The Importance of Data Quality

- And How to Manage It

Dr Mike Osborne
Managing Director
OceanWise Ltd

Information Intelligence

- How good are you at spotting fake news?
- How about poor data quality?
- What is poor data quality?
- What is data quality?
- How do I measure and manage it?

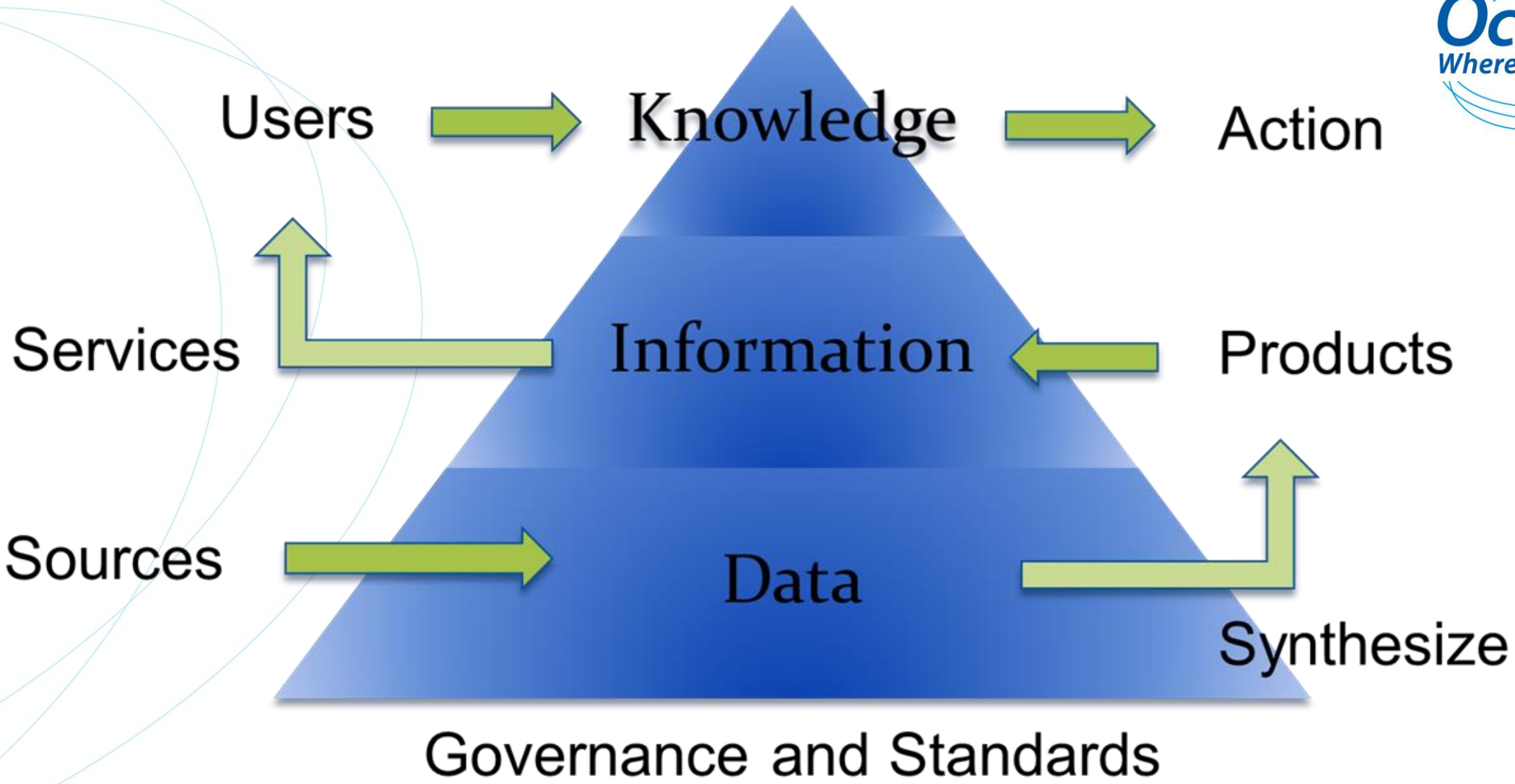
Trust in Data

- Authority and Reputation
- Crowd or User Experience
- Level of Use -> Consequences
- Supplier Relationship (Fee or Free?)
- Verification – consciously or not
- But how and is this enough?
- What is the risk of the ‘black box’?

Source: Daily Telegraph

Some Data Definitions

Data	Raw facts i.e. numbers, words, dates, images, sounds etc. without context
Information	Data put into context e.g. in a sentence or associated with field names/headings
Data Product	Data modified or aggregated for a particular purpose e.g. navigation
Data Service	Data or Data Product delivered to a User
Data Dictionary	Database of terms (to look-up and use) e.g. feature catalogue
Master Data	Core or essential data an organisation cannot do without



Ubiquitous Data-Information-Knowledge Triangle and adapted by OceanWise (2012)

International Standards

- Data Quality Management Framework (ISO 8000)
- Systems and Software Requirements and Evaluation (ISO 25001)
- Geographic Information — Data Quality (ISO 19157)
  - Part 1 General requirements
  - Part 2 XML schema implementation
  - Part 3 Quality assurance of data supply
- Quality Management (ISO 9001) – NOT data specific
- ISO Business Management High Level Structure (Annex SL)
  - Clause 8 Operating Procedures – including for data

ISO Generic Data Quality Model

Divides Data Quality into three categories:

- Inherent Quality

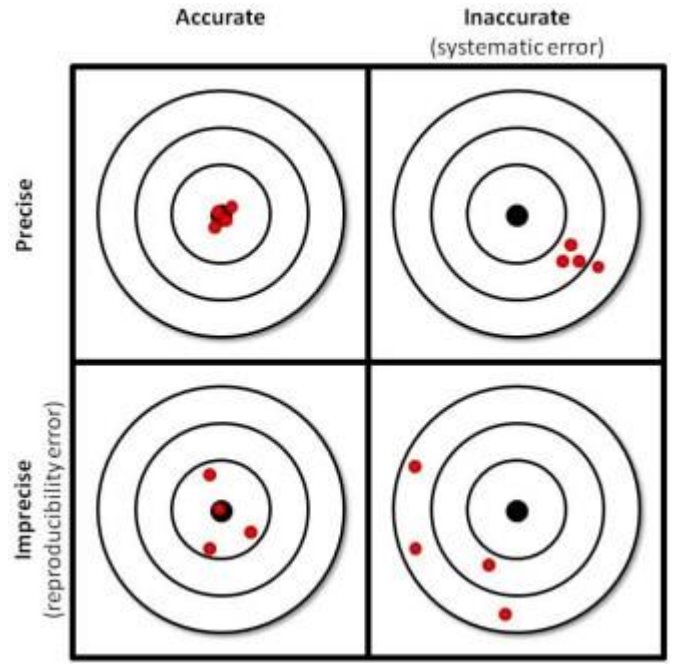
  e.g. Accuracy and Completeness

- Technology Dependent Quality

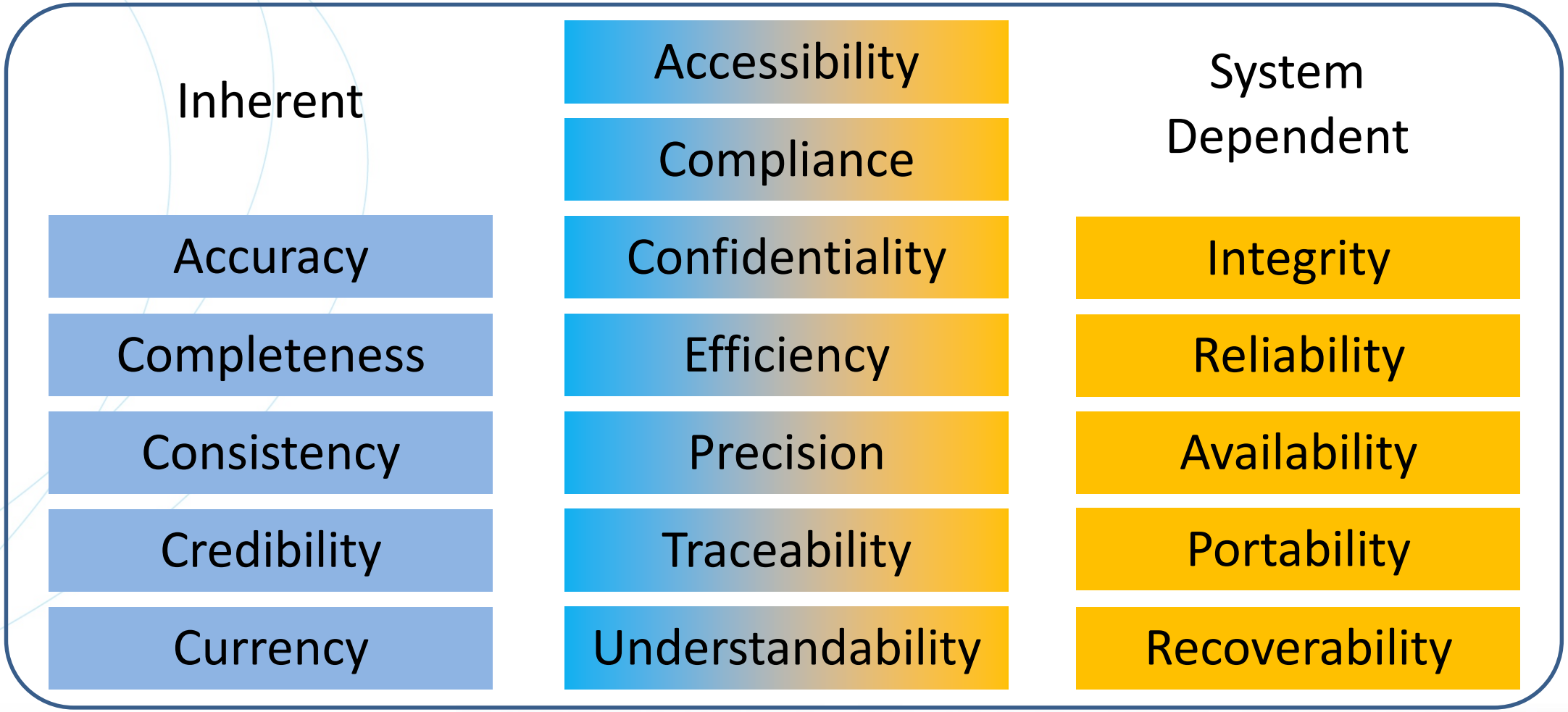
  e.g. Availability and Recoverability

- Inherent **and** Technology Dependent Quality

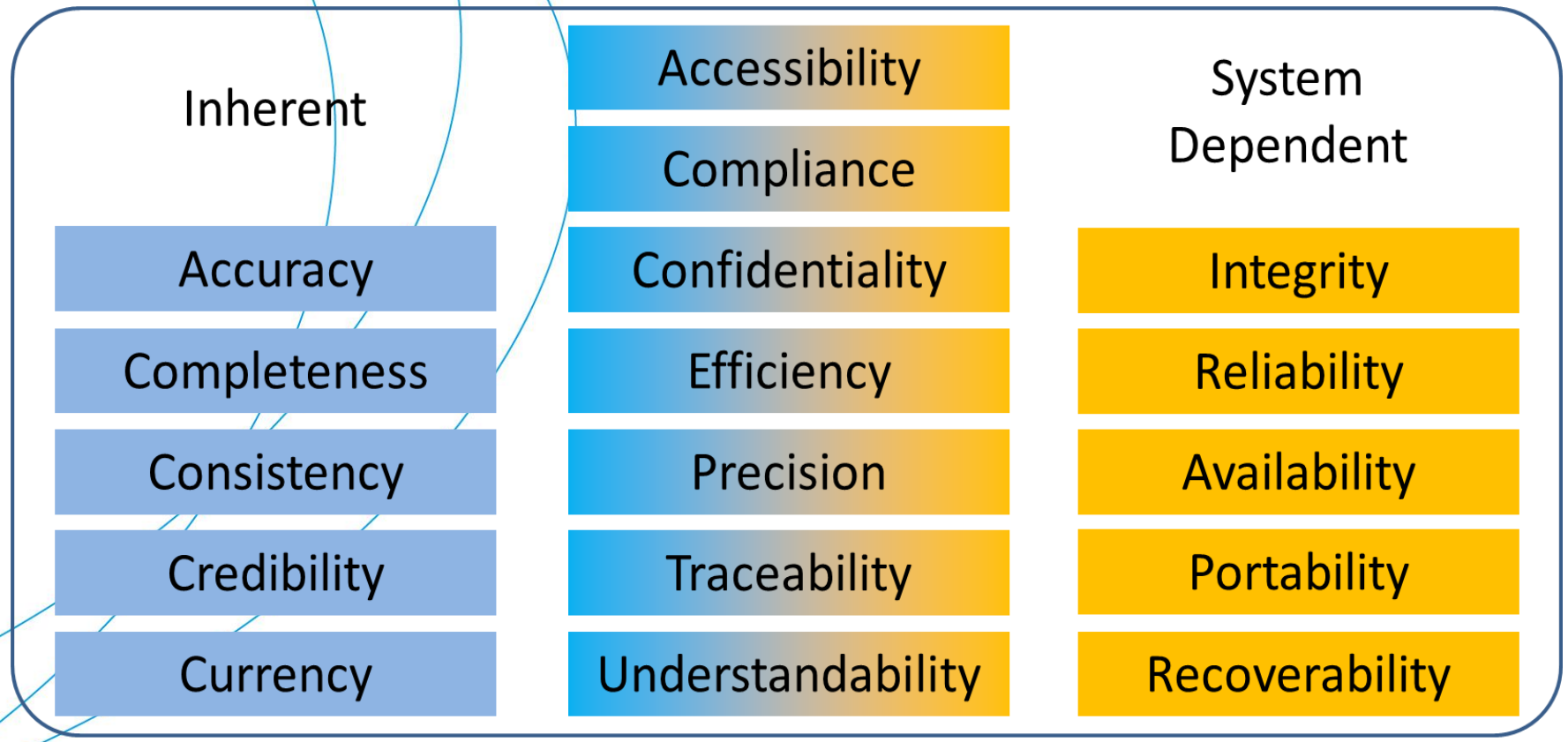
  e.g. Precision, Understandability and Traceability



ISO Generic Data Quality Model



ISO Generic Data Quality Model

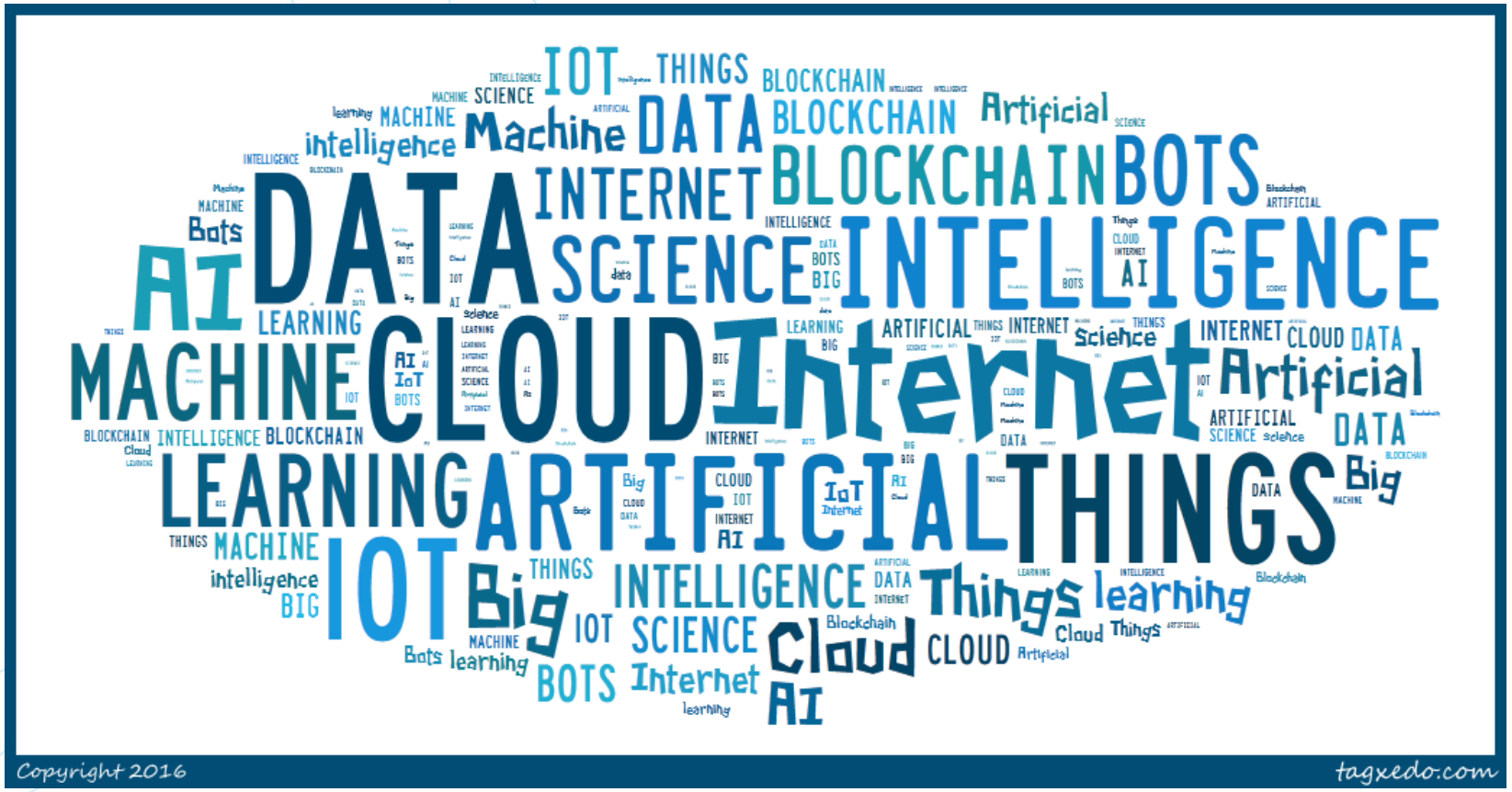


Objective versus Subjective Metrics

Subjectivity means 'Fitness for Purpose'

Purpose must be defined **and** communicated

Latest Technologies

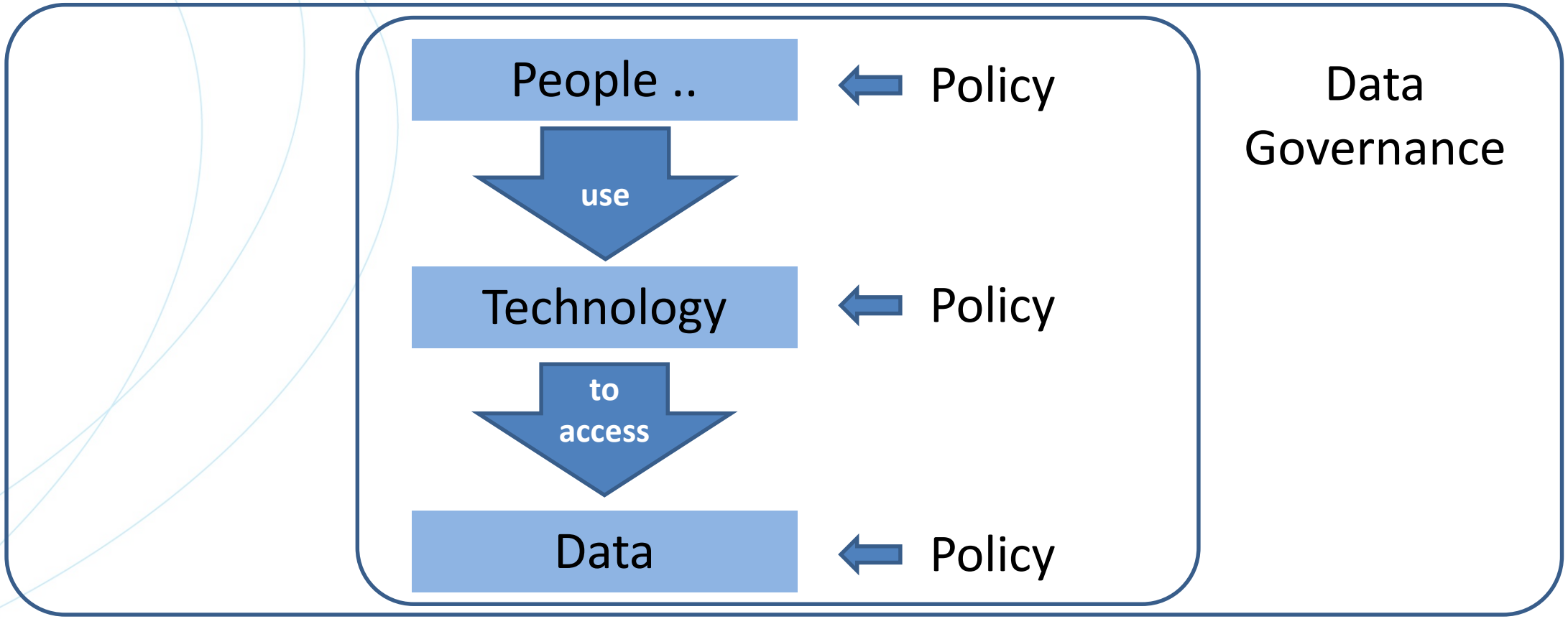


All require understanding and addressing Data Quality issues

Some Data Quality Processes

• Integrity	Checksums
• Reliability	Buffering / Acknowledgement
• Understandability	Calibration (scientific units referenced to a defined datum)
• Credibility	Validation - Sensible and in range
• Completeness	All records present or valid nulls
• Consistency	Enumerated lists / Dictionaries
• Accuracy	Unbiased or purpose defined
• Availability	Open formats (i.e. not vendor specific)

Data Quality Management System



Data Governance

- Data Governance is the execution and enforcement of authority over the management of data-related resources
- No governance = Data anarchy
- Data Governance needs to be communicated and involves internal and external stakeholders

Source: sciphilos.info

Data Governance Concepts

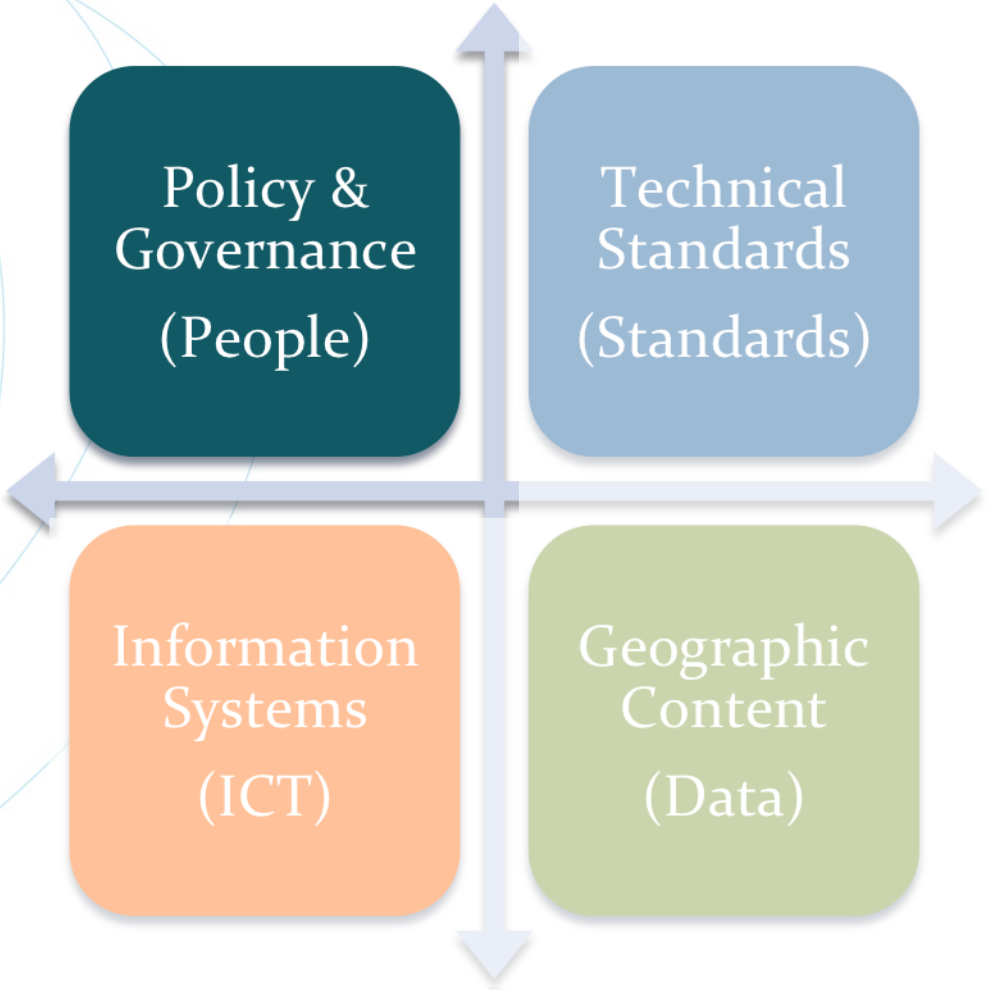
- Key data items and domains are identified and defined:
  - What are they? (Customer, Supplier, Finance etc.)
  - Where are they are held?
  - Who needs to access them and how?
- Individuals are made accountable for data within their domain –> **Data Stewards**
- Critical data is defined, indexed, measured regularly and reported on by Stewards –> **Master Data**
- As problems are identified (reported), initiatives are launched to address them –> **Data Improvement**

Data Quality Definitions

Quality Control	Verification process by which data not complying with predetermined criteria are identified, flagged or rejected
Quality Assurance	Roles, structures and processes for monitoring, measuring, reporting and remediating data quality issues
Data Quality Management	Key component of broader data governance activities via data-quality-specific policies and processes

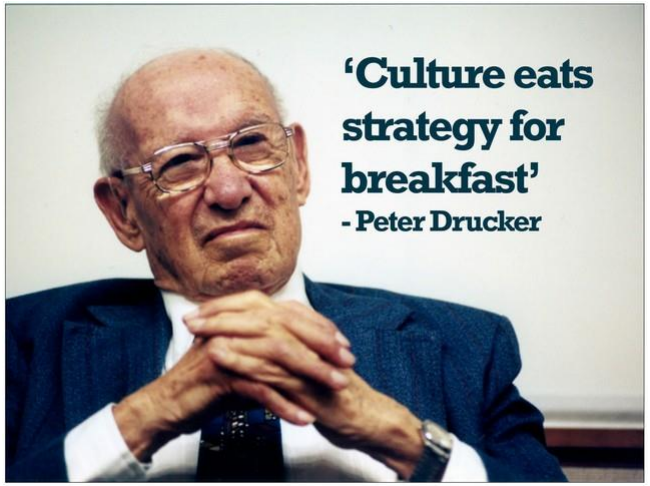
Ref: Gartner Inc. Dec 2016 "Magic Quadrant for Data Quality Tools" report

Making the Case



Four Pillars of Spatial Data Infrastructure (OceanWise, 2012)

Where's the biggest challenge?



Key Messages

- Data Quality is
  - Complex, multi faceted, inherent, technological **and** combined
  - Should be considered a business issue, not a technical or IT/IS issue
- Data Quality Management is integral to Data Governance
- Data Governance should be part of a Business Management Framework – like Quality and Health & Safety – and it can be
- Data Governance and Management are now key skills required throughout any organisation

Thank you for listening …

Talk to us about Data Management today

www.oceanwise.eu

mike.osborne@oceanwise.eu