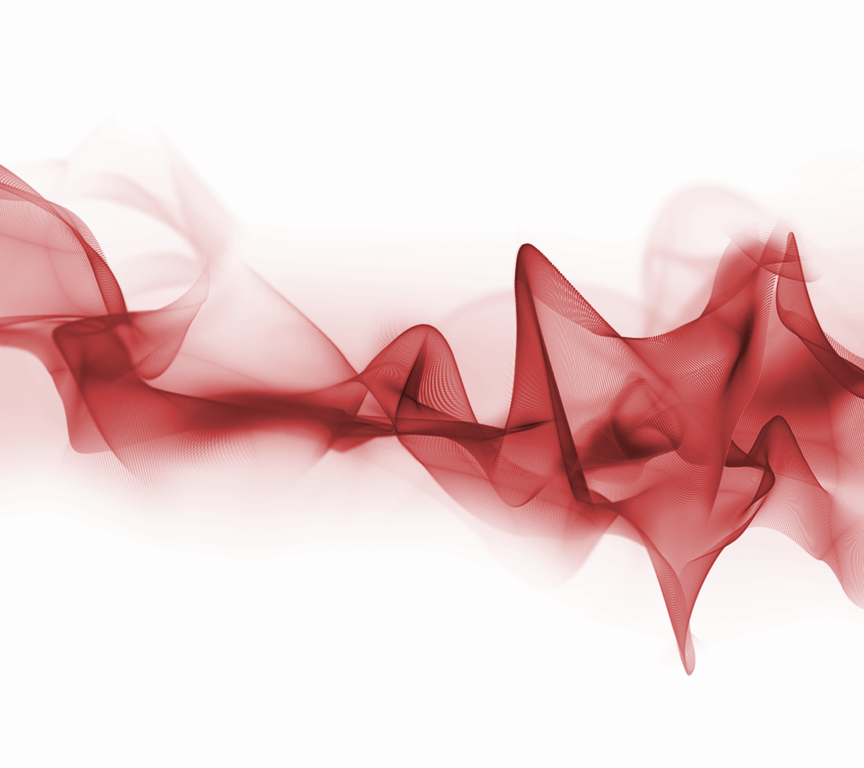
Data Quality Training

Explanations behind the Data Quality Induction Package



© Perspicacity Ltd 2018

|  |  |
| --- | --- |
| Document Control | |
| Version | V2.0 |
| Created | 2018.04.30 |
| Updated | 2019.12.20 |
| Original Author | Matthew Bishop, Perspicacity Ltd |
| Contact | [matthew.bishop@perspicacityltd.co.uk](mailto:matthew.bishop@perspicacityltd.co.uk) / 07545 878906 |

© Perspicacity Ltd 2018

This work product is licensed under the creative commons attribution 2.0 license. This allows worldwide, royalty-free, non-exclusive rights to share, distribute and create derivative work products (including for commercial gain) as long as the terms of the license are adhered to. Full details of the terms can be found at:

<https://creativecommons.org/licenses/by/2.0/uk/legalcode>

# Table of Contents

[Table of Contents 3](#_Toc27743687)

[Data Quality Strategy - Introduction 4](#_Toc27743688)

[Data Quality Strategy Series 4](#_Toc27743689)

[A little bit about Perspicacity Ltd 4](#_Toc27743690)

[Purpose of this document 5](#_Toc27743691)

[What can go wrong with data - Characteristics of Data Quality 5](#_Toc27743692)

[Data Capture Principles 6](#_Toc27743693)

[How the data quality issues are presented in the Board Report and Performance Reporting and how it links to the standalone data quality reporting 7](#_Toc27743694)

[How to use the standalone data quality reporting 7](#_Toc27743695)

[How to use the Validation Outcomes Feedback System (Validation tool) 7](#_Toc27743696)

[Data Quality Correction Principles 7](#_Toc27743697)

# Data Quality Strategy - Introduction

## Data Quality Strategy Series

This document is part of an open source series of documents to facilitate the establishment of a Data Quality strategy and function within an organisation. The full open source library can be found at <https://github.com/perspicacity-ltd/DataQualityReporting>

* The series presents a set of documents as a starter for ten
* It can be used by organisations starting on their data quality journey and those who already have a data quality function
* It contains the following components of a data quality strategy:
  + Strategy & Exec Summary
  + Policy & Standard Operating Procedures
  + Technical Specification (including link to reporting suite at <https://github.com/perspicacity-ltd/DataQualityReporting>)
  + Highlight Report Template
  + Training Materials
  + DQ KiteMark Images

## A little bit about Perspicacity Ltd

Perspicacity provides decision support consultancy, coaching, & development to the NHS. They are passionate about reducing the cost of software development to the NHS and aspire to create an active community of NHS and commercial organisations. They all share a common goal of improving the quality and efficiency of patient care through better, and more informed, decision making.

Open source helps the healthcare community to do this by sharing software development, learning from each other, and help software meet the needs of every organisation without being constrained to a single solution or paying for the same piece of work over and again across different organisations.

Although these Data Quality open source products are suitable for any organisation, healthcare or not, they are here as a result of wanting to freely share Perspicacity's collective products and ideas across the NHS and to widen the benefit of good, but usually locally isolated, projects further.

Perspicacity's open source offerings can be found at <https://github.com/perspicacity-ltd>

If you'd like to find out more, please contact Matthew Bishop on 07545 878906 or [matthew.bishop@perspicacityltd.co.uk](mailto:matthew.bishop@perspicacityltd.co.uk)

# Purpose of this document

This training material should explain to the user:

* + what can go wrong with data – characterising data quality
  + data capture principles
  + how the data quality issues are presented in the PTLs and how it links to the standalone data quality reporting
  + how to use the standalone data quality reporting
  + how to use the Validation Outcomes Feedback system (Validation tool)
  + data quality correction principles

# What can go wrong with data - Characteristics of Data Quality

When reviewing data quality, it is beneficial to use a pre-defined set of data quality characteristics to evaluate issues in order to more clearly articulate what the issue is and how it can be identified within the data. The following characteristics are proposed:

Drawing upon expertise from the Audit Commission, it is proposed the elective assurance programme utilises their six characteristics of data quality. A description of each characteristic is detailed in the appendices.

#### Accuracy (A)

Data should be sufficiently accurate for their intended purposes, representing clearly and in enough detail the interaction provided at the point of activity. Data should be captured once only, although they may have multiple uses. Accuracy is most likely to be secured if data are captured as close to the point of activity as possible. Reported information that is based on accurate data provides a fair picture of performance and should enable informed decision making.

The need for accuracy must be balanced with the importance of the uses for the data, and the costs and effort of collection. For example, it may be appropriate to accept some degree of inaccuracy where timeliness is important. Where compromises are made on accuracy, the resulting limitations of the data should be clear to their users. This must be a judgement determined by local circumstances, and is unlikely to be appropriate in the case of the data supporting published performance indicators.

#### Validity (V)

Data should be recorded and reported accordant to its definition and purpose. Where proxy data is used to compensate for an absence of actual data, organisations must consider how well this data is able to satisfy the intended purpose.

#### Reliability (R)

Data should reflect stable and consistent data collection processes across collection points and over time, whether using manual or computer-based systems, or a combination. Managers and stakeholders should be confident that progress toward performance targets reflects real changes rather than variations in data collection approaches or methods.

#### Timeliness (T)

Data should be captured as quickly as possible after the event or activity and must be available for the intended use within a reasonable time period. Data must be available quickly and frequently enough to support information needs and to influence the appropriate level of service or management decisions. It should be able to meet agreed contractual schedules for reporting.

#### Relevance (RV)

Data captured should be relevant to the purposes for which it is used. This entails periodic review of requirements to reflect changing needs. It may be necessary to capture data at the point of activity which is relevant only for other purposes, rather than for the current intervention. Quality assurance and feedback processes are needed to ensure the quality of such data.

#### Completeness (C)

Data requirements should be clearly specified based on the information needs of the organisation and data collection processes matched to these requirements. Monitoring missing, incomplete, or invalid records can provide an indication of data quality and can also point to problems in the recording of certain data items.

# Data Capture Principles

Effective data capture should use the data capture principles in the DQ Policy in order to improve the quality of data at the point of capture and reduce the chance of building a data quality problem.

# How the data quality issues are presented in the Board Report and Performance Reporting and how it links to the standalone data quality reporting

The principles of DQ measurement should be covered, explaining how we can use different data to try and identify each of the 6 characteristics of data quality. It should be explained that the data quality system is measuring many issues in the background, every night, when the data is extracted from source systems.

This can then lead on to demonstrating how all of these issues are presented in the PTLs and how clicking on the DQ indicator within a particular record in a PTL will take you through to the DQ reporting showing details of the record error(s).

The DQ reporting suite will need to be operational and in a reasonably mature position for this training material to be built.

# How to use the standalone data quality reporting

This section of the training can simply demonstrate the standalone DQ summary report, the “one measure, all records” detail report and the “one record, all measures” detail report. It should show how the user can navigate from one report to another using the links in the reports, giving a use case for each of the reports so the user can conceptualise how each report adds value.

# How to use the Validation Outcomes Feedback System (Validation tool)

The concept of “false positives” and “false negatives” should be covered, explaining that some DQ measures show a definite problem (e.g. missing information) but others can only propose that there is a potential problem (e.g. potential duplicate referrals).

This can then lead to explanation of how capturing validation outcomes feedback converts the unknown quantity of potential issues to a known quantity of correct records or corrected records.

The validation outcomes feedback system will need to be operational and in a relatively final state for this training material to be built.

# Data Quality Correction Principles

Effective data correction should use the data correction principles in the DQ Policy in order to increase the efficiency of correction and ensure the record is corrected consistently across all systems.