Smart Water
Infrastructure Group
WATER NEW ZEALAND

Data Quality in the Water Industry

Presenters:

Nicolette Voskuilen, Quentin Griffiths, Greg Preston

water NEW ZEALAND
The New Zealand Water & Wastes Association Waiora Aotearoa

Why is Data Quality Important?

Turbidity:

0.2 *vs* 0.2

Metadata – Information About the Data Point

0.2

- Units: NTU
- Sample Date: 23/08/2023
- Sample Time: 11:58am NZST
- Temperature: 15.2°C

0.2

- Units: NTU
- Sample Date: 23/8/2023
- Sample Time: 7:08
- Temperature: not recorded

Accuracy or Precision?

0.2

- Method: Turbidimeter
- Last calibration date: 21/7/2023
- Meter ID: TURB-1234
- Sample Location: Treated Water Tank Inflow

0.20

- Method: Field Sample
- Last calibration date: ?
- Meter ID: ?
- Sample Location: Reservoir

How do we FAIR?

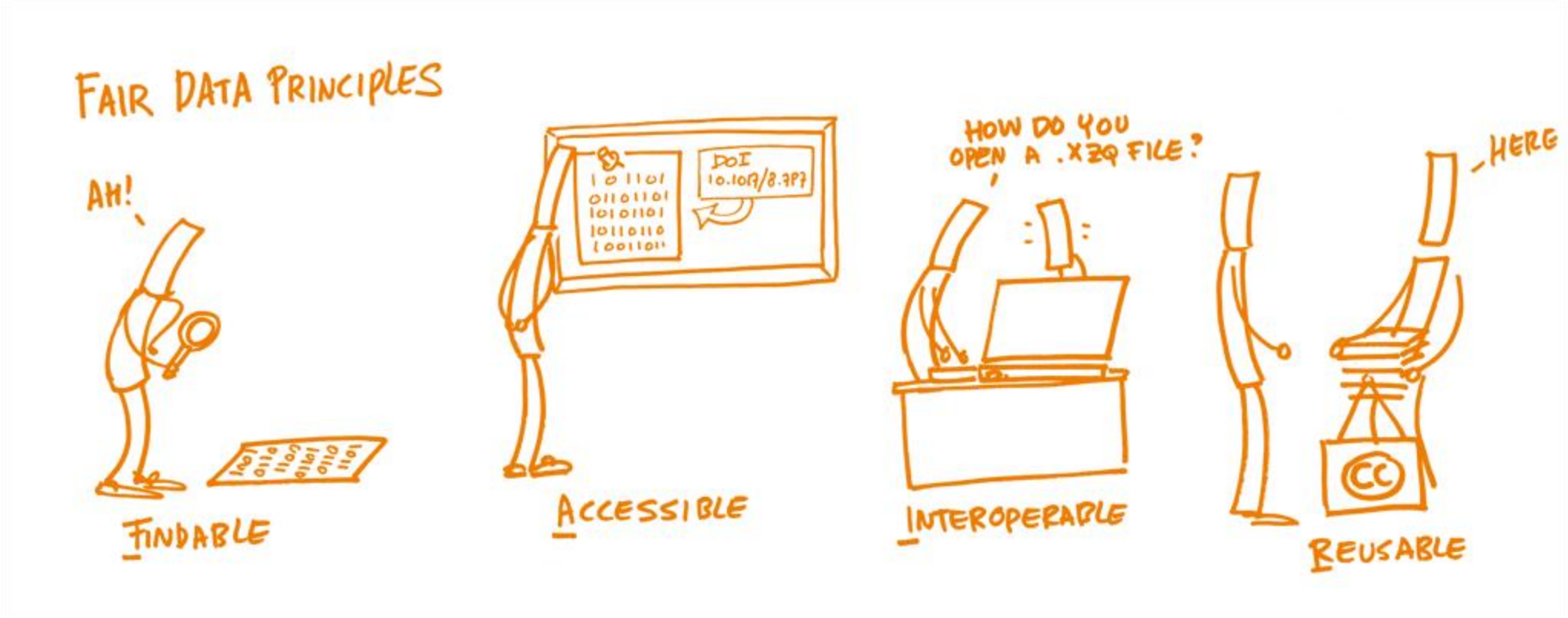


Image Credit: https://www.fosteropenscience.eu/

How do we FAIR?

Image Credit: https://www.fosteropenscience.eu/

How do we FAIR?

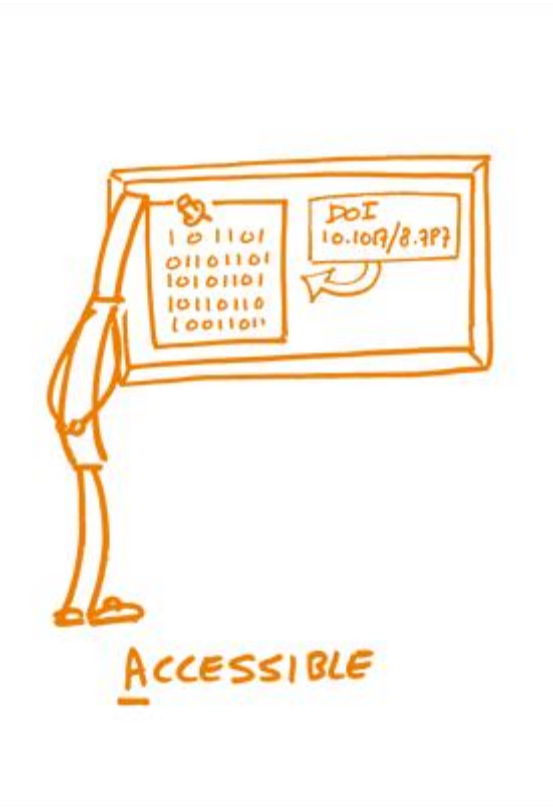


Image Credit: https://www.fosteropenscience.eu/

How do we FAIR?

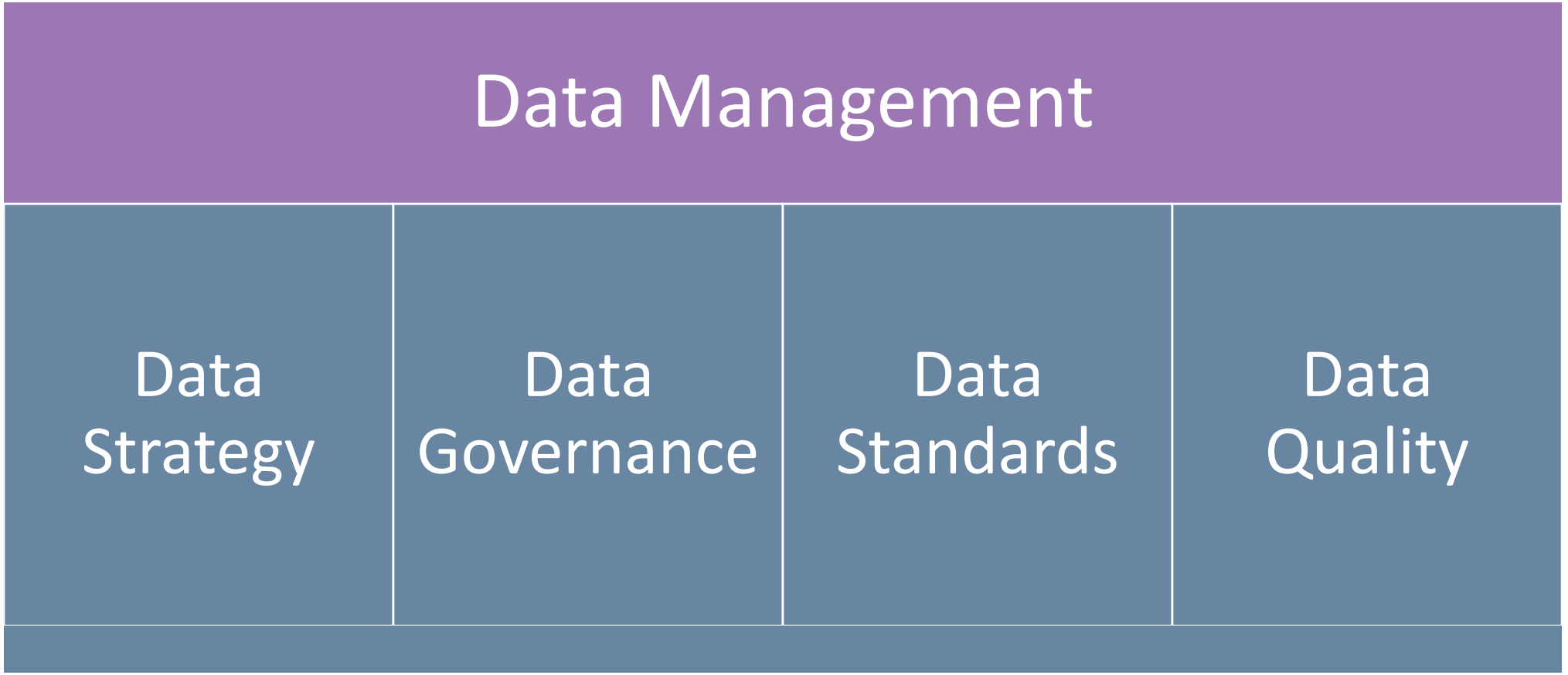
Image Credit: https://www.fosteropenscience.eu/

How do we FAIR?

Image Credit: https://www.fosteropenscience.eu/

Data Quality - Underpinning Data Management

Four pillars of data management



Strategy, Governance, Standards

Data Strategy

- How does the data align with the uses of those data?

Data Governance

- How do you ensure the processes that enable data, security, integrity, etc?

Data standards

- What metadata is defined and what format are the data in?

Data Quality Assessment Framework

Provide a methodology to:

- Assess the quality of data
- Measure and compare the quality of data between different datasets
- Communicate the quality of data in a given dataset

Data Quality Assessment Framework

Three stages:

1. Data Quality Checks: simple assessments of data with discrete outcomes (e.g. Yes/No)
2. Data Quality Metrics: a quantified aggregation of the results from the Data Quality Checks, designed as interpretable indicators of data quality
3. Data Quality Communication: a set of reporting recommendations to effectively communicate the most important aspects of the quality of data

1. Data Quality Checks

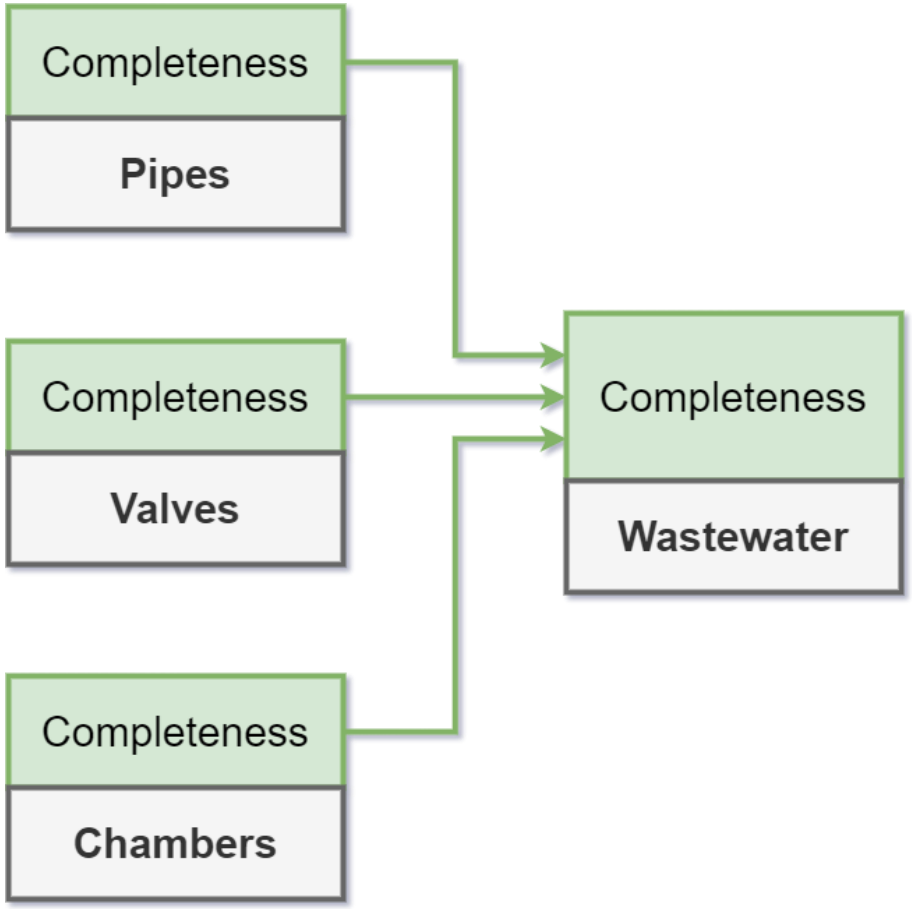
Based on Data Standard. Three types:

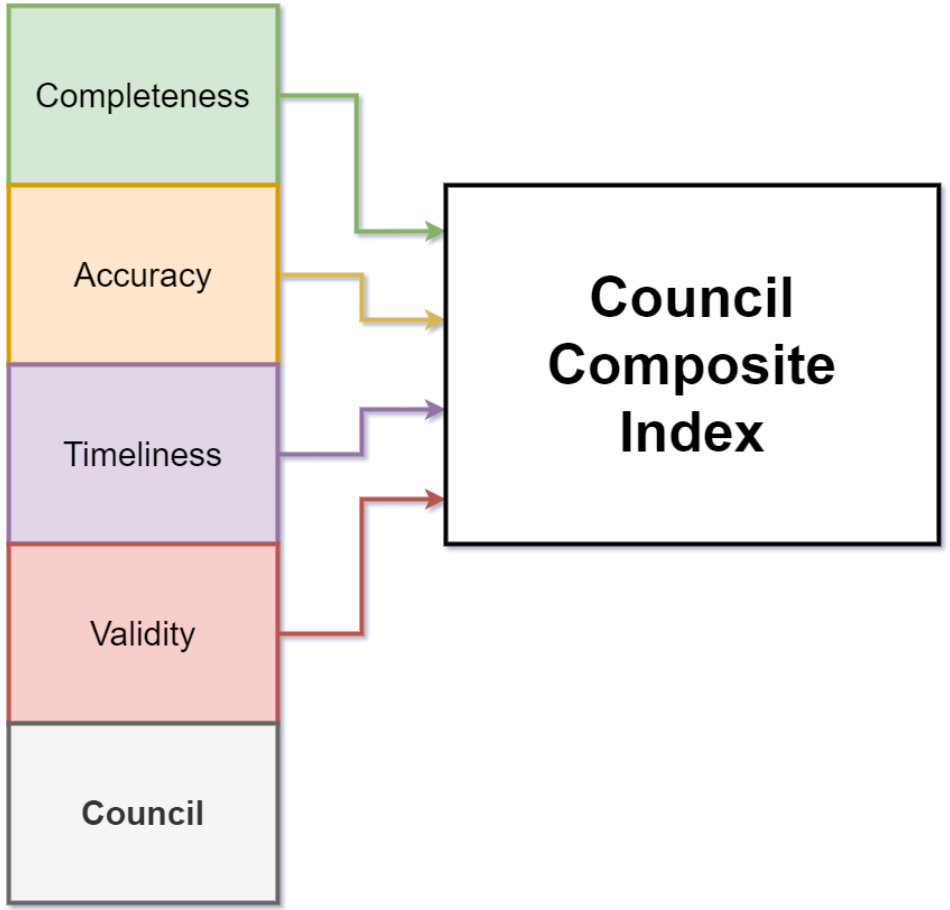
- Dataset Level Checks are carried out on each attribute in a dataset
  - e.g. Does the attribute have data in the same data type as its counterpart in the CoP?
- Asset Level Checks are carried out on each attribute, on each asset in a dataset – some are generic, some are attribute-specific
  - e.g. Generic: Is there data recorded in this attribute for this asset?
  - e.g. Specific: For installation dates, is the date a valid date?
- Geospatial Level Checks are carried out on a network level for a given pipe asset
  - e.g. Is the pipe in the correct geographic location (within a council's governing region)?

2. Data Quality Metrics

Metric	Description
Completeness	The data is comprehensive and does not contain missing values
Accuracy	The data is a true reflection of the real-world values
Validity	The data contains values in a valid type and format
Timeliness	The data is readily available when expected and needed
Data Standard Alignment	The data structure and values are aligned with the appropriate Data Standard

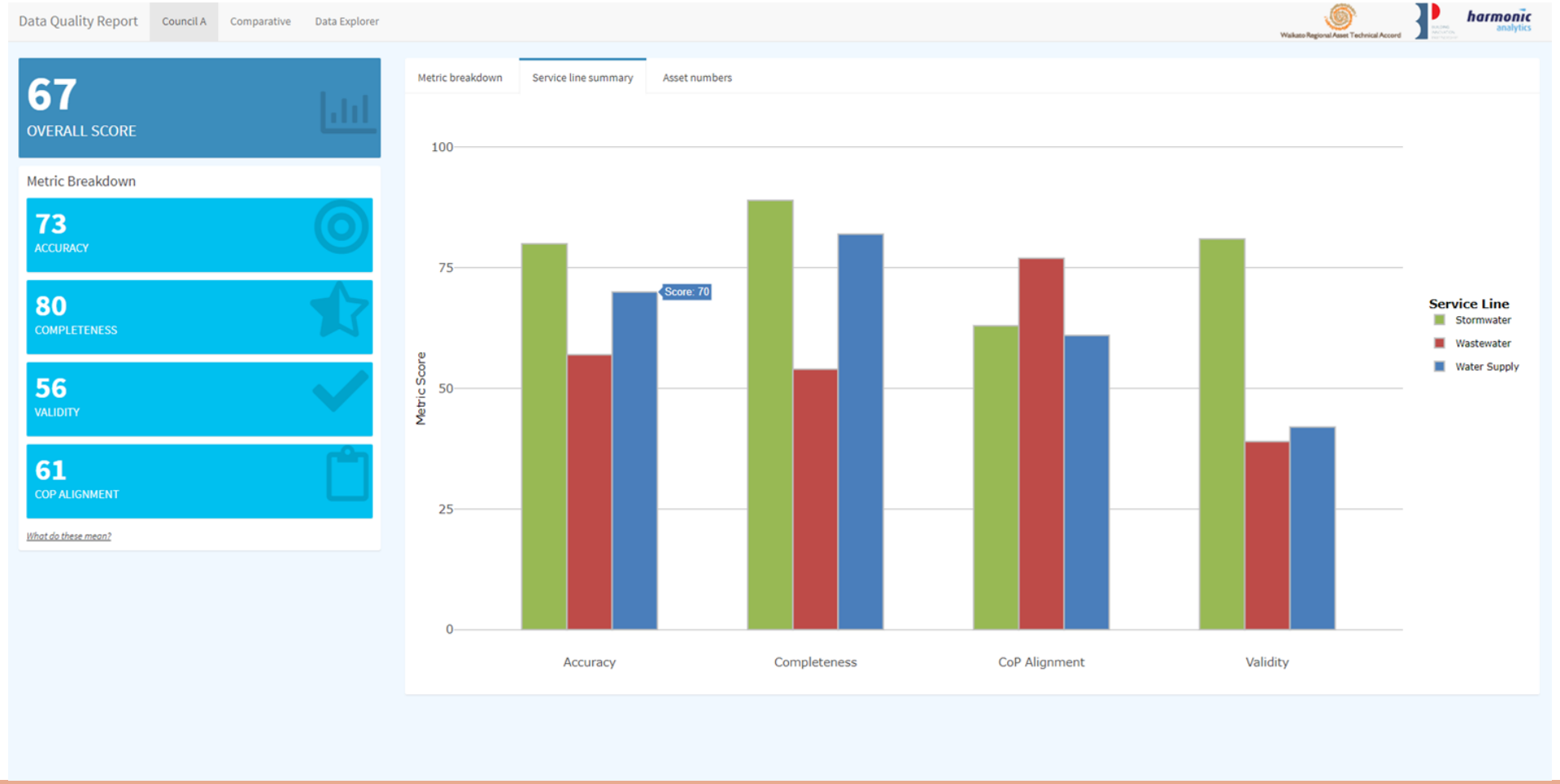
Data Quality Metrics – Asset Data



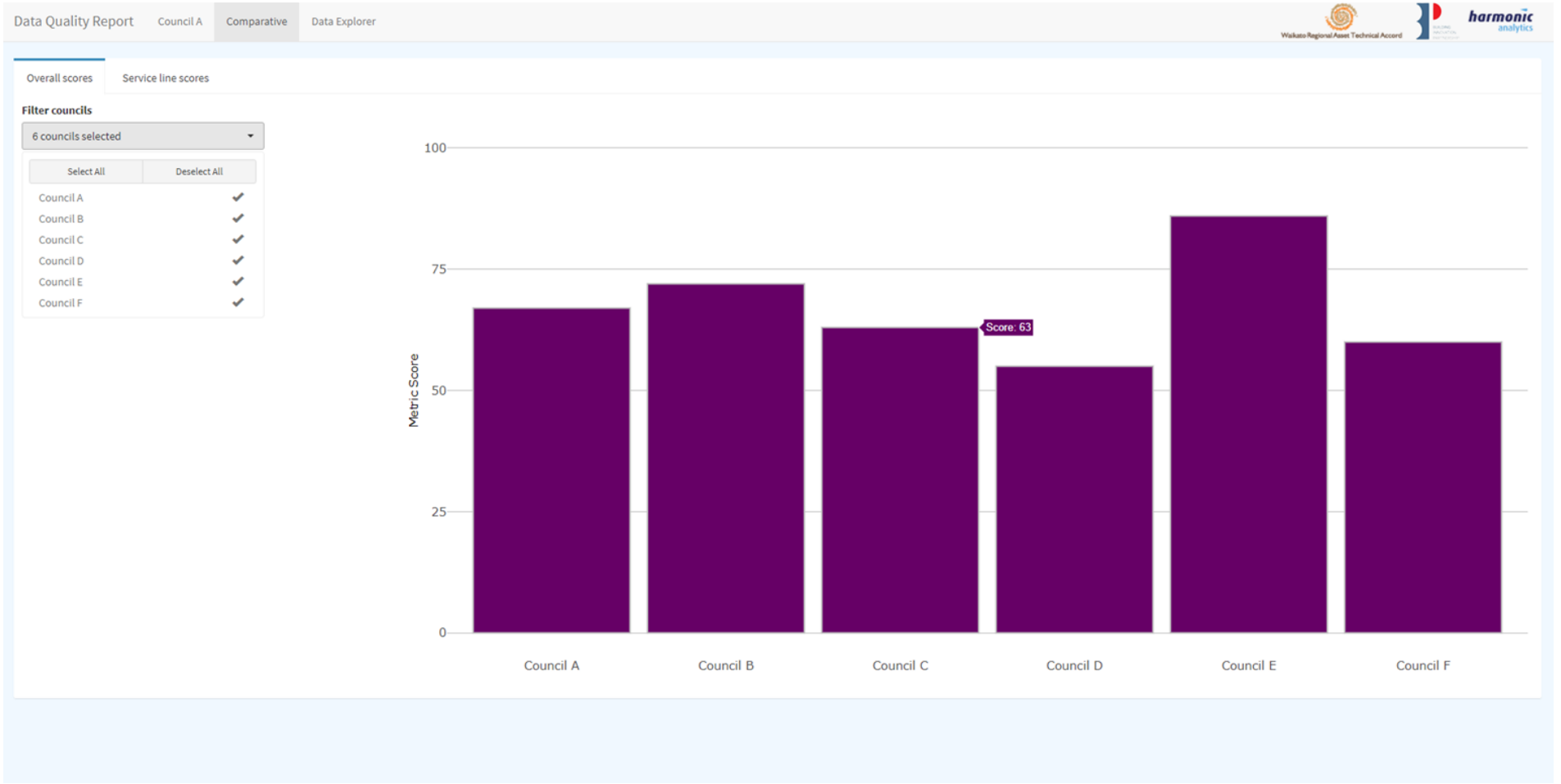


3. Communication Dashboard

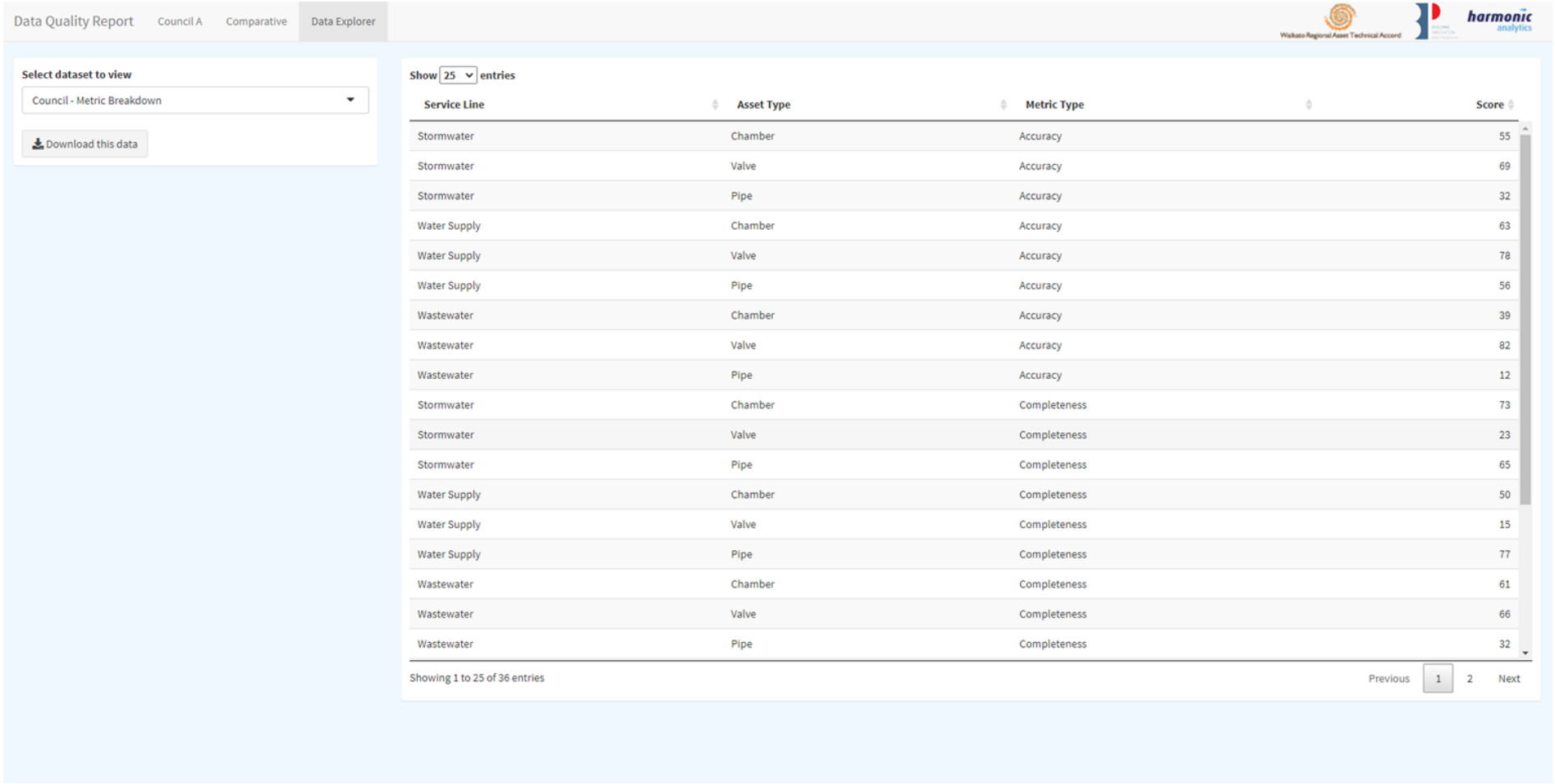
Breakdown by service type



Breakdown by Council



Breakdown by individual asset



Conclusion

Useful data is underpinned by:

- A data strategy
- Strong data governance
- Clearly defined data standards
- A robust data quality framework which will:
  - Reference the data standard
  - Set benchmarks for comparison
  - Have a range of aspects including:
    - Completeness, Accuracy, Validity & Timeliness

Smart Water
Infrastructure Group
WATER NEW ZEALAND

Thank you!
Questions? Patai?

water
NEW ZEALAND
The New Zealand Water & Wastes Association Waiora Aotearoa