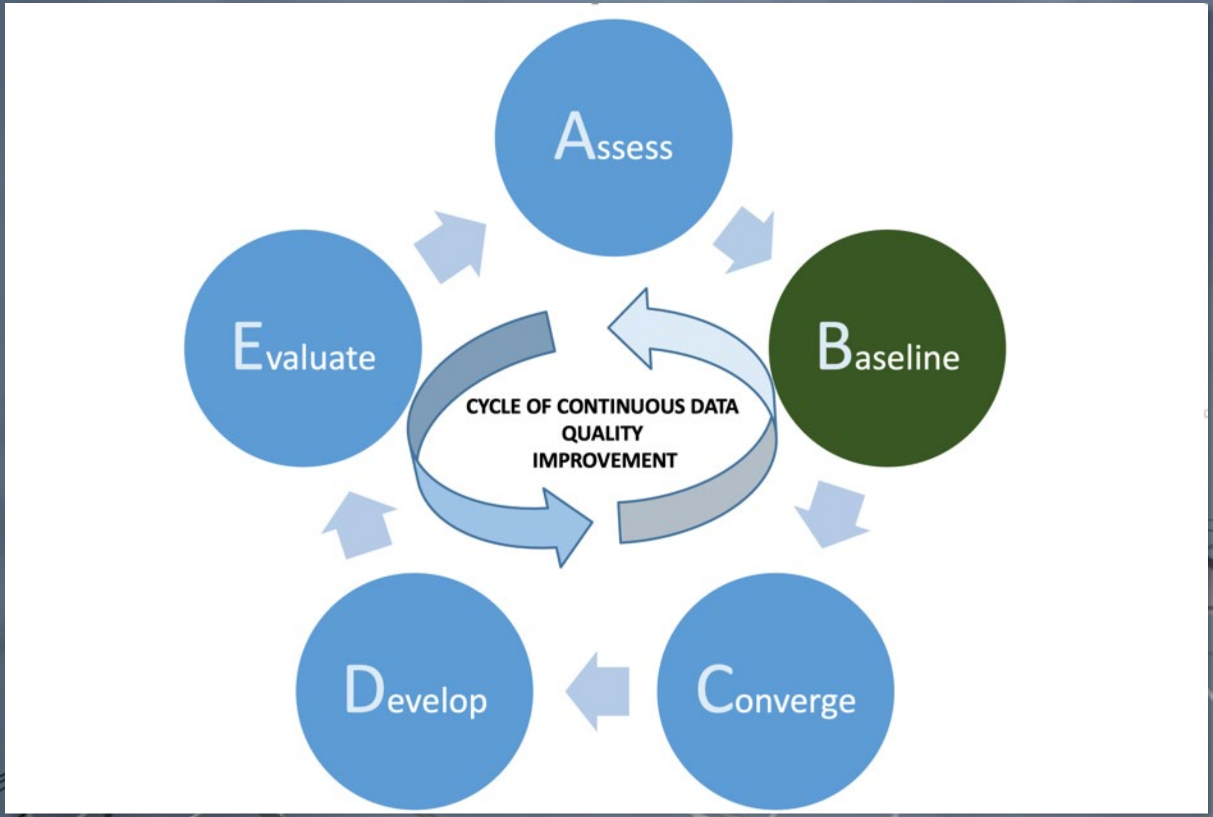
- **Industry regulations**
  (GDPR, HIPAA, BCBS 239, Spice, HIPAA, etc.)
- **Product traceability**
  (Food lineage from farm/catch)
- **Litigation due to data breaches**
- **Health and safety audits**

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 quality
- Make sure to include the "do nothing" option in your analysis



The A2E Approach:
Step 2 – Baseline

A2E Step 2: Baseline

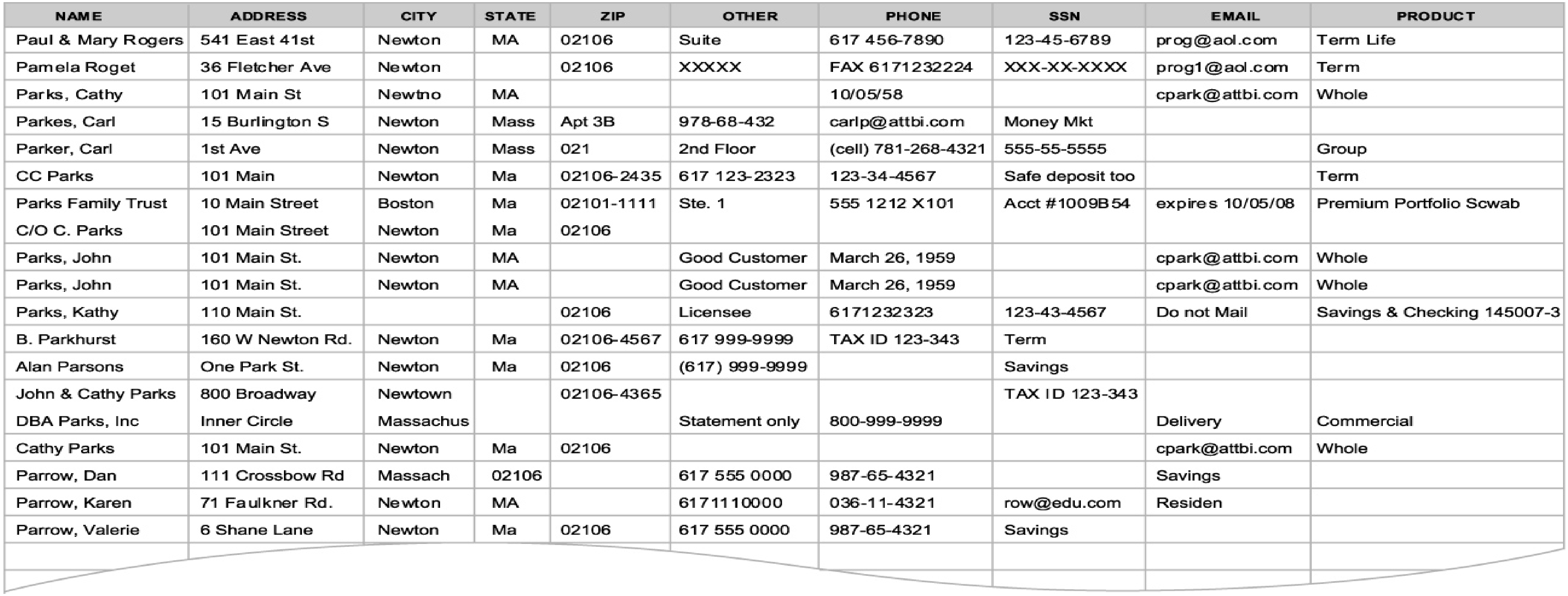
BASELINE CRITICAL DATA SOURCES

- Gives 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 Assessment
  - Profile the data (ideally use a data profiling tool) and focus on key objects and attributes
  - Assess the data according to the 7 Dimensions of Data Quality – see later
  - 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 Log, with quantification of financial costs and other business impacts

Typical CRM Data Quality Issues



NAME	ADDRESS	CITY	STATE	ZIP	OTHER	PHONE	SSN	EMAIL	PRODUCT
Paul & Mary Rogers	541 East 41st	Newton	MA	02106	Suite	617 456-7890	123-45-6789	prog@aol.com	Term Life
Pamela Roget	36 Fletcher Ave	Newton		02106	XXXXX	FAX 6171232224	XXX-XX-XXXX	prog1@aol.com	Term
Parks, Cathy	101 Main St	Newtno	MA			10/05/58		cpark@attbi.com	Whole
Parkes, Carl	15 Burlington S	Newton	Mass	Apt 3B	978-68-432	carlp@attbi.com	Money Mkt		
Parker, Carl	1st Ave	Newton	Mass	021	2nd Floor	(cell) 781-268-4321	555-55-5555		Group
CC Parks	101 Main	Newton	Ma	02106-2435	617 123-2323	123-34-4567	Safe deposit too		Term
Parks Family Trust C/O C. Parks	10 Main Street 101 Main Street	Boston Newton	Ma Ma	02101-1111 02106	Ste. 1	555 1212 X101	Acct #1009B54	expires 10/05/08	Premium Portfolio Scwab
Parks, John	101 Main St.	Newton	MA		Good Customer	March 26, 1959		cpark@attbi.com	Whole
Parks, John	101 Main St.	Newton	MA		Good Customer	March 26, 1959		cpark@attbi.com	Whole
Parks, Kathy	110 Main St.			02106	Licensee	6171232323	123-43-4567	Do not Mail	Savings & Checking 145007-3
B. Parkhurst	160 W Newton Rd.	Newton	Ma	02106-4567	617 999-9999	TAX ID 123-343	Term		
Alan Parsons	One Park St.	Newton	Ma	02106	(617) 999-9999		Savings		
John & Cathy Parks DBA Parks, Inc	800 Broadway Inner Circle	Newtown Massachus		02106-4365	Statement only	800-999-9999	TAX ID 123-343	Delivery	Commercial
Cathy Parks	101 Main St.	Newton	Ma	02106				cpark@attbi.com	Whole
Parrow, Dan	111 Crossbow Rd	Massach	02106		617 555 0000	987-65-4321		Savings	
Parrow, Karen	71 Faulkner Rd.	Newton	MA		6171110000	036-11-4321	row@edu.com	Residen	
Parrow, Valerie	6 Shane Lane	Newton	Ma	02106	617 555 0000	987-65-4321	Savings		

Missing Data

Misfielded data

Lack of data standards

Free format text

Multiple names

Mixed business & personal names

Duplicate data

No complete customer view

Source: Harte Hanks Trillium Software

ACTIVITY

From this extract of a real HR data source, list all the real or suspected data quality problems you can identify

Hint: There are at least 13 problems

ACTIVITY – Extract of HR data source

EMPLOYEE NO	SURNAME	FIRST NAME	GENDER	DATE OF BIRTH	ROLE CODE
802540	Smith	Brian	Female	31/01/56	PM16
YN4176B	Gregg		Male	07/09/80	9999
811609	Patel	Priya	XXXX	25/12/78	AL60
22298	Bothroyd	Bridget	Female	28/08/09	TBD
802540	Smith	Bryan	Male	31/01/56	PM10
855265	Hayes	Leslie	Female	00/00/00	AL76
	Taylor	Kevin	Unknown	12/30/69	US18

Note: Records extracted and anonymised from an actual HR database

ACTIVITY – Extract of actual HR data source

EMPLOYEE NO	SURNAME	FIRST NAME	GENDER	DATE OF BIRTH	ROLE CODE
802540	Smith	Brian	Female	31/01/56	PM16
YN4176B	Gregg		Male	07/09/80	9999
811609	Patel	Priya	XXXX	25/12/78	AL60
22298	Bothroyd	Bridget	Female	28/08/09	TBD
802540	Smith	Bryan	Male	31/01/56	PM10
855265	Hayes	Leslie	Female	00/00/00	AL76
	Taylor	Kevin	Unknown	12/30/69	US18

Key:

🟥	Potential Data Quality Problem
🟨	Potential Duplicate Record

ANSWER: Total number of potential Data Quality problems is 13 or 19, depending on whether Smith is a duplicate record

Quantifying Data Problems: The Value of Data Profiling Tools

- Data profiling tools automate the process of assessing and reporting on the quality of data sources
- The benefits of data profiling include:

➢ Fast processing of large data sets

➢ Complete analysis of an entire data set, so identifies all outliers

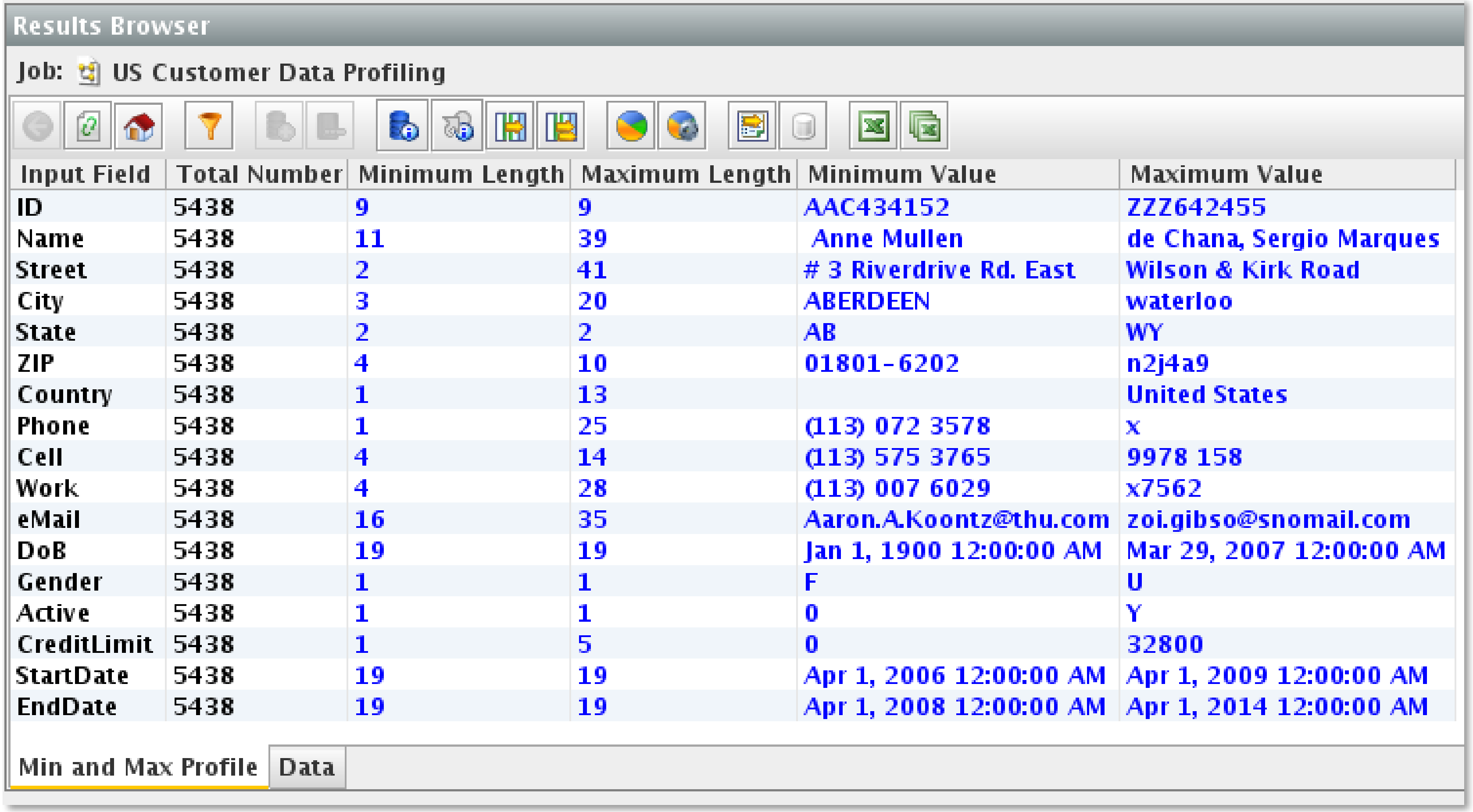
➢ Some profilers enable drill down to individual records / rows

➢ Automatic generation of metadata

➢ Checks conformance of the dataset with business rules (pre-built or added)

➢ 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

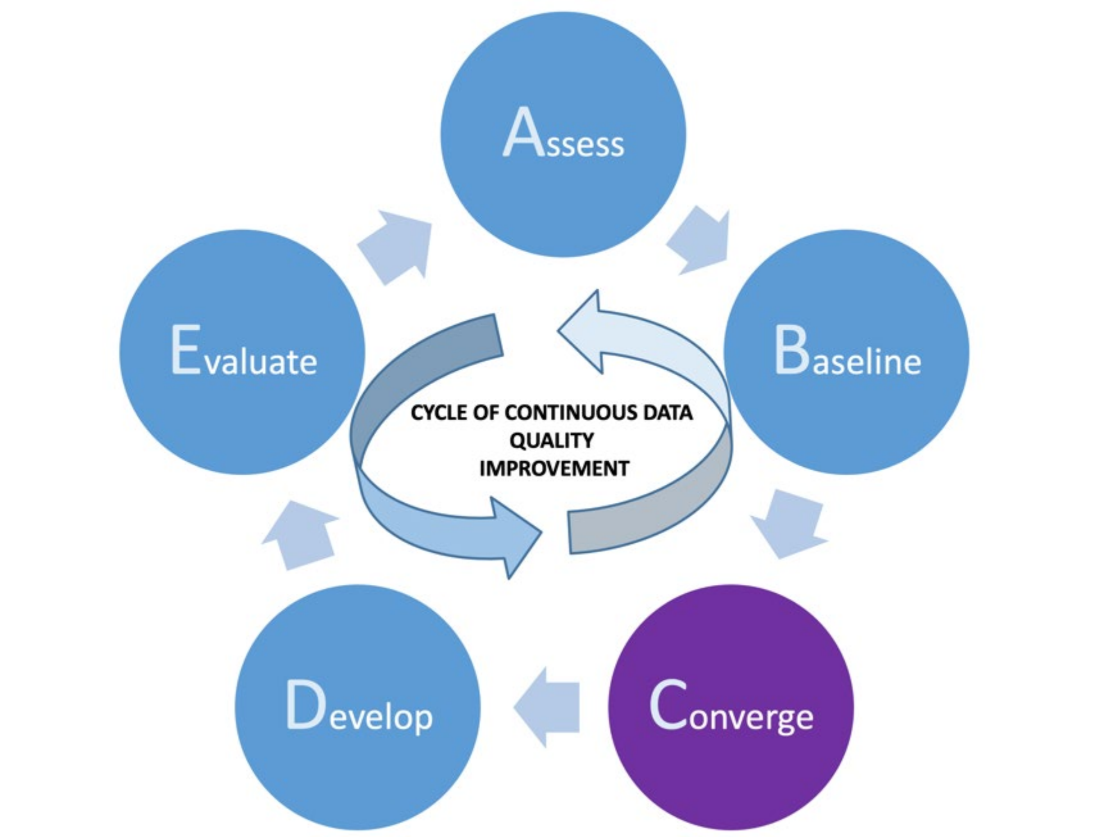
Input Field	Total Number	Minimum Length	Maximum Length	Minimum Value	Maximum Value
ID	5438	9	9	AAC434152	ZZZ642455
Name	5438	11	39	Anne Mullen	de Chana, Sergio Marques
Street	5438	2	41	# 3 Riverdrive Rd. East	Wilson & Kirk Road
City	5438	3	20	ABERDEEN	waterloo
State	5438	2	2	AB	WY
ZIP	5438	4	10	01801-6202	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

Example partial Data Profiling report

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 led by the Data Stewards responsible for the specific data areas (see later)
- 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 (covered later)

The A2E Approach: Step 3 – Converge



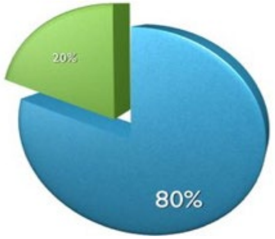
A2E Step 3: Converge

PRIORITISE & FOCUS ON SPECIFIC ISSUES & OPPORTUNTIES

- 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
- Prioritise projects (e.g. Priority Grid)
- Run pilots / proofs of concept
- Identify and run initial DQ improvement projects

POTENTIAL OUTPUTS & TOOLS

- Prioritised 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 & Activities

WHY?

- Ensure clear focus & priority to manage limited time
- Important to make an early impact to gain credibility
- Need to gain support of Data Champions & Steering Group
- Build confidence & experience in DQ principles & approaches

WHAT?

- Look for 'quick wins'
- Data Quality improvement projects often a good source of quick wins as benefits easier to quantify
- Create a use case for promoting the value of the role
- Build a foundation for longer term success

HOW?

- Prioritise Issues List & agree priorities with the data stakeholders
- Specify any support (financial and / or resources) required
- Create & lead a team to deliver the improvements

WHEN?

- Don't spend too long doing analysis and reflection
- Start a pilot / PoC DQ project as the top priority
- Use the pilot / PoC to introduce DQ principles & practices
- Apply to further pilots / PoCs

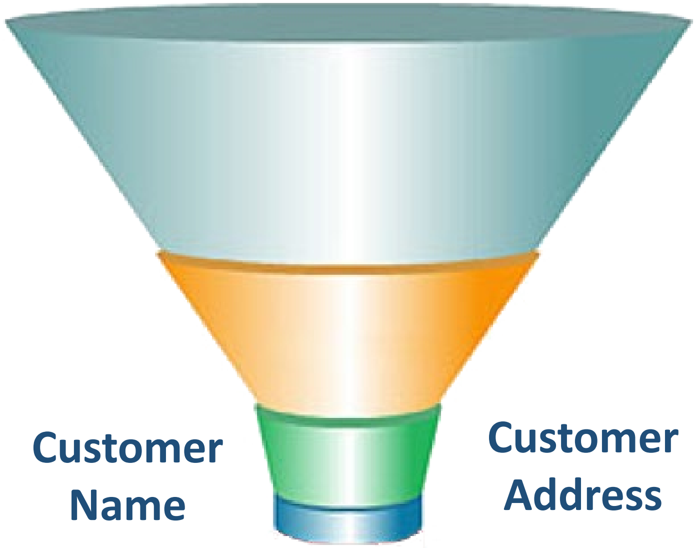
Data Management & Measures: Key Data Identification (1)

Focus on high priority data items and attributes

Identify Key Business Drivers

e.g. Launch a New Service – Marketing Campaigns require better customer information

Filter Data Elements Aligned with Business Drivers



Customer Name

Customer Address

Focus Improvement Efforts on Key Data

Customer email Address

Date of Birth

UK Region Code

Targeted Projects in to Deliver Priority Improvements

Use case: Consumer Energy Company

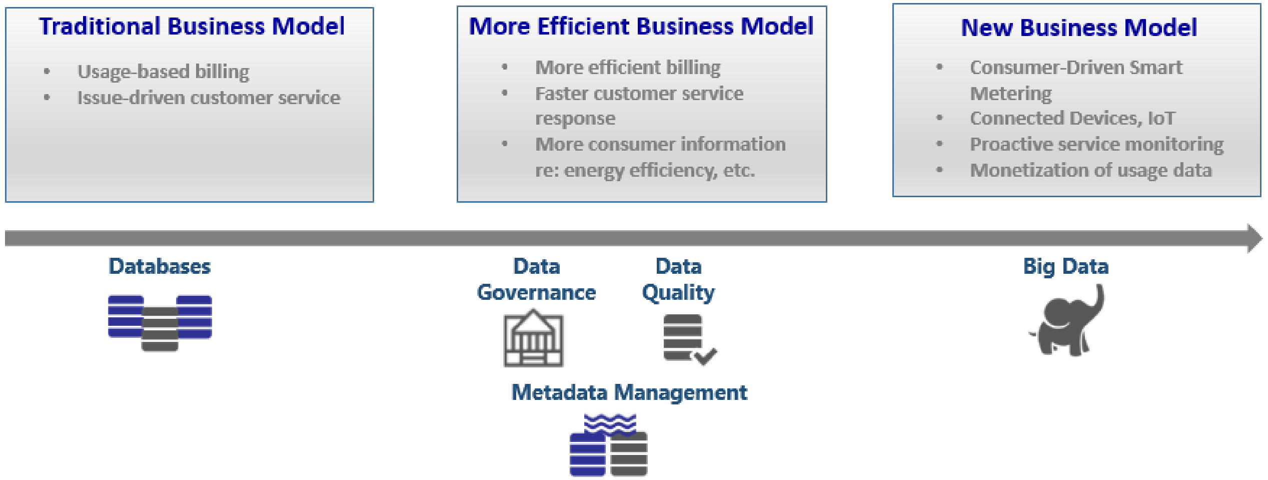
Business Transformation through Quality Data

Objectives

- For the consumer energy sector *Big Data and Smart Meters are transforming the ways of doing business* and interacting with customers.
  - Moving away from traditional data use cases of metering & billing.
  - Smart meters allow customers to be in control of their energy usage.
    - Control over energy usage with connected systems
    - Custom Energy Reports & Usage
    - Smart Billing based on usage times
- As energy usage declines, *data is becoming the true business asset* for this energy company.
  - Monetisation of non-personal data is a future consideration.
- While the Big Data Opportunity is crucial, equally important are the traditional data sources
  - Data Quality critical for operational and data warehouse data
  - Data Governance critical for analysing data in relation to business processes & roles
  - With high volumes of data, critical data elements needed to be prioritised

Result

- Data Governance in Place for Critical Data
  - Business-critical data elements identified
  - Definitions created
  - Data Governance Program analysing data in relation to business processes & roles
- Tools & Technologies Implemented for Data Quality
  - New Data Quality Tools in place for operational and DW data



Traditional Business Model

- Usage-based billing
- Issue-driven customer service

More Efficient Business Model

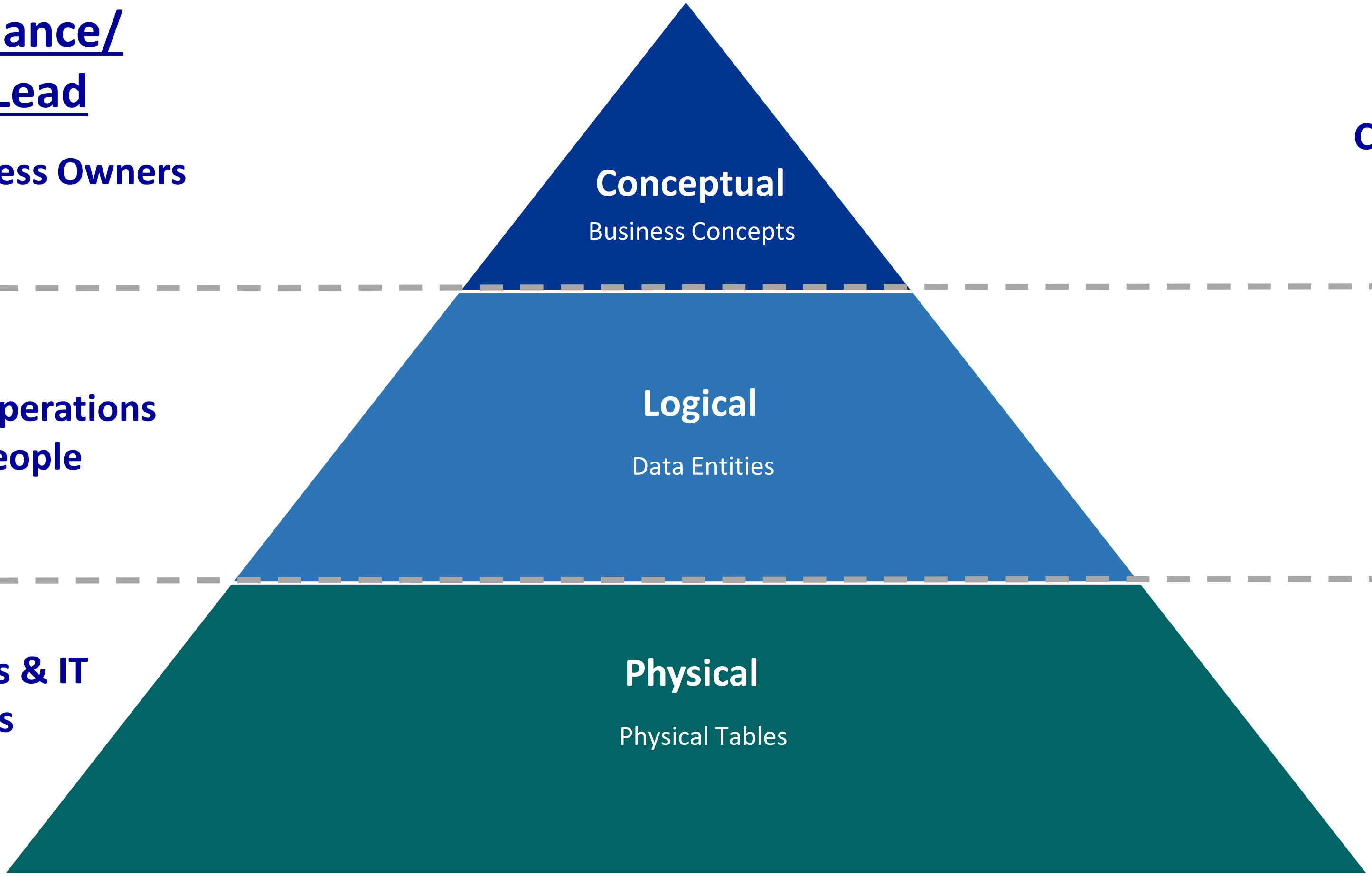
- More efficient billing
- Faster customer service response
- More consumer information re: energy efficiency, etc.

New Business Model

- Consumer-Driven Smart Metering
- Connected Devices, IoT
- Proactive service monitoring
- Monetization of usage data

Data Management & Measures: Key Data Identification (2)

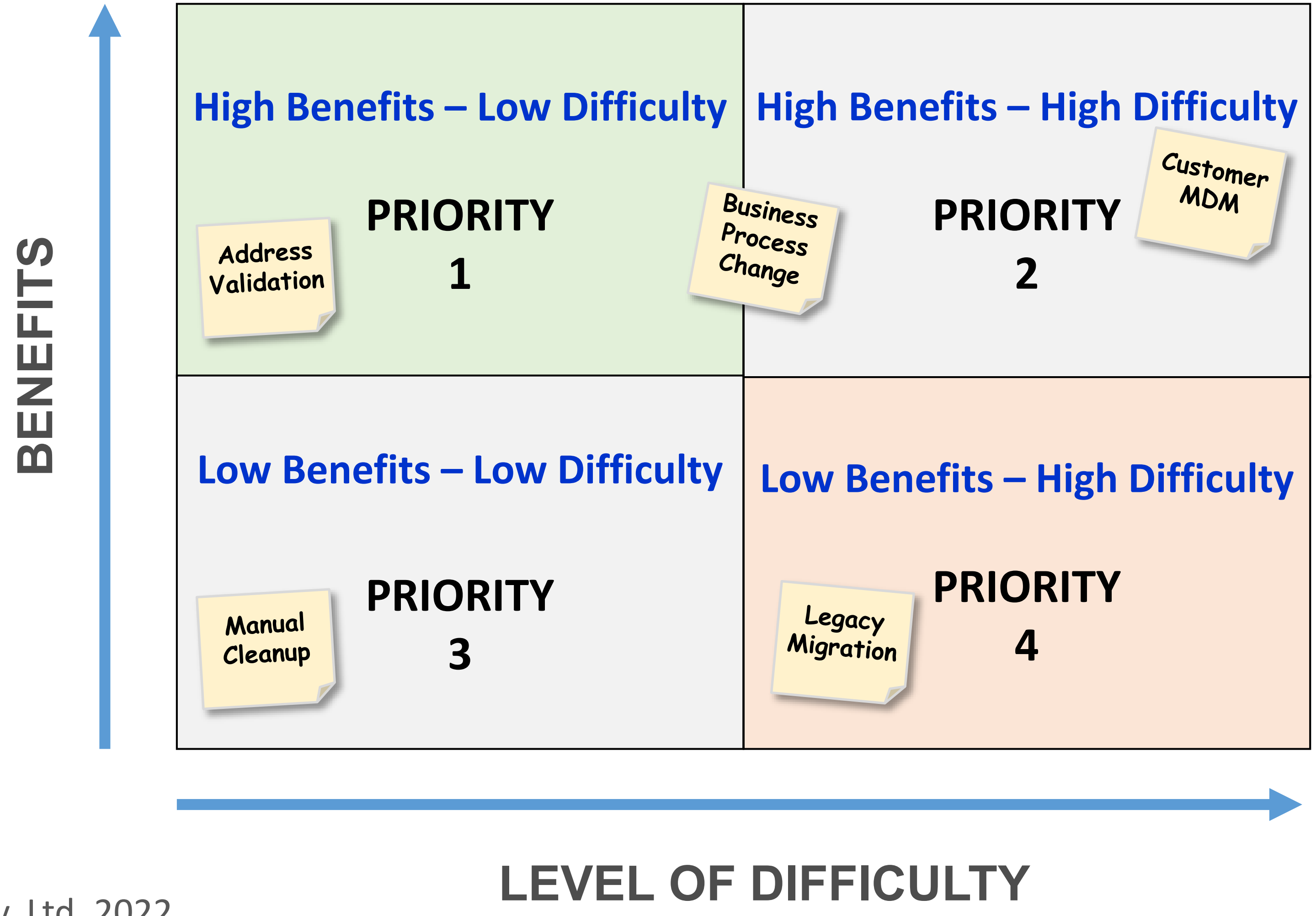
Focus on key Data Architecture entities and attributes



Governance/ SME Lead		Purpose
Senior Business Owners	**Conceptual** Business Concepts	**Communication & Definition** of Business Terms & Rules
Business Operations & IT People	**Logical** Data Entities	**Clarification & Detail** of Business Rules & Data Structures
Business SMEs & IT Specialists	**Physical** Physical Tables	**Technical Implementation** on a Physical Database

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.



High Benefits – Low Difficulty

PRIORITY 1

Address Validation

Business Process Change

High Benefits – High Difficulty

PRIORITY 2

Customer MDM

Low Benefits – Low Difficulty

PRIORITY 3

Manual Cleanup

Low Benefits – High Difficulty

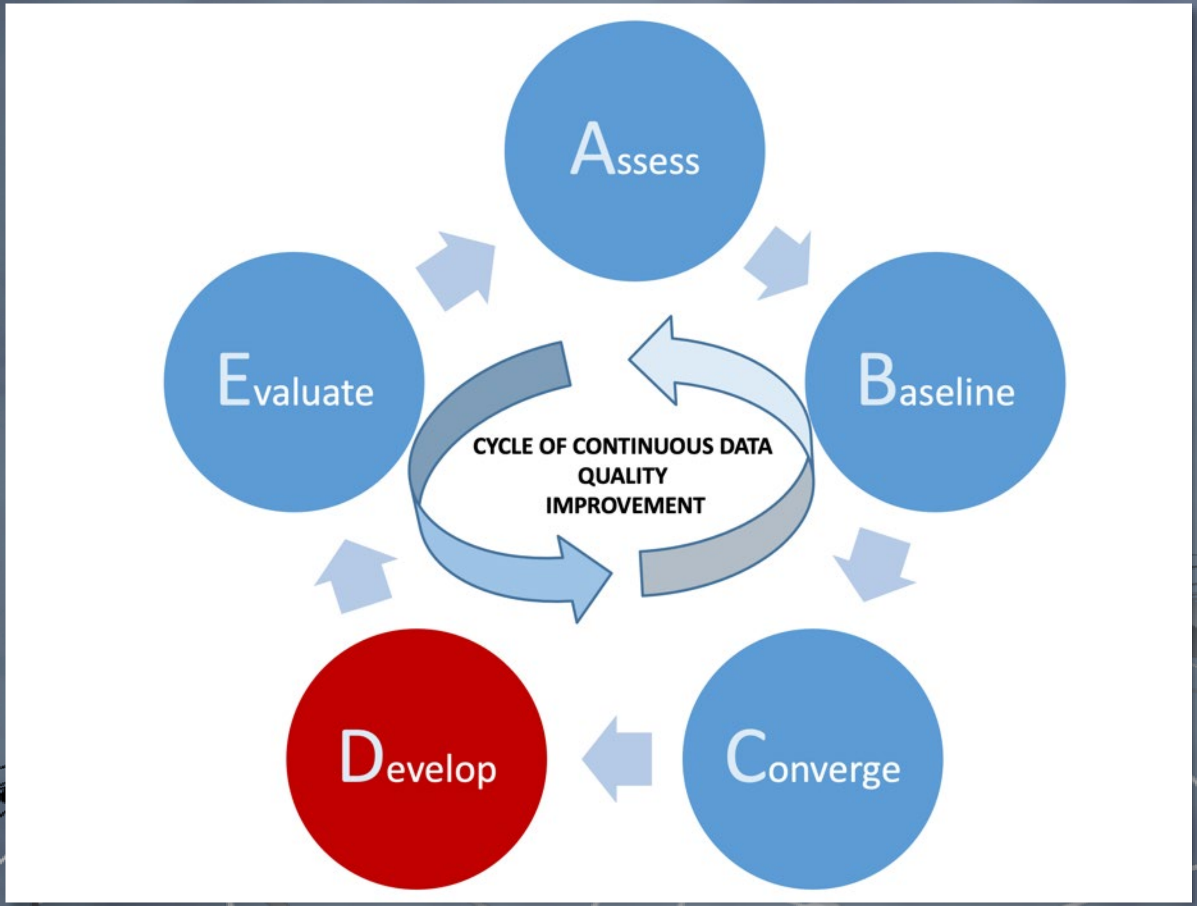
PRIORITY 4

Legacy Migration

AdeptEvents

Data Issues & Opportunities Log: Suggested Template

ID	Short Name	Brief Description	Impact of the Problem / Potential Opportunity (Business & IT)	Boston Grid Priority Score (*) 1 – High Benefits / Low Difficulty 2 – High Benefits / High Difficulty 3 – Low Benefits / Low Difficulty 4 – Low Benefits / High Difficulty
1				
2				
3				



The A2E Approach: Step 4 – Develop

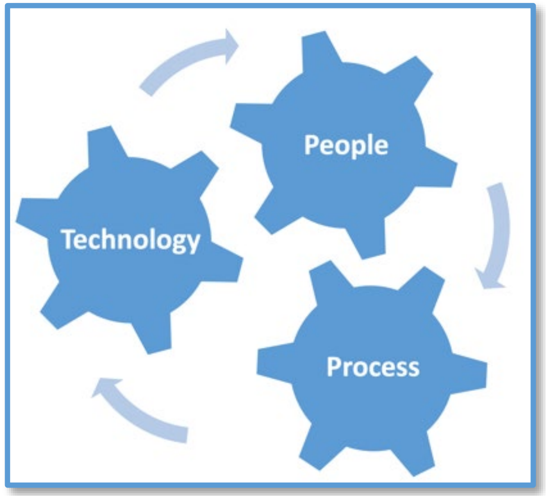
A2E Step 4: Develop

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-analyse 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

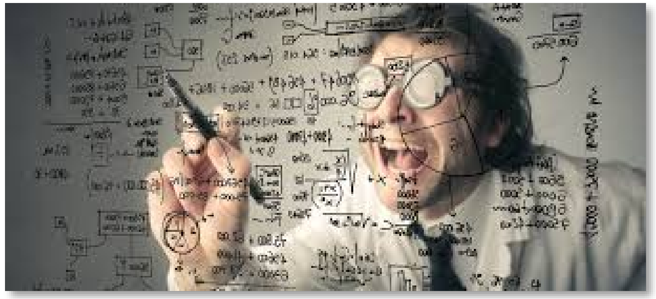
- Root Cause Analysis diagrams
- Updated business cases & case study
- Data Quality KPIs and thresholds based on the 7 Data Quality Dimensions and identified Business Rules
- Data Improvement Plans



Creating Data Quality Working Groups: Getting Wider Engagement & Collaboration

People Who Feel the Pain of Poor Data

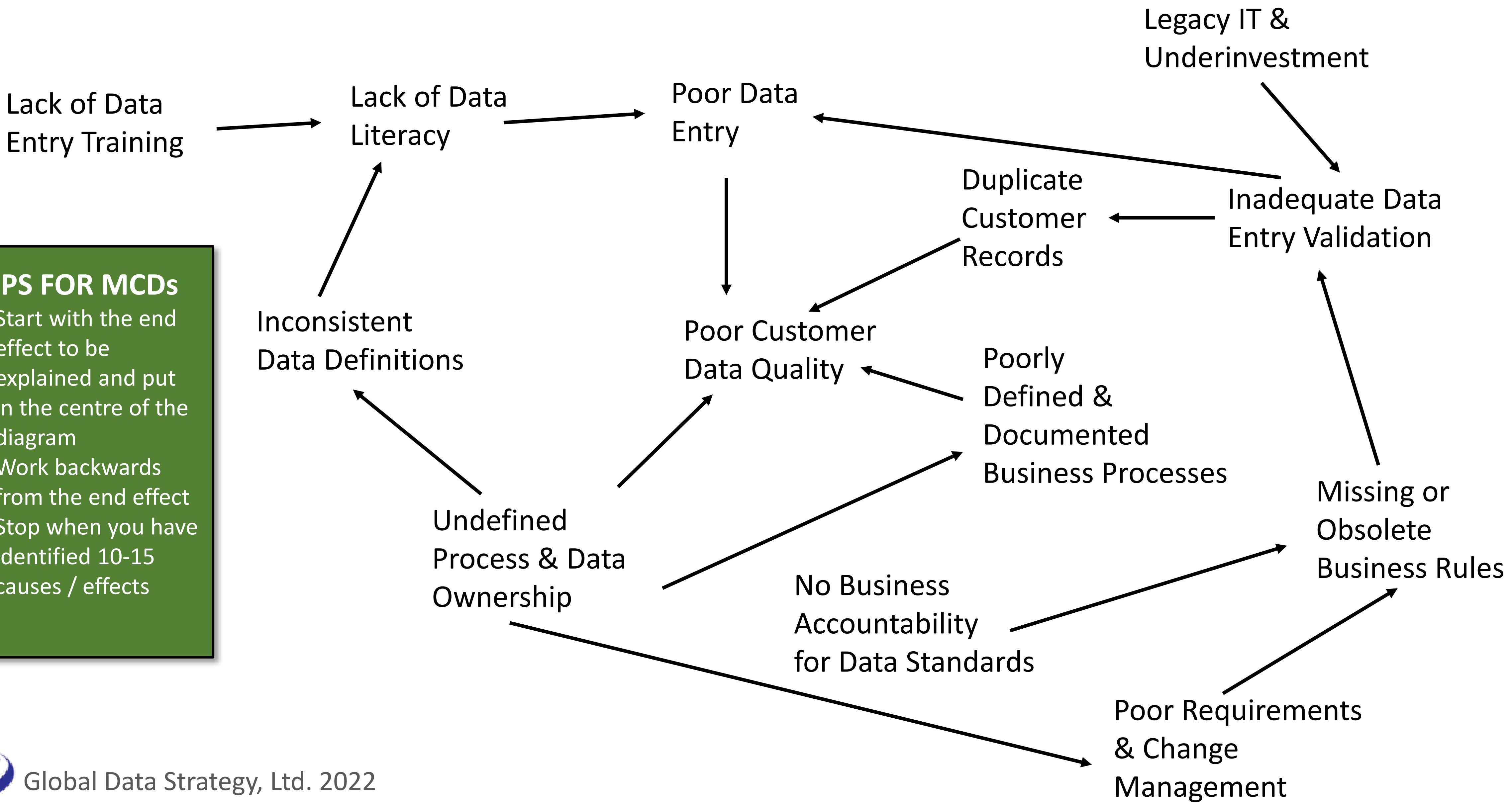
Enthusiasts Seeking a New Challenge



Data Domain Experts / Geeks
(The ‘Go To’ person)

Willing Volunteers

Multiple Cause Diagrams (Root Cause Analysis) – Example



Lack of Data Entry Training

Lack of Data Literacy

Poor Data Entry

Legacy IT & Underinvestment

Duplicate Customer Records

Inadequate Data Entry Validation

Inconsistent Data Definitions

Poor Customer Data Quality

Poorly Defined & Documented Business Processes

Missing or Obsolete Business Rules

Undefined Process & Data Ownership

No Business Accountability for Data Standards

Poor Requirements & Change Management

TIPS FOR MCDs

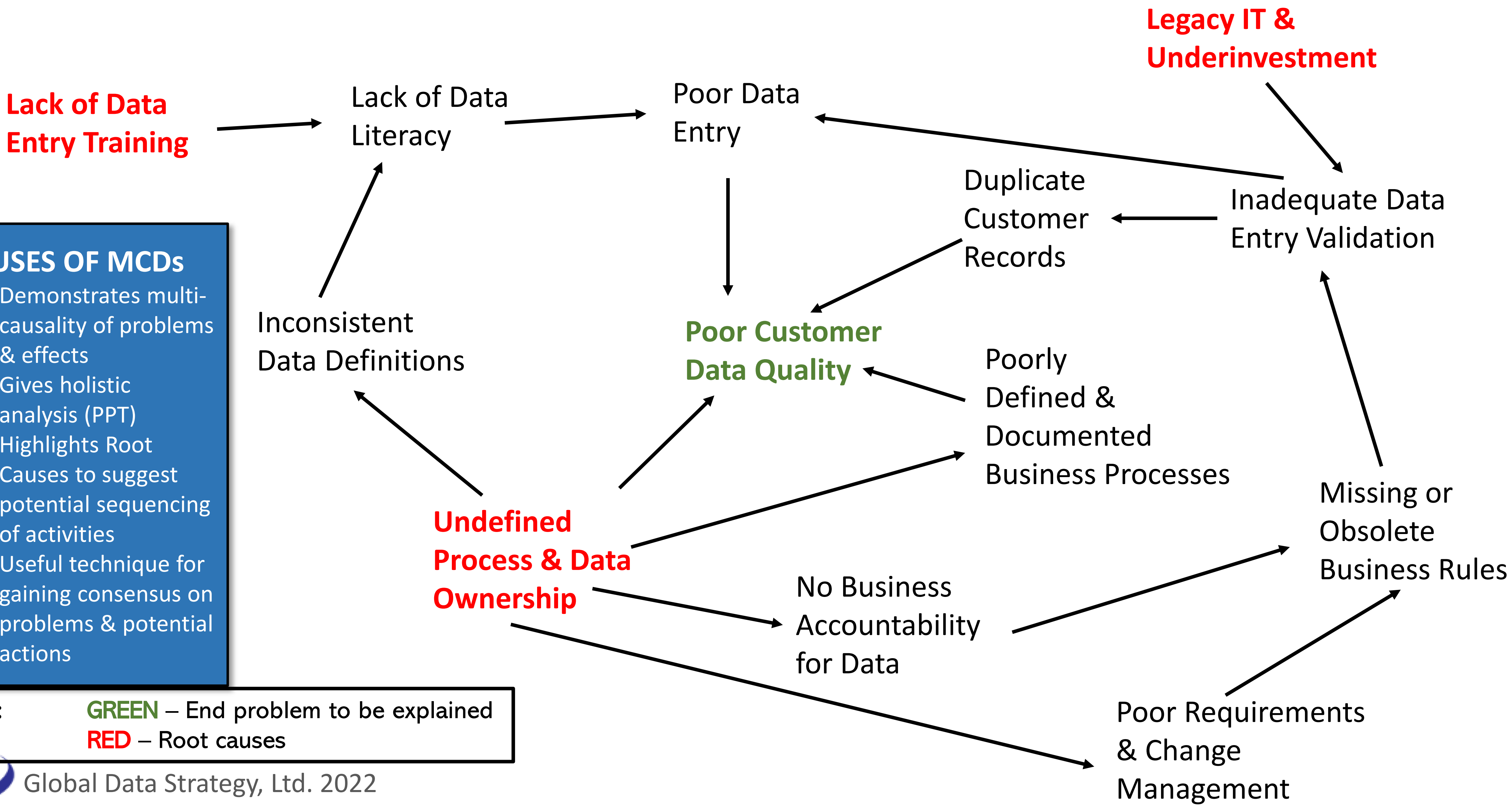
- Start with the end effect to be explained and put in the centre of the diagram
- Work backwards from the end effect
- Stop when you have identified 10-15 causes / effects

Multiple Cause Diagrams (Root Cause Analysis) – Value

USES OF MCDs

- Demonstrates multi-causality of problems & effects
- Gives holistic analysis (PPT)
- Highlights Root Causes to suggest potential sequencing of activities
- Useful technique for gaining consensus on problems & potential actions

Key: GREEN – End problem to be explained
RED – Root causes



Legacy IT & Underinvestment

Lack of Data Entry Training

Lack of Data Literacy

Poor Data Entry

Duplicate Customer Records

Inadequate Data Entry Validation

Inconsistent Data Definitions

Poor Customer Data Quality

Poorly Defined & Documented Business Processes

Missing or Obsolete Business Rules

Undefined Process & Data Ownership

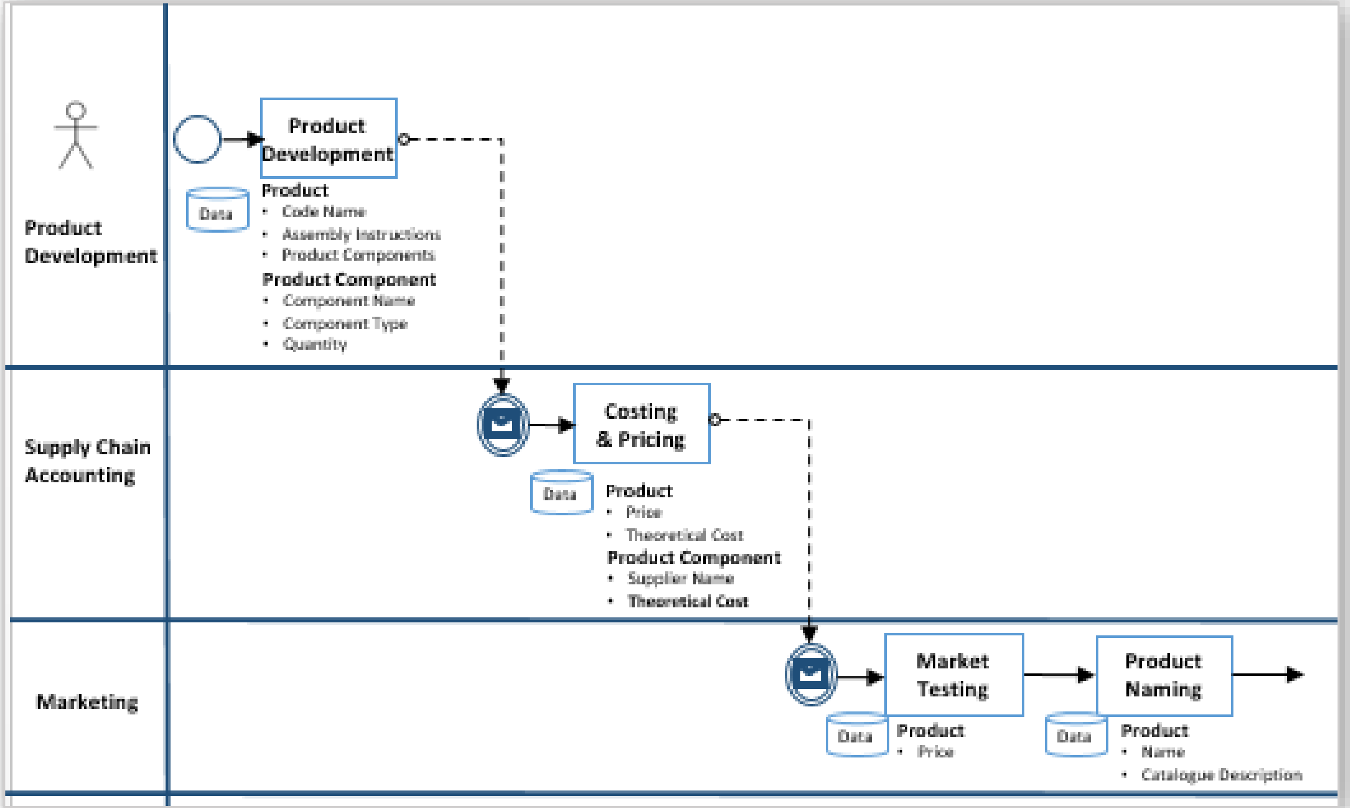
No Business Accountability for Data

Poor Requirements & Change Management

Process Models

Identifying key data dependencies in core business processes

- Process models are a helpful tool for describing core business processes.
  - “Swimlanes” outline organizational considerations
  - Data can be mapped to key business processes to understand creation & usage of information.



Use Case: BT’s Enterprise Information Programme

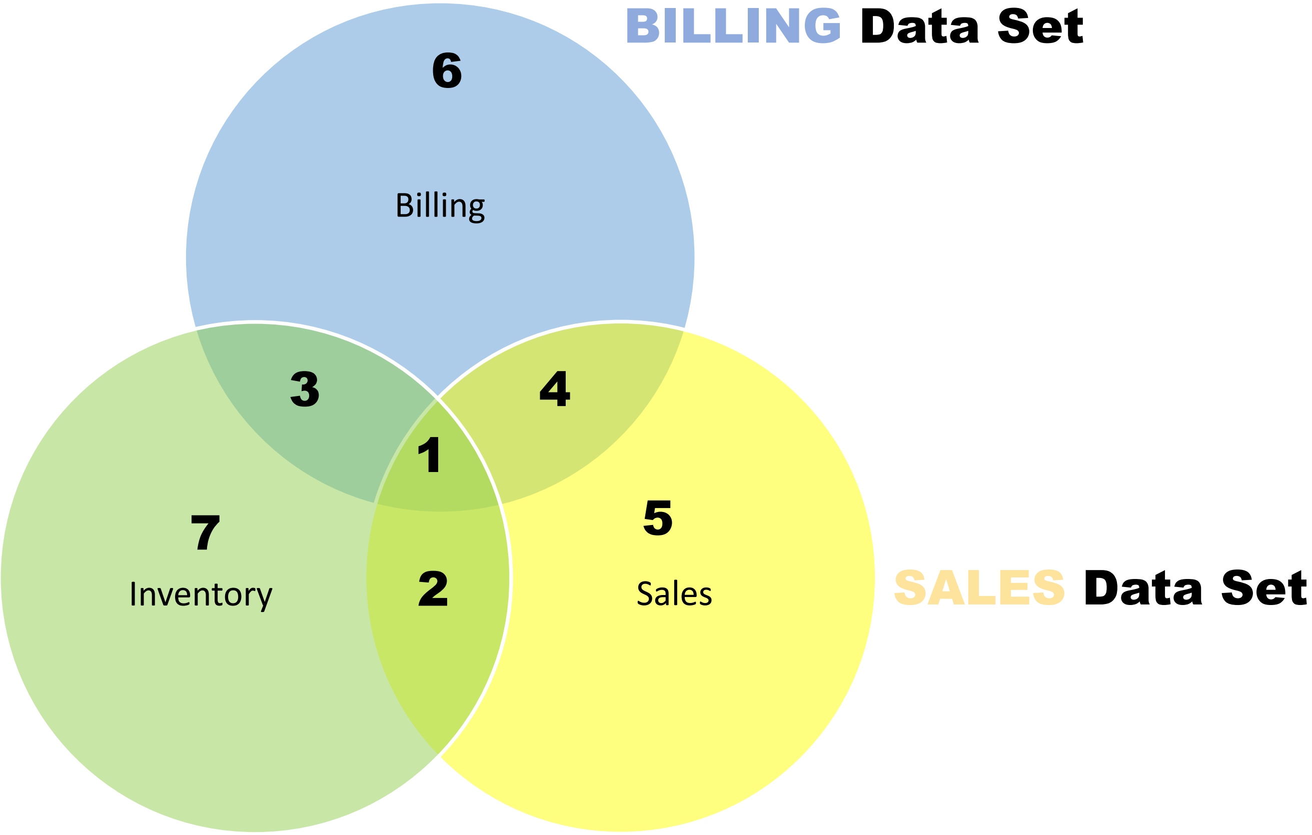
- 10 year data quality improvement programme
- Encompassed over 75 separate data quality improvement projects ranging from tactical data cleanses to master data management
- Focus on a number of data domains including:
    - Customer
    - Product
    - Billing
    - Finance
    - Personnel
    - Network Inventory

Network Inventory Improvement in BT Wholesale

- Sub-programme of overall Enterprise Information Programme
- BT Wholesale identified many problems with mismatches in billing and physical / logical inventory
- Focused on a range of network inventory areas:
  - PSTN
  - Private circuits
  - Broadband
  - Virtual private networks
  - etc.

One Technique – Matching Source Data in BT Using Venn Concepts

Data Quality Problem:
Suspected that Sales, Inventory & Billing Data Sets used for private circuits were not consistent



Approach:
Used Data Profiling tool to analyse the three key sources in parallel

1. Matched on all 3 data sources – no action
2. In Sales & Inventory, but not billed – investigation needed (revenue leakage?)
3. In Billing & Inventory, but no Sales record – investigation needed (omission?)
4. In Sales & Billing, but no Inventory record – investigation needed (omission?)
5. In Sales only – investigation needed (cancelled sale not cleared?)
6. In Billing only – investigation needed (Customer billed in error?)
7. In Inventory only – investigation needed (obsolete inventory record?)

The Importance of KPIs

“You Can’t Manage What You Can’t Measure”

- Most businesses set strategic goals they desire to achieve and measure these goals against Key Performance Indicators (KPIs).
  - These KPIs provide a concrete, objective way to measure progress towards these goals
- To use Finance as a comparison, they have a number of KPIs they use to manage **financial assets**
  - Revenue Projections
  - Budget Goals & Limits
  - Expense Ration, etc.

- We need to do the same with **data assets**
  - % complete
  - % accuracy
  - Timeliness
  - RoI
  - Cost Savings

Measuring Data Improvements

Align Data Quality Metrics to Business Improvements

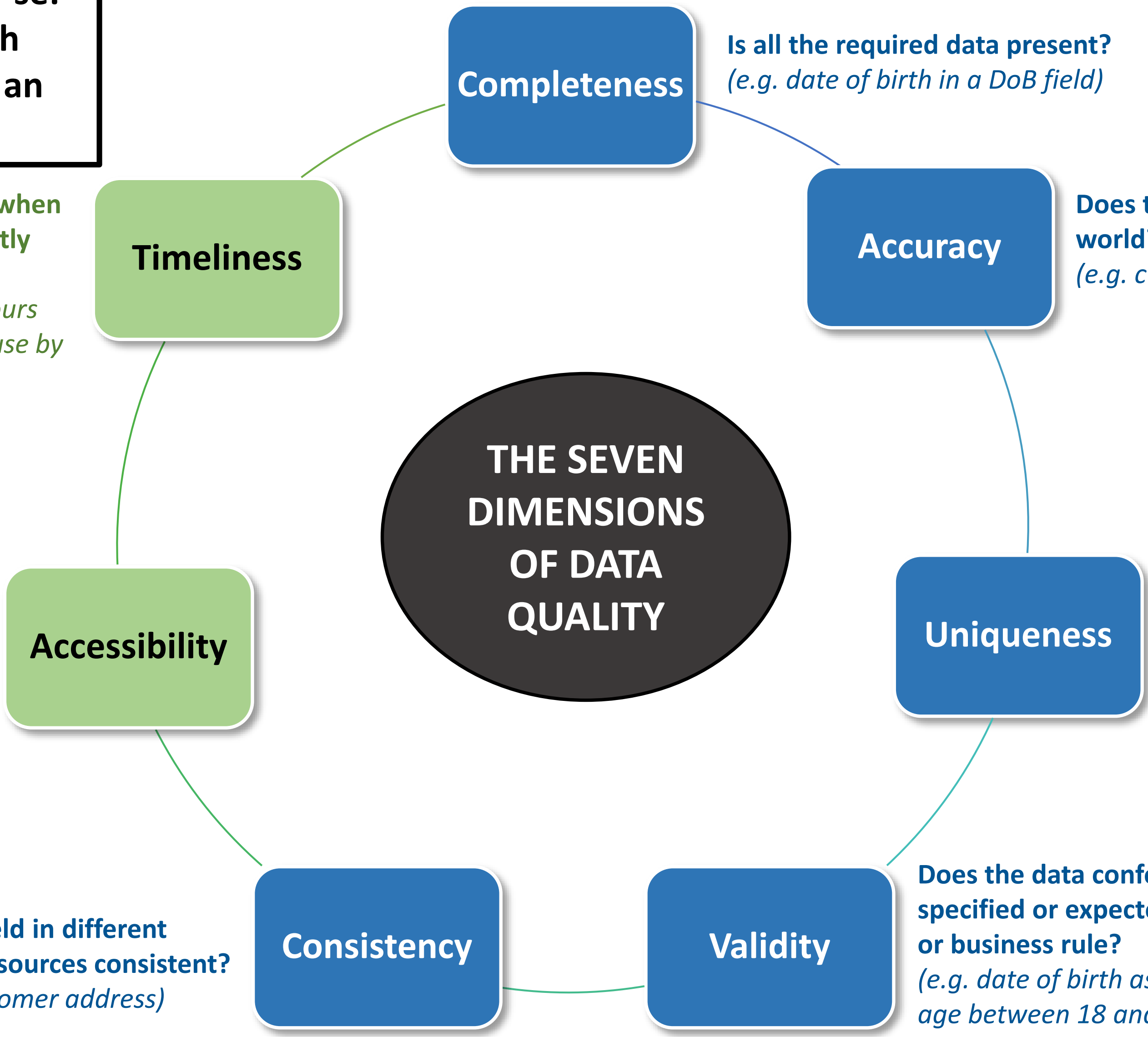
- KPIs & Measures aligned with concrete business drivers
  - Helps prioritise efforts
  - Assists with the “Why do I Care?” issue
  - Basis for showing benefits and results

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	🟥	• Correct # of customers for sales estimations • Better single view of customer for integrated social media campaign • Reduce cost of physical mailing by £20K	• Cost savings • Brand Reputation • Marketing Innovation
Incorrect Salutation (Mr, Ms, etc.)	5,000	1,000	🟨	• Customer satisfaction & Brand reputation harmed by incorrect salutation • Targeted marketing campaigns by gender	• Brand Reputation • Campaign Effectiveness
Incorrect address/location	10,000	500	🟨	• Lower return rate on physical mailings • Better targeted marketing by region.	• Cost Savings • Campaign Effectiveness
Missing Sales Rep Assigned	500	100	🟩	• Ability for Sales to execute on customer leads • Revenue growth	• Sales Effectiveness

Baselining & Setting KPIs: the 7 Dimensions of Data Quality

You cannot measure DQ per se: you need to measure each dimension and roll up into an overall DQ measure



THE SEVEN DIMENSIONS OF DATA QUALITY

Completeness
Is all the required data present?
(e.g. date of birth in a DoB field)

Accuracy
Does the data reflect the real world?
(e.g. current customer address)

Uniqueness
In a data source, is the entry unique or are there unintended duplicate records?
(e.g. same client organisation spelled several different ways in multiple CRM records)

Validity
Does the data conform to a specified or expected format and / or business rule?
(e.g. date of birth as DD/MM/YYYY; age between 18 and 120 years)

Consistency
Where data is held in different sources, are the sources consistent?
(e.g. current customer address)

Accessibility
Do the users who need to use the data have access to it?
(e.g. Finance team and invoice data held in data warehouse)

Timeliness
Is the data available to users when they need it and is it sufficiently timely to meet their needs?
(e.g. invoices sent in last 24 hours available on the data warehouse by 9am the next day)

Key:

🟦 CONTENT DIMENSIONS

🟩 CONTEXT DIMENSIONS

Data Improvement: The Importance of Business Rules

”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

- In a data context, business rules are used to define and enforce the standards that data must conform to
- Therefore 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

Example Data Related Business Rules

FORMAT RULES

- A UK National Insurance Number must be in the format: aa nn nn nn a
- An employee must have a unique Employee ID in the format: aa nnnn
- Date of birth should be in North American format of MM/DD/YYYY
- A full US zip code must be in the format nnnnn-nnnn
- Internet router identifier must be in the format Aaa_Nan_Naa

Example Data Related Business Rules

CONTENT RULES

- Every Sales Representative must be assigned to one and only one Sales Region
- A valid email address must be entered by a customer to enable a customer's order to be accepted
- Gender codes must have the valid value of Male, Female or Unknown
- A supplier must have at least one associated geographical address
- Product Price should be Product Unit Cost + 25%

How Do You Identify Business Rules?

- Business rules can be discovered or derived from:
  - Data models (Business / Logical / Physical)
  - Business documentation (e.g. Process Descriptions, User Instructions)
  - IT Documentation (e.g. requirements specifications, system manuals)
  - Source code (e.g. If 'A Then B' statements)
  - Master and / or Reference Data Sources (e.g. currency codes, product master data)
  - Documented metadata (e.g. Business Glossaries, Data Dictionaries, Metadata Repositories)
  - Data profiling outputs
  - **Talking to key stakeholders**:
    - Data owners and data stewards (if in place)
    - Data producers and consumers
    - Other business and IT subject matter experts

VITAL IMPORTANCE OF STAKEHOLDER ENGAGEMENT:

- Business rules are frequently implicit (i.e. locked in people's heads) and not formally documented
- Where business rules are documented, documentation is often out of date and not updated in line with system changes

ACTIVITY

Referring back to the HR data source sample earlier, specify:

1. 3 Format Business Rules
2. 3 Content Business Rules

that can be used to create and enforce ‘fit for purpose’ data improvement on this data source

FIT FOR PURPOSE

ACTIVITY – Extract of actual HR data source

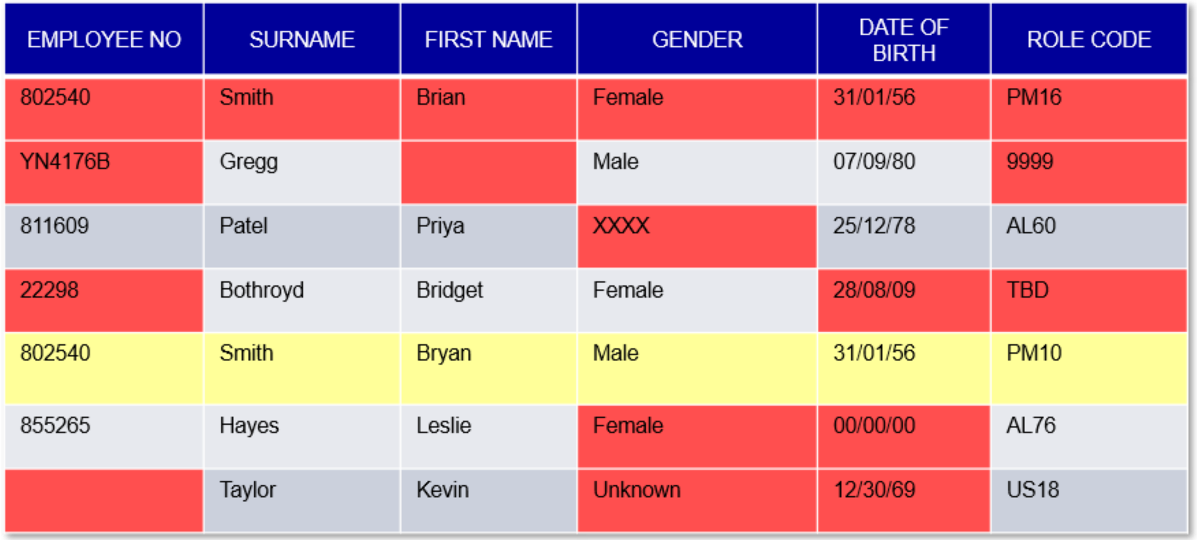
EMPLOYEE NO	SURNAME	FIRST NAME	GENDER	DATE OF BIRTH	ROLE CODE
802540	Smith	Brian	Female	31/01/56	PM16
YN4176B	Gregg		Male	07/09/80	9999
811609	Patel	Priya	XXXX	25/12/78	AL60
22298	Bothroyd	Bridget	Female	28/08/09	TBD
802540	Smith	Bryan	Male	31/01/56	PM10
855265	Hayes	Leslie	Female	00/00/00	AL76
	Taylor	Kevin	Unknown	12/30/69	US18

Key:

🟥	Potential Data Quality Problem
🟨	Potential Duplicate Record

ANSWER: Total number of potential Data Quality problems is 13 or 19, depending on whether Smith is a duplicate record

Using Business Rules to steer and enforce Data Quality standards



EMPLOYEE NO	SURNAME	FIRST NAME	GENDER	DATE OF BIRTH	ROLE CODE
802540	Smith	Brian	Female	31/01/56	PM16
YN4176B	Gregg		Male	07/09/80	9999
811609	Patel	Priya	XXXX	25/12/78	AL60
22298	Bothroyd	Bridget	Female	28/08/09	TBD
802540	Smith	Bryan	Male	31/01/56	PM10
855265	Hayes	Leslie	Female	00/00/00	AL76
	Taylor	Kevin	Unknown	12/30/69	US18

Example potential format business rules	Example potential content business rules
Employee No. must be in format nnnnnn. Blank Employee Numbers are only allowed if new starter is awaiting Emp. No. allocation	Employee No. should be unique. Only one Employee No. should be allocated to any individual employee
First Name must not be blank	Gender should align with First Name derived from Common Names Reference file
Date of Birth must be in format nn/nn/nn	Allowable Genders are FEMALE, MALE, SELF-DETERMINED or UNKNOWN
Role code must be in format AAnn	Date of Birth must be expressed as DD/MM/YY and in the range 01/01/1940 to 12/12/2006

Deploying Business Rules - Approaches

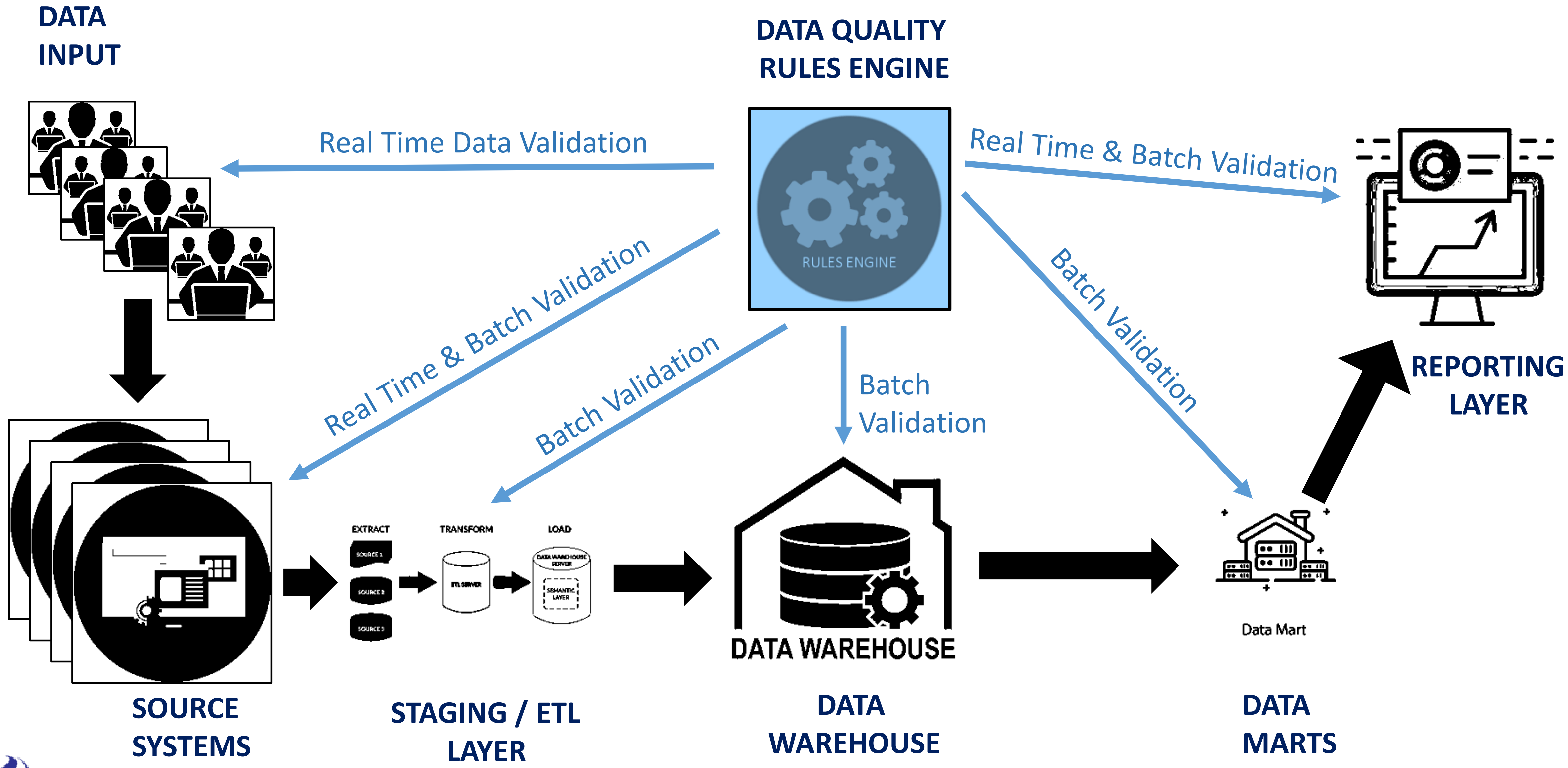
Data Entry Guidelines, Business Glossary & Training

Master & Reference Data Management

Application Code (e.g. data input validation)

Data Quality Tool: DQ Business Rules Engine

Step 3: Automating Data Quality Business Rules via a DQ Rules Engine



DATA INPUT

DATA QUALITY RULES ENGINE

Real Time Data Validation

Real Time & Batch Validation

Real Time & Batch Validation

Batch Validation

Batch Validation

Batch Validation

REPORTING LAYER

SOURCE SYSTEMS

STAGING / ETL LAYER

DATA WAREHOUSE

DATA MARTS

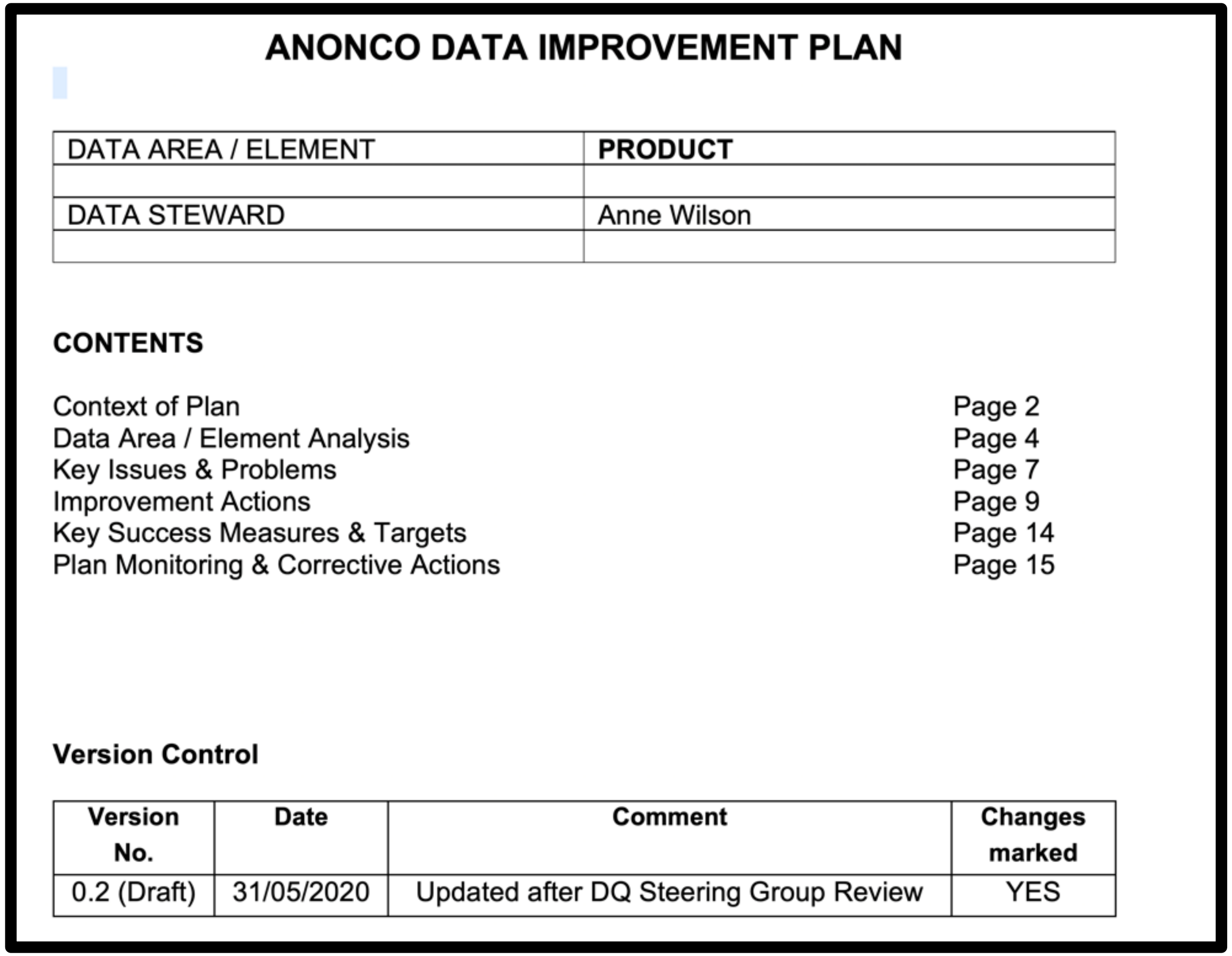
Data Improvement Plan

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
- Prioritises improvement activities
- Can be used to track improvements and communicate successes
- Can evolve to align with the changing needs of the business

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



ANONCO DATA IMPROVEMENT PLAN

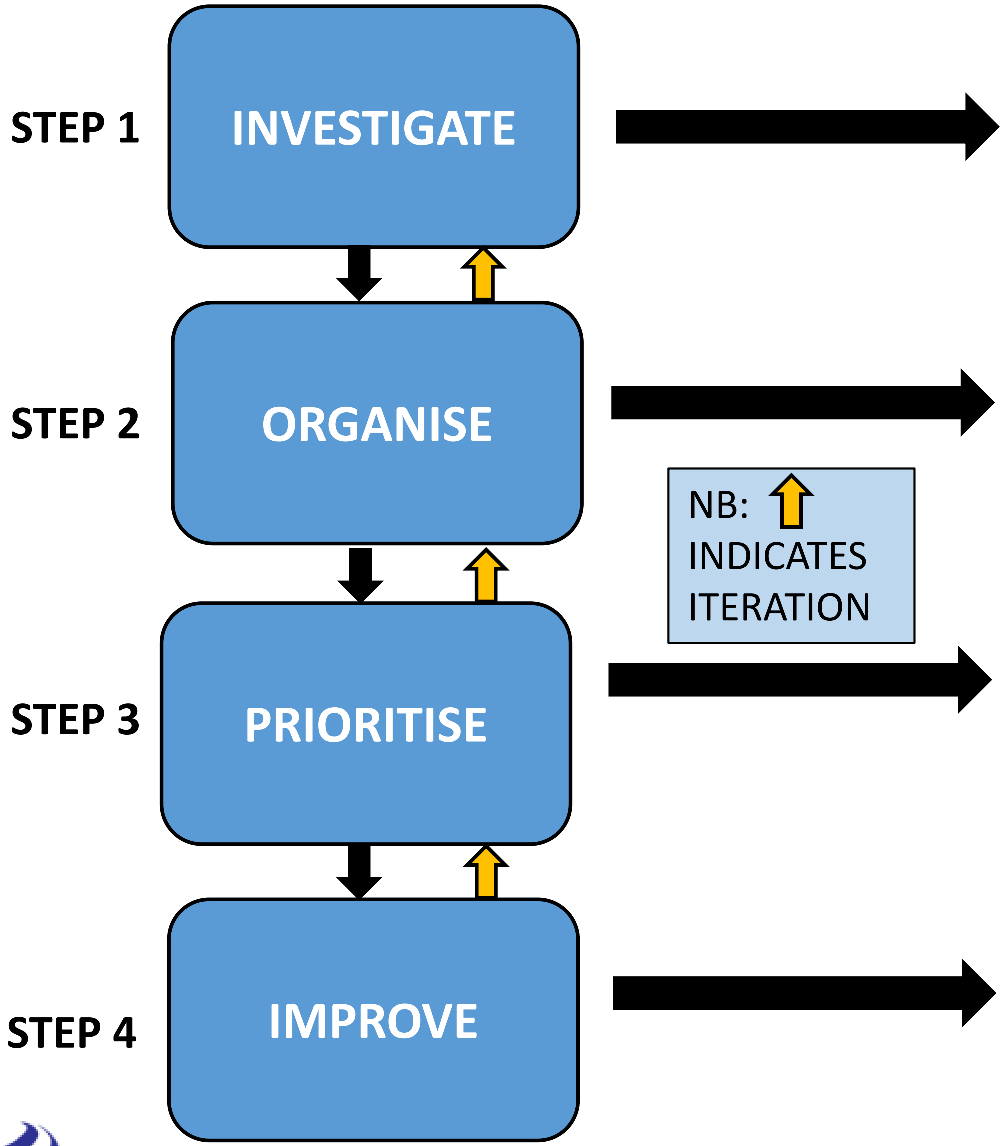
DATA AREA / ELEMENT	**PRODUCT**
DATA STEWARD	Anne Wilson

CONTENTS

Version Control

Version No.	Date	Comment	Changes marked
0.2 (Draft)	31/05/2020	Updated after DQ Steering Group Review	YES

Creating a DIP for a Data Domain or Problem Area – 4 Steps



STEP 1 INVESTIGATE

- Define data domain and its uses (Processes and Functions)
- Identify data stakeholders (Creators, Modifiers, Consumers)
- Engage with stakeholders (e.g. interviews, workshops, documents)
- Identify key data fields (i.e. what data <u>really</u> matters)
- Baseline data (Systems & Quality) & Data Governance maturity
- Identify data problems & impact

STEP 2 ORGANISE

- Set up Data Domain Working Group (Business, IT and key stakeholders)
- Create a log of data problems, opportunities and business impact
- Initially identify and define potential improvement initiatives (People / Process / IT)

NB: ⇧ INDICATES ITERATION

STEP 3 PRIORITISE

- Prioritise data problems
- Define improvement projects
- Create improvement team(s) (from Working Group & others)
- Produce Motivation Model & business case(s) for action
- Finalise initial Data Improvement Plan for Steering Group endorsement

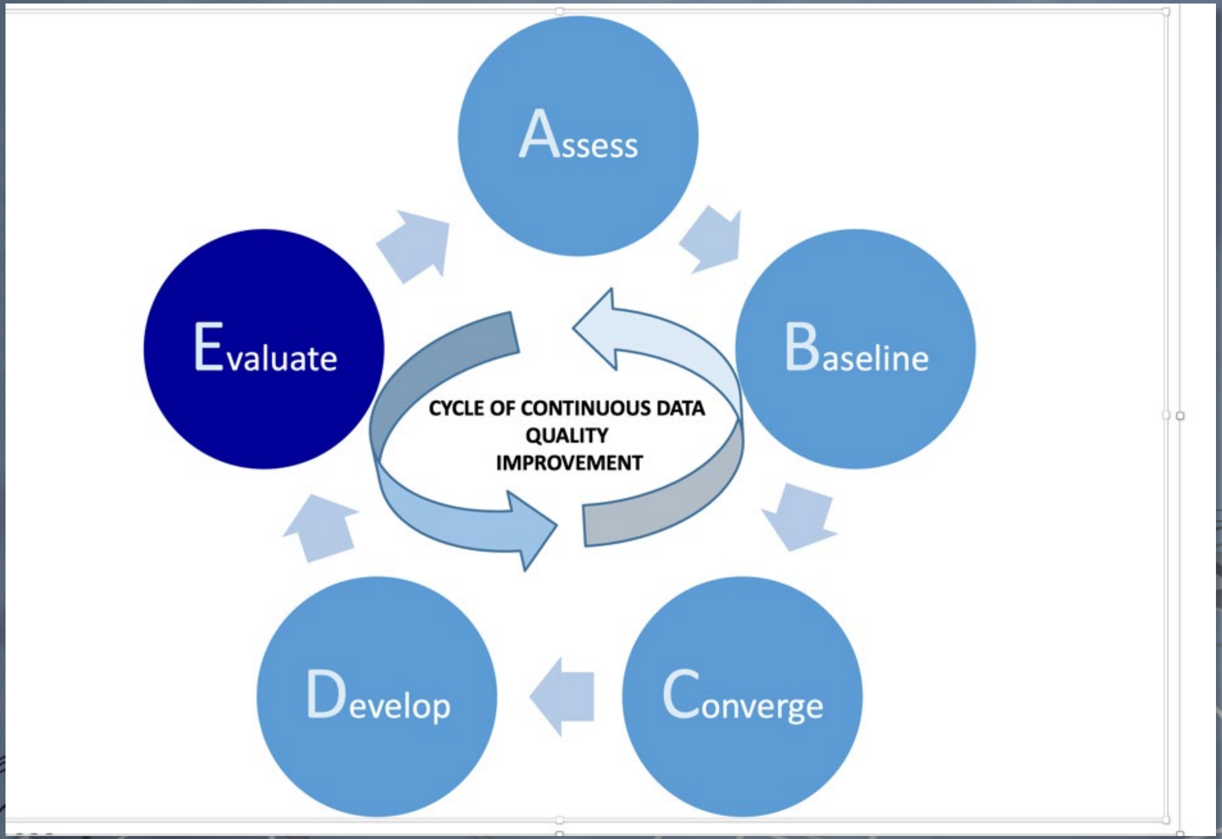
STEP 4 IMPROVE

- Launch improvement initiatives
- Set KPIs and success measures
- Perform root cause analysis and propose & evaluate changes
- Design and implement improvements (People / Process / IT)
- Produce improvement plans, monitor progress and measure data improvements
- Log benefits, publicise successes, and identify lessons learnt

DIP – Primary Roles & Responsibilities

- **DATA OWNER / SENIOR BUSINESS LEAD**
  - Is formally responsible for the Data Domain DIP
  - Champions its value and importance across the business
  - Acquires funding, resourcing and support to ensure the DIP is delivered
  - Ensures that the DIP evolves in line with changing business goals and shifting priorities
  - Supports the Business Data Steward in leading the DIP
- **BUSINESS DATA STEWARD / BUSINESS DATA SUBJECT MATTER EXPERT**
  - Leads the DIP on behalf of the Data Owner
  - Liaises with data stakeholders to secure active participation in the plan
  - Creates and leads a Data Domain Working Group to deliver the plan
  - Appoints specific data improvement leads and participants
  - Reports delivery progress and escalations to the Data Owner
  - Communicates successes to the data stakeholder community and the wider business
- **BUSINESS DATA STAKEHOLDERS**
  - Represents data consumers and producers as appropriate
  - Plays an active part in creating and / or reviewing the DIP
  - Participates in improvement projects and the Data Working Group
  - Highlights new potential problems and improvements for consideration by the Data Working Group
  - Helps secure the commitment of the wider business to enable changes
- **IT / TECHNICAL DATA SUBJECT MATTER EXPERTS**
  - Appoints a DIP lead / Technical Data Steward(s) to support the Data Owner and Business Data Steward to deliver the DIP
  - Ensures technical considerations are included in the DIP
  - Provides technical help and support (e.g. data baselining, data profiling etc.) to progress DIP improvement initiatives

Roles & Responsibilities



The A2E Approach: Step 5 – Evaluate

A2E Step 5: Evaluate

EVALUATE & SUSTAIN GAINS

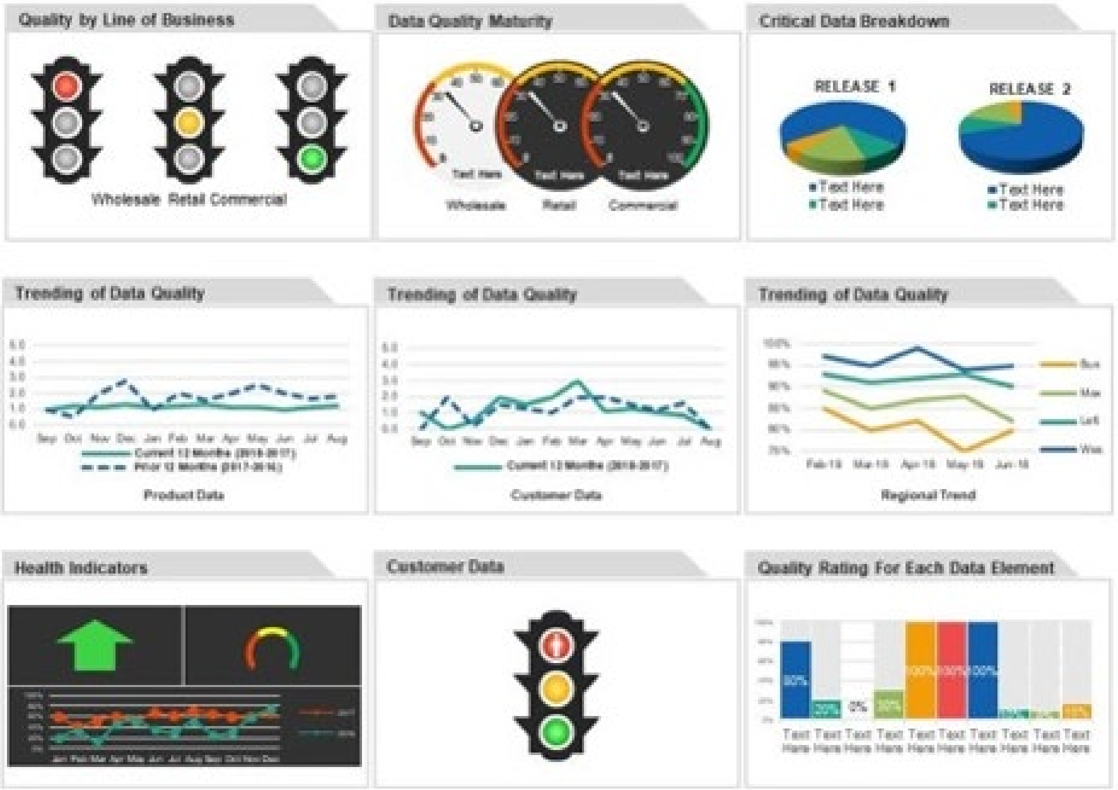
- Embed Data Quality improvement as a business as usual activity
- Evolve Data Quality improvement teams into wider Data Governance structure:
- Track Data Quality improvements via Data Quality Dashboards
- Monitor financial and business benefits over time
- Evangelising benefits – part of your job is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

Monitor & Report Business Rule Adherence

- When Business Rules are implemented can be used to:
  - Check continued adherence of existing data
  - Enforce the rules on new data to prevent new problems
- Best monitored via Data Quality Dashboards
  - Provide regular reports on adherence of data to Business Rules
  - Set KPIs to drive continuous data improvement
  - Identify data quality trends
  - Highlight areas where corrective action required
  - Indicate where / if Business Rules may need to be amended to meet changing business needs
- When reporting always try to relate data quality to business outcomes
  - Address the 'so what' objection
  - Puts a financial or other benefit on continued data quality improvement


Quality by Line of Business
Wholesale Retail Commercial
Data Quality Maturity
Text Here
Text Here
Text Here
Wholesale
Retail
Commercial
Critical Data Breakdown
RELEASE 1
RELEASE 2
Text Here
Text Here
Text Here
Text Here
Trending of Data Quality
Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug
Current 12 Months (2016-2017)
Prior 12 Months (2017-2016)
Product Data
Trending of Data Quality
Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug
Current 12 Months (2016-2017)
Customer Data
Trending of Data Quality
Feb-18 Mar-18 Apr-18 May-18 Jun-18
Regional Trend
Health Indicators
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Customer Data
Quality Rating For Each Data Element
80%
20%
0%
30%
100%
100%
100%
Text Here

Data Quality Dashboard

Data Governance & Data Quality: the synergies

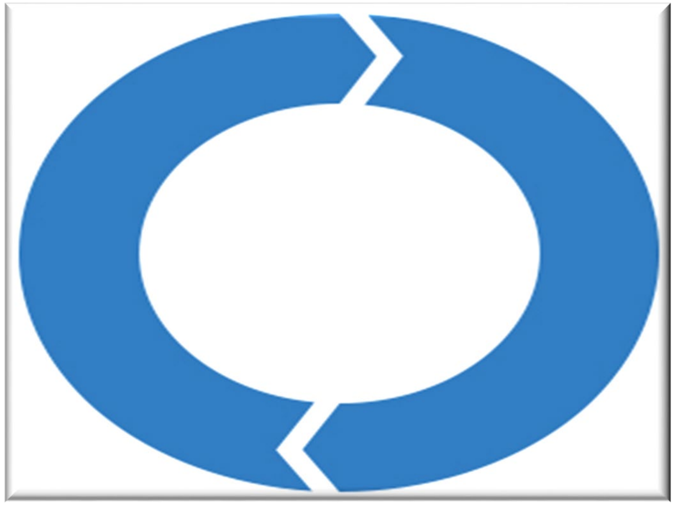
- Data quality improvement is the primary reason why most organisations implement data governance:
  - Realise that data quality is a business problem and not an IT problem
  - Recognise that data quality improvement cannot be sustained without business leadership
  - Better IT systems & tools can help but are not sufficient in themselves
  - Have learnt that data quality is NOT a synonym for data cleanse; data cleanse is a repeated cost of failure and usually does not remove the root causes of poor data quality
  - Data is volatile and so data quality has to be a perpetual business as usual activity enabled through data governance

- Better data quality is usually the most effective way of demonstrating the value of data governance:
  - Enhanced data quality can often deliver ‘quick wins’
  - The benefits of reducing the ‘costs of failure’ caused by poor data quality can be significant & measurable
  - A sound data quality foundation enhances the value & success of other data management investments (e.g. Business Intelligence, Data Science, Analytics, CRM et al)

Data Quality

Data that is demonstrably fit for business purposes

Drives the need for

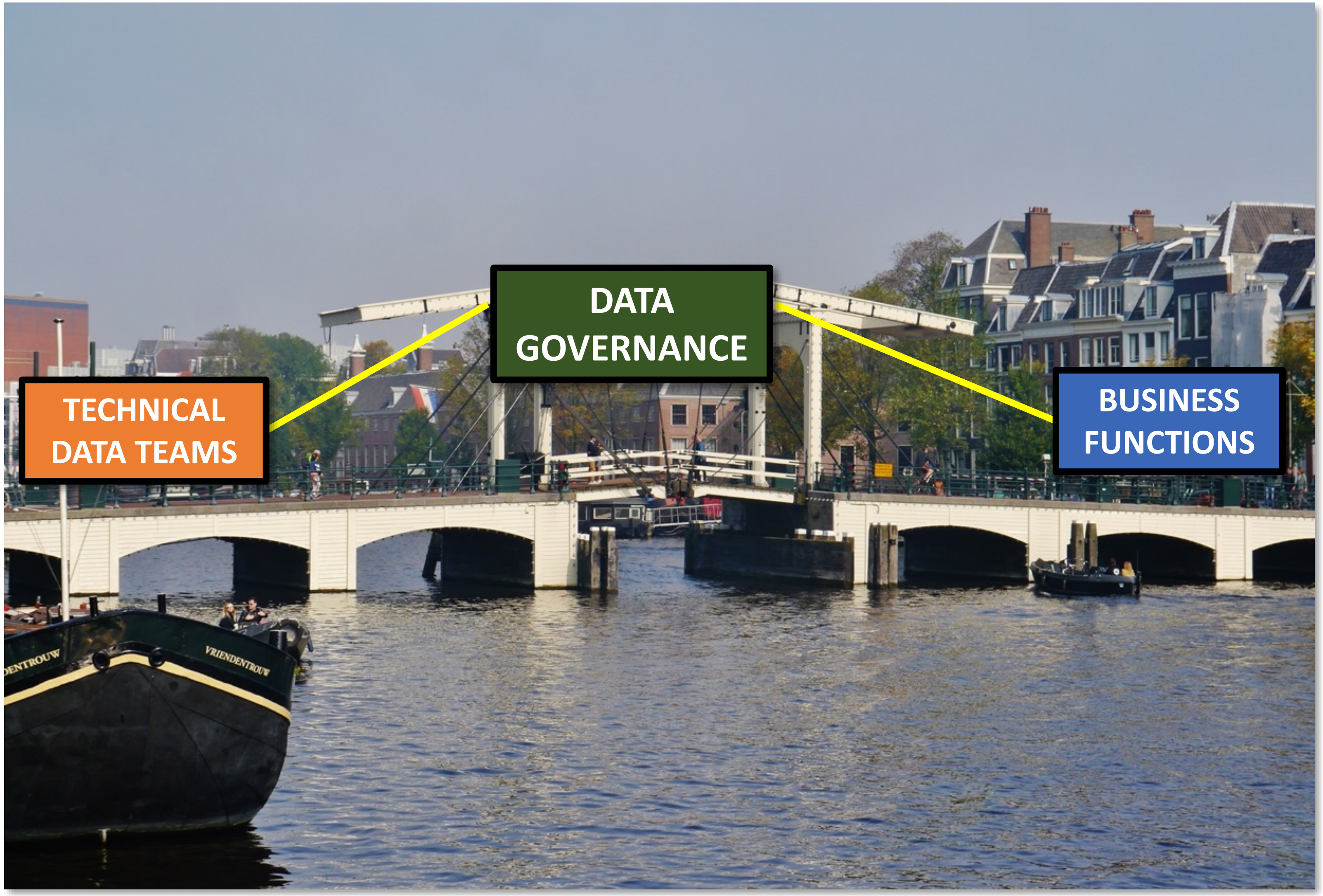


Provides the means to deliver

Data Governance

A business led continuous process to improve data for the benefit of all data stakeholders

Data Governance: The Bridge Between Business & IT



DATA GOVERNANCE

TECHNICAL DATA TEAMS

BUSINESS FUNCTIONS

Typical Key Data Governance Roles

Executive Sponsor

- Promotes Data Driven Culture
- Champions Best Practices
- Advocate with ELT and Board
- Escalation Point for Key Issues

Data Owner

- Represents the data needs for a particular functional area
- Defines key KPIs & data elements
- Defines key business rules
- Sanctions Data Quality Metrics & Thresholds

Business Data Steward

- Responsible for the day-to-day management and quality of data
- Subject Matter Expert (SME) for a given business domain
- Aligns with the Lead Data Steward to support business rules and to align with key KPIs

Technical Data Steward

- Digital/IT expert for a given business unit
- Subject matter expert for a given system and its usage
- Aligns with Business Data Stewards to ensure technical needs are met

Data Governance Lead*

- Acts as a cross-functional lead for the data governance effort, working with both business and IT roles
- Chair of the Data Governance Steering Committee

Data Architect*

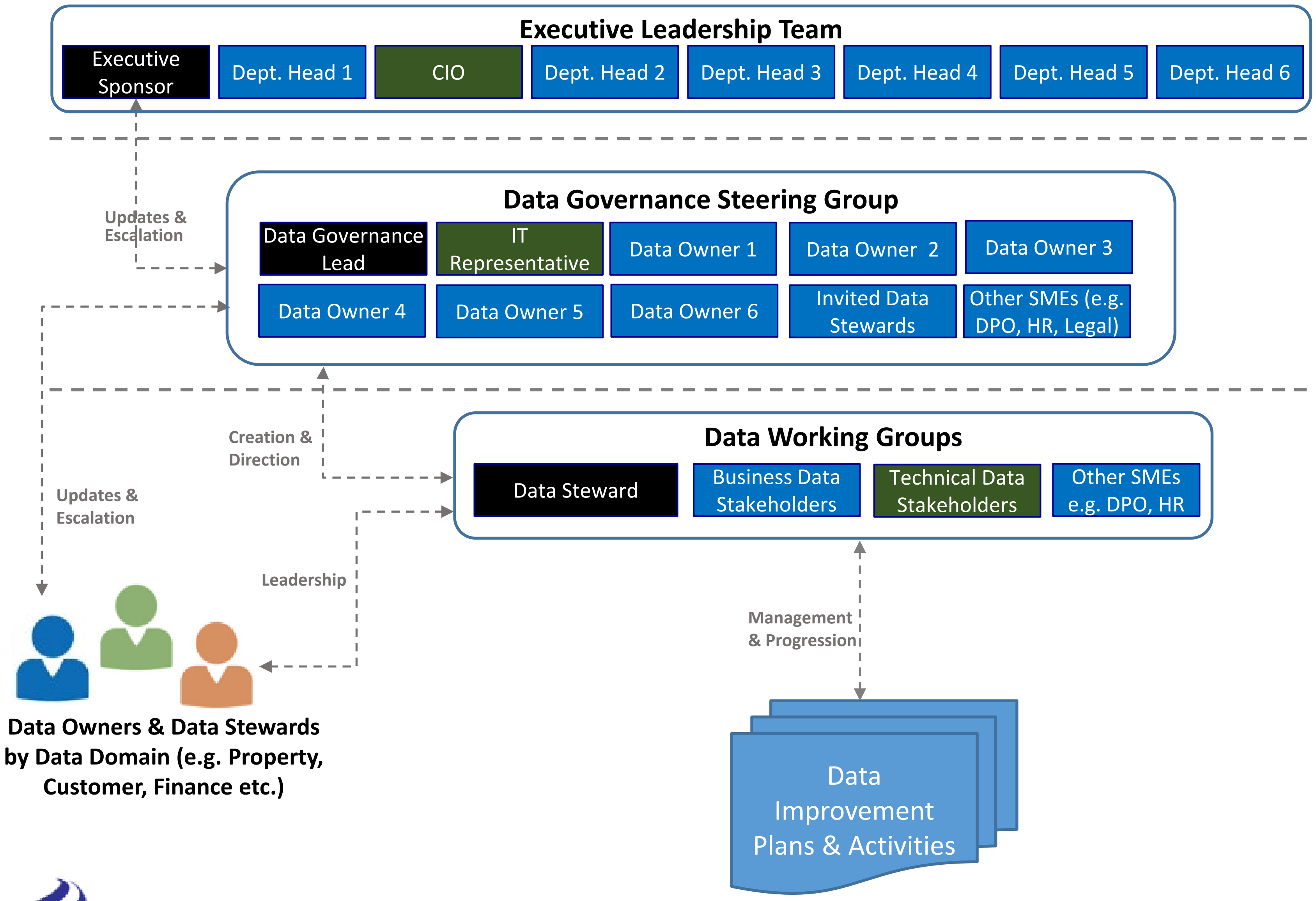
- Oversees the holistic data architecture for the organisation, including data models, data standards, data integration, etc.
- Works with both business and technical stakeholders to ensure that systems implementations align with key business rules & needs

Data Security Lead*

- Ensures that the organization adheres to the adequate security standards to support industry regulations and best practices
- Works with the Data Governance Lead and Data Architecture to ensure that data implementations support business needs in a secure way.

* Typically a full-time role

Typical Data Governance Organisational Structure



Executive Leadership Team

Executive Sponsor | Dept. Head 1 | CIO | Dept. Head 2 | Dept. Head 3 | Dept. Head 4 | Dept. Head 5 | Dept. Head 6

Updates & Escalation

Data Governance Steering Group

Data Governance Lead | IT Representative | Data Owner 1 | Data Owner 2 | Data Owner 3 | Data Owner 4 | Data Owner 5 | Data Owner 6 | Invited Data Stewards | Other SMEs (e.g. DPO, HR, Legal)

Creation & Direction

Updates & Escalation

Data Working Groups

Data Steward | Business Data Stakeholders | Technical Data Stakeholders | Other SMEs e.g. DPO, HR

Leadership

Management & Progression

Data Owners & Data Stewards by Data Domain (e.g. Property, Customer, Finance etc.)

Data Improvement Plans & Activities

Executive Level

- Executive support
- Data advocacy
- **Vital to have ELT level sponsor and champion**

Strategic Level

- Sets strategic direction for Data Governance
- **Owns the Data Governance Roadmap**
- Identification of working groups as needed
- Funding within budgeted amounts for data quality & governance
- Identification of data stewards for key data areas
- Arbiter in the case of conflicting needs, definitions, or priorities around cross-functional use of data.

Tactical Level

- **Create Data Improvement Plans for Data Areas etc.**
- Propose and progress data improvement initiatives
- Report progress to Data Governance Steering Group
- Escalate cross-domain issues and barriers to DG Steering Group

Key

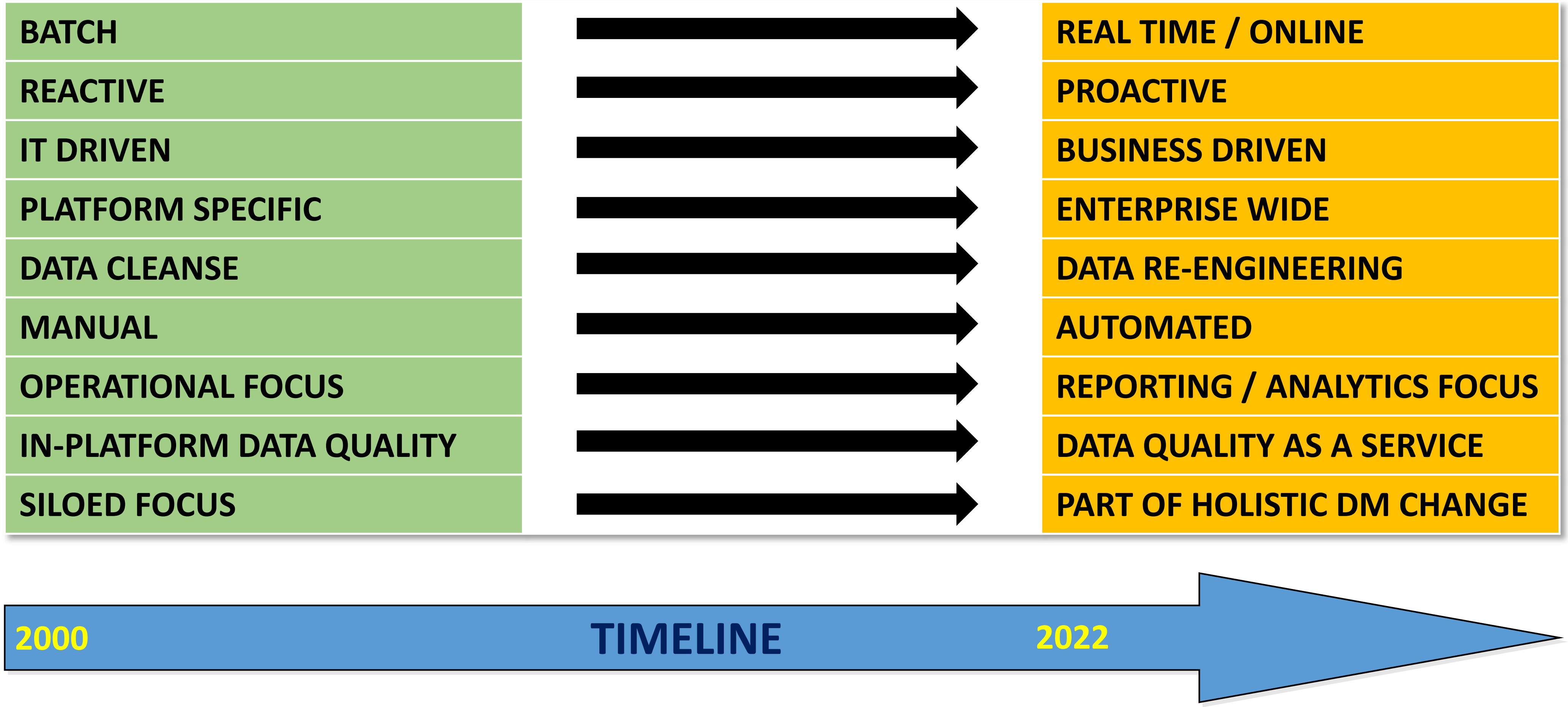
■	Lead Role
🟩	Technical Role
🟦	Business Role

The A2E Approach: Summary Benefits

- Views Data Quality as a holistic problem requiring holistic solutions:
  - People, Process, Technology
  - Supports incremental and continuous data quality improvement
- A reusable methodology that can be applied at multiple organisational levels:
  - Organisation
  - Function (e.g. Finance, Order Fulfilment etc.)
  - Department (e.g. Sales, Marketing, HR etc.)
  - Data Domain (e.g. Customer, Product, Billing)
  - Specific Data Quality Issue (e.g. missing contact data in CRM platform)
- Identifies focus and priority:
  - Identifies 'quick wins' to accelerate momentum
  - Ensures resources are targeted at highest priority problems

Data Quality: The Future

Evolution of data quality since 2000… Evolving Approaches



BATCH	→	REAL TIME / ONLINE
REACTIVE	→	PROACTIVE
IT DRIVEN	→	BUSINESS DRIVEN
PLATFORM SPECIFIC	→	ENTERPRISE WIDE
DATA CLEANSE	→	DATA RE-ENGINEERING
MANUAL	→	AUTOMATED
OPERATIONAL FOCUS	→	REPORTING / ANALYTICS FOCUS
IN-PLATFORM DATA QUALITY	→	DATA QUALITY AS A SERVICE
SILOED FOCUS	→	PART OF HOLISTIC DM CHANGE

The Future of Data Quality

New approaches for data quality in digital organisations

- **Increased focus on real time data validation** and improvement at the point of data creation, ingestion or use
  - Automated digitised processes and self-service will fail if the data is not fit for purpose
  - Data validation and improvement must be done in real time on large data volumes & varieties
  - ‘After the fact’ data cleanse and improvement is now too late
  - Opportunity to exploit IoT and AI to develop self-checking data quality capabilities through machine learning
- **Business users need more control** over the creation & management of business rules
  - They need the ability to create business rules dynamically when preparing data
  - Different data users will potentially require the application of varying business rules
  - The paradigm where business rules are created and held centrally by IT is obsolete
- **End user self-service data quality** functionality is essential
  - Data preparation & formatting
  - Data parsing & cleansing
  - Data enhancement & enrichment
- Toolsets must support a **wider variety of platforms & data types**
  - Legacy and Big Data environments (e.g. Data Warehouses and Data Lakes)
  - Real-time and batch
  - Structured / semi-structured / unstructured data types – via Data Profiling, Data Preparation & Metadata tagging

Summary & Conclusions

Course & Learning Objectives

- Understand what ‘fit for purpose’ data is, and is not
- Describe the dimensions of data quality
- Know the main causes of poor data quality
- Highlight the impact of poor data quality on individuals and organisations
- Understand the relationship between data quality and other data management disciplines, with particular emphasis on data governance
- Highlight the shortcomings of traditional ways of tackling poor data quality and the importance of a holistic approach, involving people, process and technology
- Learn the five steps of the A2E methodology and how to apply it to identify, prioritise and address data quality problems
- Specify and apply the main activities and deliverables of each of the five steps
- Be able to understand and develop business rules to baseline data quality and to set improvement thresholds
- Be aware of software tools that can help to support and automate the A2E approach

DISCUSSION

- How well do you feel the course has met its stated objectives?
- How well have your personal objectives been met?

Summary & Conclusions

- Data quality is complex because **businesses and organisations are complex**
- Addressing data quality issues requires a **holistic approach combining people, process, and technology change**
- Data governance is needed to sustain data quality improvement **– it orchestrates the people, processes and organisational structures** required to improve data quality
- Build quantifiable Data Improvement Plans to **show demonstrable RoI and implement a culture of continuous data quality improvement**
- It's vital to **deliver frequent incremental improvements** to maintain business interest and backing
- Data quality is a multi-dimensional issue for organisations so **tackle it through multi-dimensional approaches such as A to E**

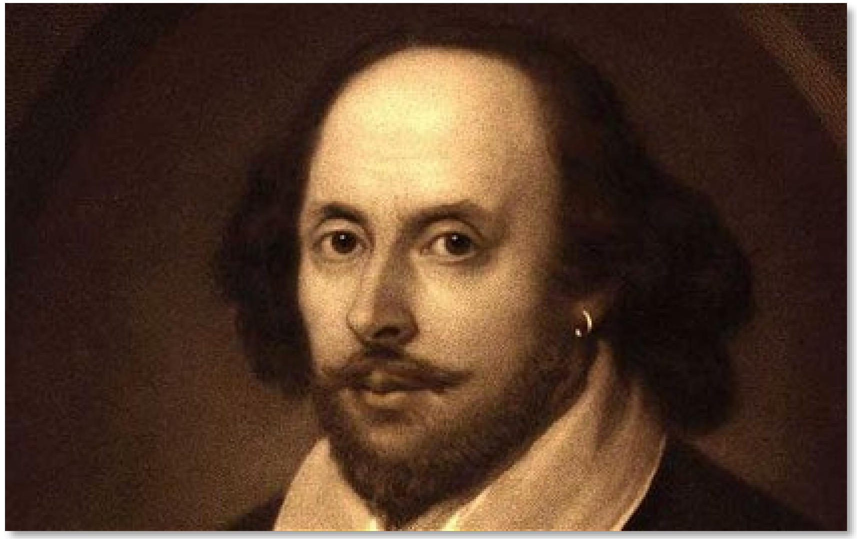
Contact Info

- Email: nigel.turner@globaldatastrategy.com
- Twitter: @NigelTurner8
- Website: www.globaldatastrategy.com
- LinkedIn: uk.linkedin.com/in/nigelturnerdataman

Final Thoughts

“It is not in the stars to hold our destiny but in ourselves”

William Shakespeare
(Julius Caesar)



“The harder I practise, the luckier I get.”

(Gary Player, Golfer)

Good luck in your Data Quality journey and hope our paths cross again…