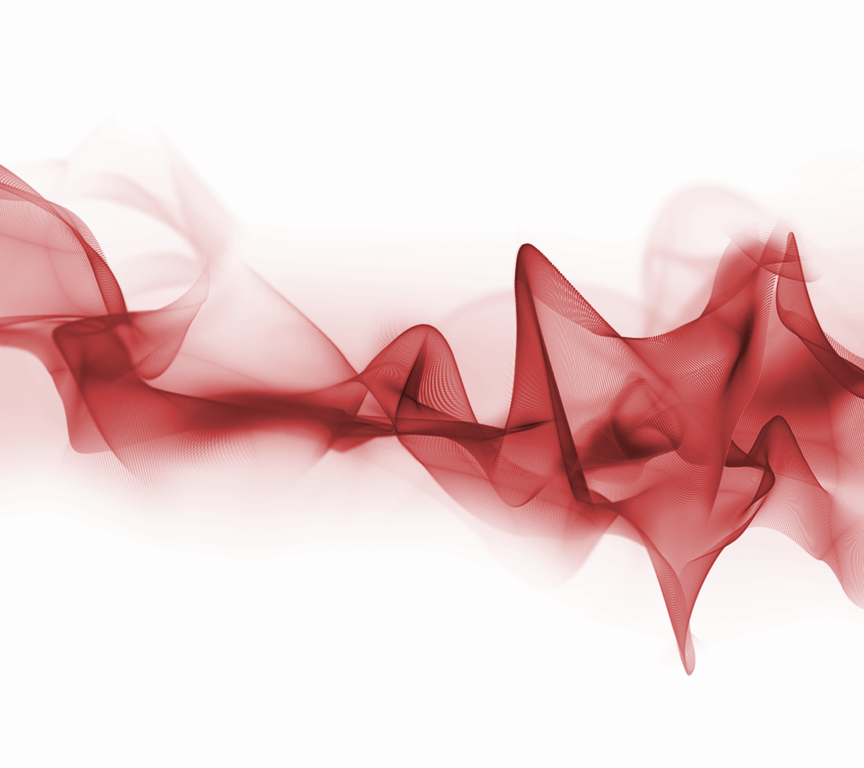
Data Quality Technical Development

Pre-Cursory Materials to Develop a DQ Reporting Suite

DQ Reporting Specification



© 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](#_Toc27735214)

[Data Quality Strategy - Introduction 4](#_Toc27735215)

[Data Quality Strategy Series 4](#_Toc27735216)

[A little bit about Perspicacity Ltd 4](#_Toc27735217)

[Overview 5](#_Toc27735218)

[Integration into the Trust Board Report 7](#_Toc27735219)

[Integration into Performance Reporting 8](#_Toc27735220)

[Data Quality Summary Reporting 8](#_Toc27735221)

[Data Quality Discovery Reporting 9](#_Toc27735222)

[Record-level Detail Reporting – One measure, all records 9](#_Toc27735223)

[Record-level Detail Reporting – One record, all measures 10](#_Toc27735224)

# 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

# Overview

At the core of the reporting framework is the need to build a library of specific measures that compare a multitude of data items across multiple systems, underpinned by the wide-scale integration of data sources into a data warehouse or specific DQ reporting system.

This library of measures consists of datasets that identify every single row of data that has a data quality issue or a potential issue. There will be thousands of issues in reality, but they should only be added to the library because there is a business need to improve them for a judgment or improvement purpose.

The DQ measures in the library should be updated every day, and have a clear process to correct issues raised by the measure so that anyone can point at a number of measures and know exactly what needs to be done to fix them. The data quality measurement needs to be embedded into all other reporting from the data warehouse from aggregate measures at the highest level to the lowest level of detail, evidencing the data quality of board KPIs through to identifying which records in operational lists have data quality issues. At any point from the top to the bottom, the reporting system should intuitively direct report consumers to potential data quality risks with a single click and reaching the detail right from the highest level of reporting should take no more than three clicks – this will mean an executive can point at a KPI on the board report with a DQ issue and ask their team to click on the link to follow through to the problem.

Apart from the DQ standard mark, data quality reporting should only raise to attention by exception, when there is an issue to be resolved. Work lists should only link to DQ when the record has an error, scheduled automatic error alerts should be sent directly to users who have complicit responsibility, and scheduled automatic escalation communications should be sent up the chain of command for measures that have exceeded control tolerances.

The DQ Assurance Reporting Framework is the primary tool of the Data Assurance team. Its importance in transforming the data quality cannot be underestimated – without it the Data Assurance team will be severely limited in their potential impact.

This example data quality reporting specification proposes a full reporting suite, comprising 6 basic components:

* + Integration into the Trust Board report
  + Integration into Performance reports
  + A summary report  
    This report allows a overview of records that require validation and/or correction
  + A discovery report  
    This report allows exploration of where data quality issues may be particularly prevalent in order to focus and prioritise validation and re-training efforts where the greatest yield is likely to be achieved
  + A detail report for a single measure  
    This work flow based report provides the details for all records within a particular measure that require validation and/or correction. This is, ideally, a single report that works dynamically for all measures rather than a collection of multiple reports.
  + A detail report for a single record   
    This work flow based report provides the details for all measures that apply to a particular record that requirea validation and/or correction. This is, ideally, a single report that works dynamically for all measures rather than a collection of multiple reports.

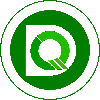
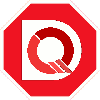
The DQ calculation process should execute on a daily basis, calculating incidence & prevalence (omitting snapshots for problems in between measures / audits at a lower frequency) for each item on the DQ measures list.

For incidence only measures (e.g. timeliness measures) prevalence will be reverse calculated. For prevalence measures without incidence data, incidence will be reverse calculated by comparing snapshots in the data warehouse or triangulating validation outcomes data against changes in prevalence.

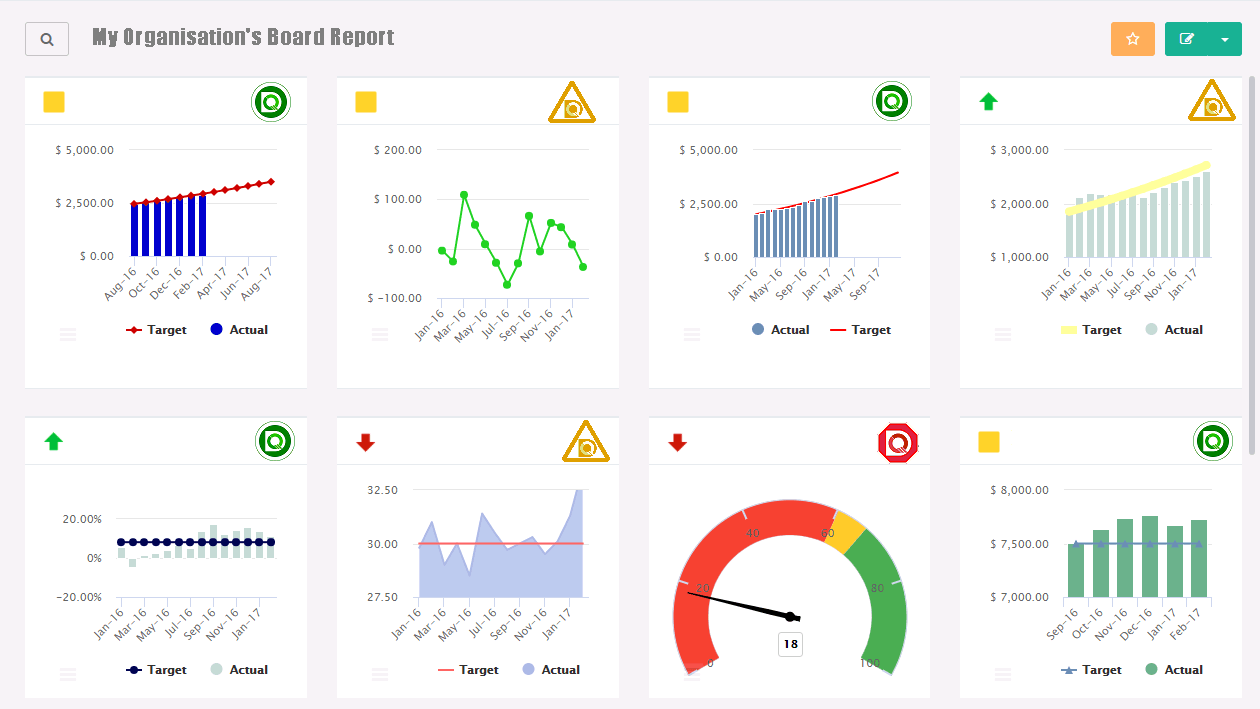
# Integration into the Trust Board Report

At the surface level, incorporation of data quality reporting into the board report looks like a set of DQ standard marks next to the existing KPIs. Clicking on any standard mark will take you to the list of measures that are used to assess the data quality of the KPI, a process known as drill-down. Clicking on any of those data quality measures will drill-down to the detailed records that have that data quality error, preferably with details of who made the error and when it was made.

Behind the scenes, each KPI will have a one-to-many relationship between DQ measures and that KPI measure. A weighting algorithm will calculate a score for each measure based on current prevalence, annualised incidence and an impact multiplier – this score will be used to decide whether the DQ standard for the KPI is red, amber or green and the appropriate DQ standard mark will be displayed accordingly.



These DQ standard marks should be clickable, taking the user through to the Data Quality Summary Reporting and pre-filtered to the set of measures that are implicated by the KPI that was clicked on.

*Example Board Report KPIs with embedded DQ*

# Integration into Performance Reporting

Like in the board report, data quality reporting should be incorporated in the performance review packs using the DQ standard mark against KPIs with drill-down through the list of associated DQ measures to the DQ detail.

In order to facilitate an explicit view of data quality, a DQ summary section should also be included. This will expose each DQ measure that impacts a KPI owned by the directorate. The following information will also be included for each measure:

* + • number of KPIs used by the measure and it’s total weighting
  + • current prevalence and incidence
  + • current escalation level
  + • stage of progress (see DQ management report for detail of stages of progress)
  + • domain
  + • point of failure

# Data Quality Summary Reporting

The summary for data quality reporting needs to contain the following:

* + the measure identifier
  + the measure name (ideally with a rollover popup showing the measure description)
  + the current absolute prevalence
  + the team responsible for correction

Additional items that are useful are:

* + a sparkline or indicator showing the current trend of the prevalence
  + the current absolute incidence
  + a sparkline or indicator showing the current trend of the incidence

Clicking on a measure within this report should take the user to the record-level detail report for one measure, all records.

# Data Quality Discovery Reporting

The discovery report is an aggregate report to intuitively filter and explore the data quality measures. It should use a technology that is instantly responsive (such as QlikView, Tableau or PowerPivot) and encourage the report reader to ask “where and when” questions about data quality. When they have discovered something of interest, report readers should be able to link across to the measure detail reporting to see the associated detail.

The report should show the:

* + absolute prevalence (errors at a snapshot in time)
  + relative prevalence (% errors of all records at a snapshot in time)
  + absolute incidence (errors created over a time period)
  + relative incidence (% errors of all records created over a time period)
  + weighted risk score

Broken down by:

* + domain
  + board KPI
  + DQ measure
  + escalation level
  + stage of progress (see DQ management report for detail of stages of progress)
  + point of failure
  + time (year, month, week, day, hour, minute, month of the year, day of the week, and hour of the day)

# Record-level Detail Reporting – One measure, all records

Reporting of the detail for a single measure needs to contain the following:

* + the name and description of the measure
  + a description of steps required to correct the data quality issue
  + all the data items required to correct the data quality issue
  + the user who made the error, or the last user to touch the record
  + the time the error was made, or the time the record was last updated
  + a list of all other DQ issues for the same record with a hyperlink to the “One record, all measures” report for that record, so the user can drill-across to the detail for all issues at once

# Record-level Detail Reporting – One record, all measures

In order to satisfy the “Correct it once, and once only” principle, reporting for a single record needs to show the detail for all associated DQ measures. Each DQ measure needs to contain the same detail as found in the “One measure, all records” report, filtered to the single record we are looking at. Drill across will link back to the “One measure, all records” report.