

Better Data Quality = Better Cancer Care



CANCER RESEARCH UK

RADNET MANCHESTER

Michael Adcock¹; Anthony Lighterness¹; Catherine O'Hara¹; Lauren Scanlon¹; Gareth Price^{1, 2}

¹The Christie NHS Foundation Trust; ²University of Manchester

Contact: Michael.Adcock2@nhs.net; Anthony.Lighterness@nhs.net

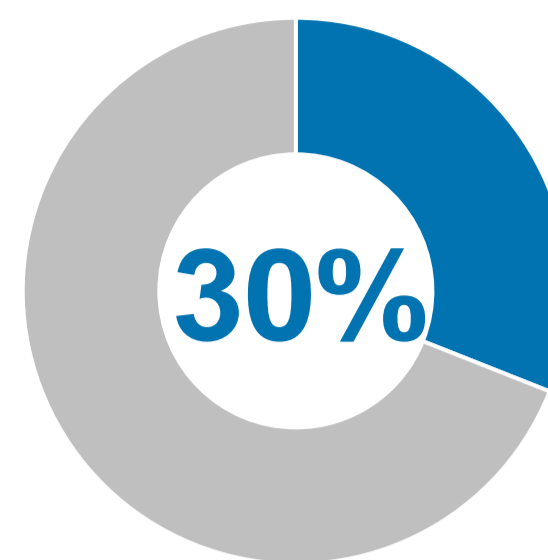
MANCHESTER CANCER RESEARCH CENTRE

NHS The Christie NHS Foundation Trust

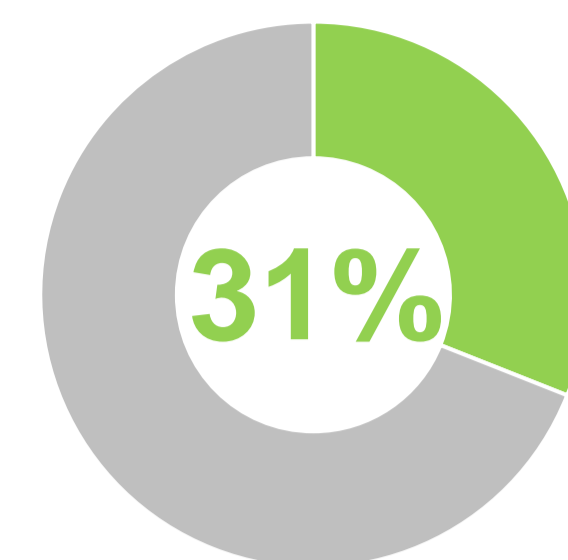
MANCHESTER 1824 The University of Manchester

Background

Data-driven analytics and modelling can improve patient care and organisational operations. However, data quality issues hinder progress and costs money. Current data quality profiling methods are scarce and mostly limited to "missingness" or "completeness". We present a new framework for approaching data quality within healthcare settings.



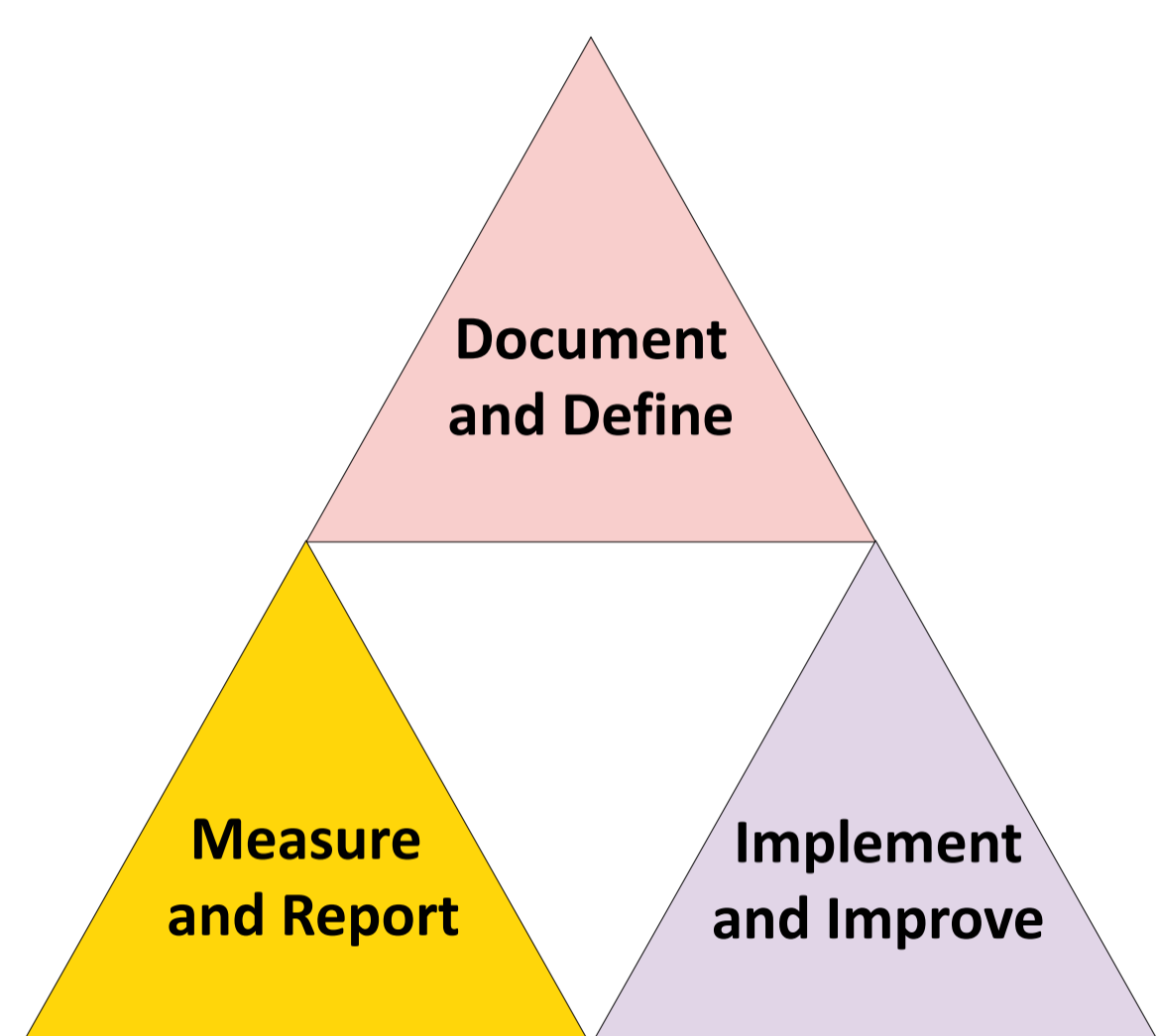
Estimated proportion of revenue spent on handling data quality issues¹



Proportion of AI tools that fail due to data quality issue²

Aim: To establish an innovative framework to quantify data quality automatically and at scale across any structured clinical data

The Framework



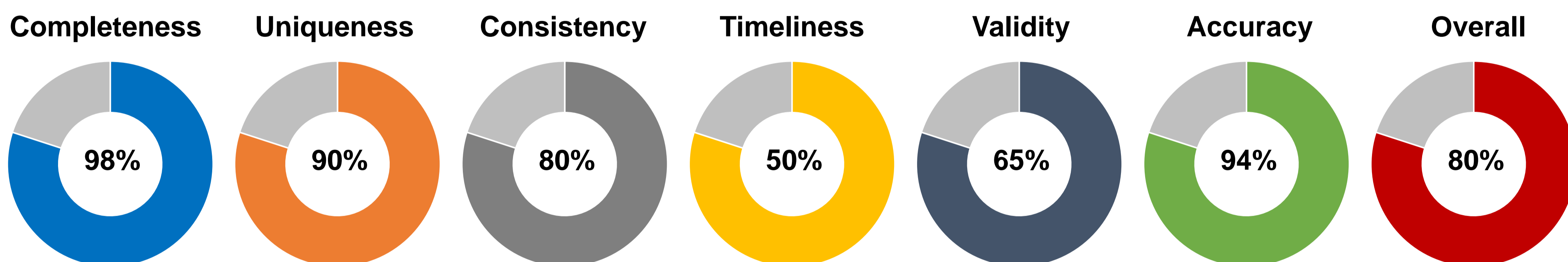
- 1) Define how data should behave, be stored and used.
- 2) Measure the data quality metrics and feed to data end-users such as analysts, researchers & managers.
- 3) Make recommendations for organisational, behavioural, and technological improvements to improve data quality scores.

Data Quality Dimensions

Completeness Missing values in records, columns and tables. Missing records from tables.	Uniqueness Duplication of unique entities, records, columns.	Consistency Data should not contradict itself.
Timeliness Data should be up to date and received / updated in a timely fashion.	Validity The degree to which data conforms to expected ranges and formats. i.e. is the data plausible?	Accuracy Closeness between a value and reality.

What aspects need to be measured to inform data usage? For healthcare, it is essential that our dimensions cover a wide breadth and are quantifiable. Our framework is closely aligned to current UK government guidelines³.

Example Data Quality Dashboard



Dashboards are an ideal way to easily visualise and assess the quality of any defined dataset based on data quality scores for each dimension. Example use cases for data quality dashboards are for assessing suitability of data for internal or external report generation, judgment on whether data is suitable to answer research questions and to identify areas in data that require improvement.

Impact

This framework provides a systematic, scalable and generalisable method to profile data quality allowing for the opportunity to enrich gaps in data and identify systemic root causes of poor data quality. As changes are made, time series aspects of dashboards allow for tracking of data quality improvements over time. Data quality profiling can be incorporated into "confidence" of analytics for research or reporting. Importantly improving data quality improves the quality of patient care. Finally profiling and improving data quality allows us progress further to being a data mature organisation and unlock the full potential of our data.

References

1. GOV.UK. *Hidden costs of poor data quality*. <https://www.gov.uk/government/news/hidden-costs-of-poor-data-quality>
2. Venture Beat. *Why machine learning strategies fail*. <https://venturebeat.com/business/why-machine-learning-strategies-fail/amp/>
3. GOV.UK. *The Government Data Quality Framework*. <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>

Acknowledgements

This work was supported by Cancer Research UK RadNet Manchester [C1994/A28701].

<https://www.crukcentre.manchester.ac.uk/radnet>

