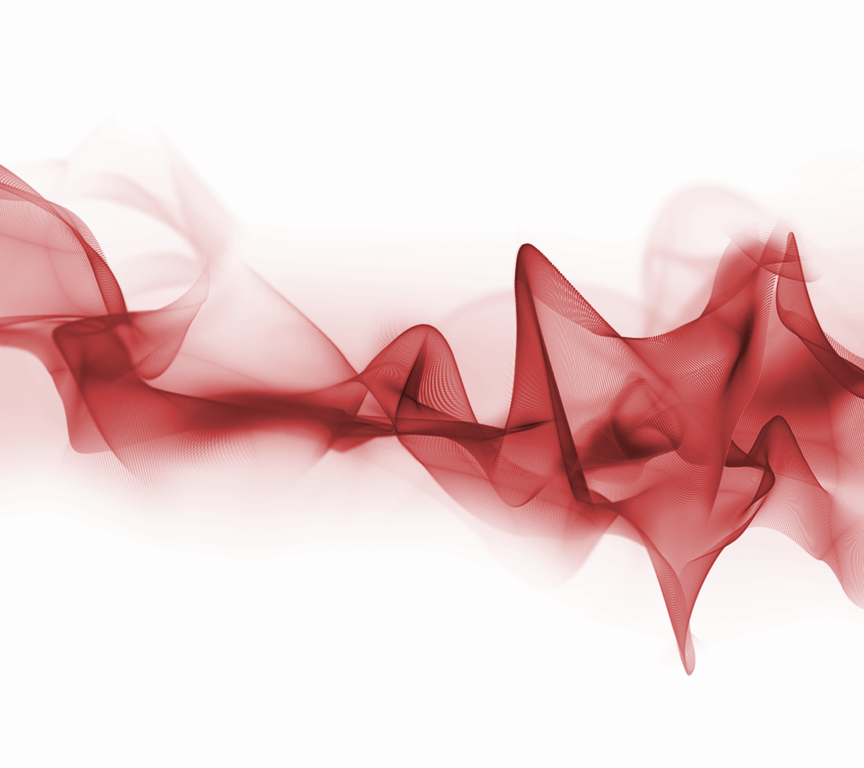
Data Quality & Data Warehouse Programme

Data Quality Strategy

“Better Data for Better Decisions”



© Perspicacity Ltd 2019

|  |  |
| --- | --- |
| Document Control | |
| Version | V2.0 |
| Created | 2019.01.12 |
| Updated | 2019.12.20 |
| Original Author | Matthew Bishop, Perspicacity Ltd & Jill McGregor |
| Contact | [matthew.bishop@perspicacityltd.co.uk](mailto:matthew.bishop@perspicacityltd.co.uk) / 07545 878906 |

© Perspicacity Ltd 2019

This work product is licensed under the creative commons attribution 4.0 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/4.0/legalcode>

It is part of an open source library which can be found at <https://github.com/perspicacity-ltd/DataQualityReporting>

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

**Table of Contents**

[Data Quality Strategy - Introduction 4](#_Toc8664467)

[Foreword 5](#_Toc8664468)

[Case for Change 6](#_Toc8664469)

[Strategic Objectives 8](#_Toc8664470)

[Governance 12](#_Toc8664471)

[Challenges 16](#_Toc8664472)

# 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

# Foreword

## The growing use of data

Since the 1970’s, the use of computers and data have played an ever increasing part in the running of an organisation. Computer systems have gone from bringing marginal benefits in specific areas of a business to defining and controlling the core of business processes and management decision making.

In boardrooms and management offices, data and information have radically changed decision processes from gut-feeling and time-consuming manual counts to automated reporting and complex analyses, increasingly delivered by machine learning or artificial intelligence. In the perpetual search for efficiency, the large opportunities come increasingly from complex processes facilitated by real-time process reporting or delivered using robotic process automation. These opportunities have a critical feature in common – their use of data and their reliance upon its quality.

Although improvements in engineering have provided improved hardware at cheaper costs, allowing high quality automated data gathering to become ubiquitously available, the quality of data capture will always fall prey to one particular weakness – that to err is human. Until we are able to install better hardware into the humans in our data processes, there will be a perpetual need to go back and correct the mistakes humans make in their data. Without this, the efficacy of machine learning, artificial intelligence or robotic process automation will be limited by the very people they are designed to serve.

## Death by 1000 cuts

This limitation is not new to the world of analytics and management reporting – many organisations have already invested, and there continues to be a growing trend in investment, into highly-skilled analytics capabilities to bring together intelligence from different systems and data sources. However, these analysts are also limited by the quality of the information they use. Their work nearly always comes after computer system users have established data capture processes, along with creeping normality of mistakes and workarounds that leaves data in a sub-optimal state of quality. It is this creeping normality that can render decisions misguided or, even worse, counterproductive.

This document sets out our aspiration for data quality within the organisation; to transform the governance and the strategic objectives surrounding how decisions are made. An accompanying framework document supports this strategy by describing the plan for delivery in terms of the data assurance team establishment, the training development and the technical requirements. Similarly, a supporting Data Quality Policy details the core principles that underpin good data quality, the roles and responsibilities and the governance structure.

Both I and the executive leadership team are committed to meeting the 6 core principles of data quality, and transforming the quality of our data through this strategy, to ensure that the intelligence we use to help shape our service is indeed of a standard suitable for a world class service.

# Case for Change

## Previous strategies

Previous iterations of data quality work within the organisation have laid the ground work for a Data Quality strategy. They have defined the challenge of ensuring the standards for data quality are met consistently and fully throughout the organisation and drafted an approach to delivering assurance.

## Need for clearly defined principles of good data quality, and for the roles, responsibilities and accountabilities to achieve them

This strategic visitation brings a clear and specific data quality policy which sets out how the Trust will practically put its strategy into practice. It includes clear definitions of the principles, roles and responsibilities for data quality.  
A data quality policy accompanies this document.

## Need for data quality assurance review framework

There is a need for a federated model of responsibility for data quality with a single point to record or monitor where the organisation gains assurance over all aspects of its data quality. Presenting that within the Board performance report, clearly evaluating the underlying data confidence for each of the Board KPIs, will encourage appropriate caution in the use of KPIs as measures of performance.  
In order to ensure an organisation wide oversight, an information assurance framework is required. This will comprise of, at least, a systematic review of data quality across all the significant data sources and a KPI review structured around the KPIs in the Board performance report. These reviews will document and record the level of assurance over the data quality underpinning the systems and KPIs, respectively. Inclusion of a RAG rating or a dial for data confidence next to each indicator in the performance report will support the decision making at Board. The data quality team will support prioritisation of systems and KPIs based on perceived risk to data quality.

Once errors have been identified, a formal mechanism will be defined to ensure that the identified errors are amended at source. For this to be an efficient process it will need a real-time data quality measurement system that identifies potential records for correction, with a clearly defined formal feedback mechanism to highlight errors discovered during reporting and validation processes.

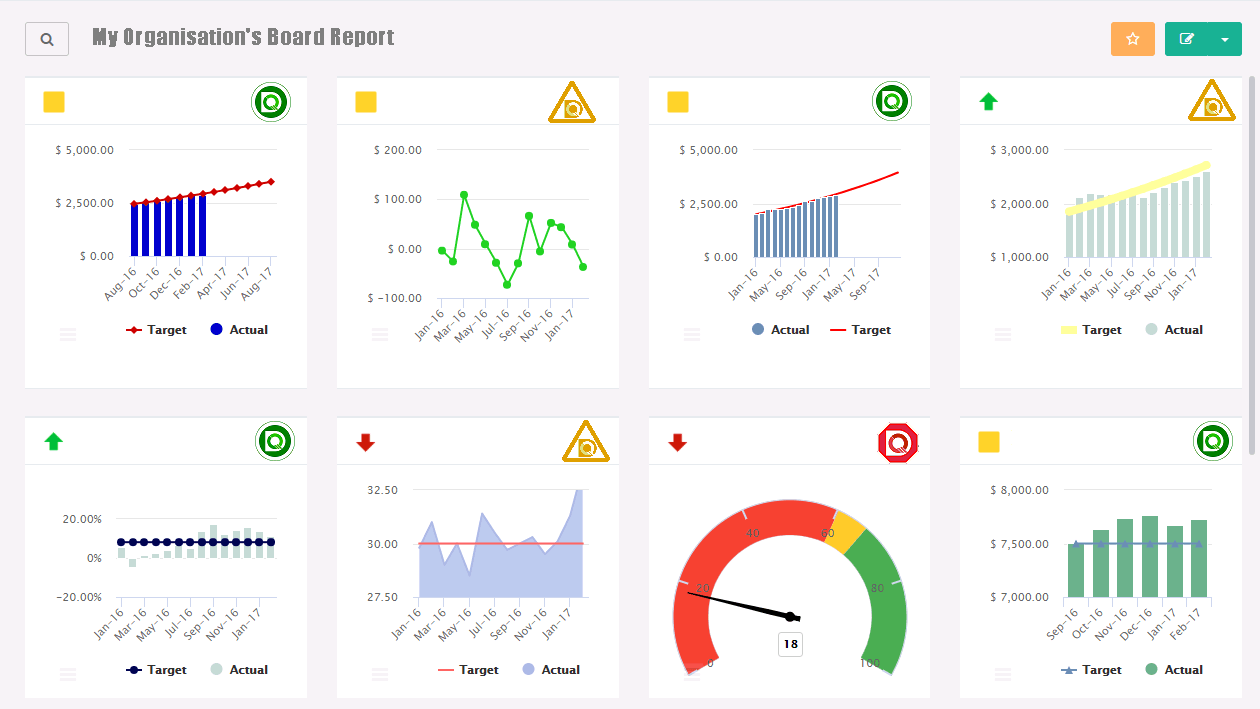
## Need to improve awareness of data quality and deliver data quality education and training

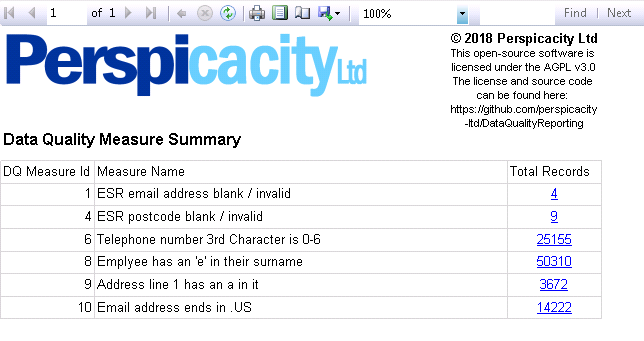
Core training and periodic refreshers are important to ensure that changes and updates in practice are communicated, especially for staff undertaking regular data entry; once a real-time data measurement system is established, the requirement for refresher training should be based on the number of errors generated in a period, rather than an elapsed period of time, to achieve greater efficiency in training effort. Completion of refresher training should be monitored and reported to divisional and team managers.

## Vision for data quality

The vision for data quality sees a data assurance team which instils a data quality mindset across the organisation, using purposely devised training and an assurance review framework which seamlessly connects the organisational decision making to the users who collect and correct data.

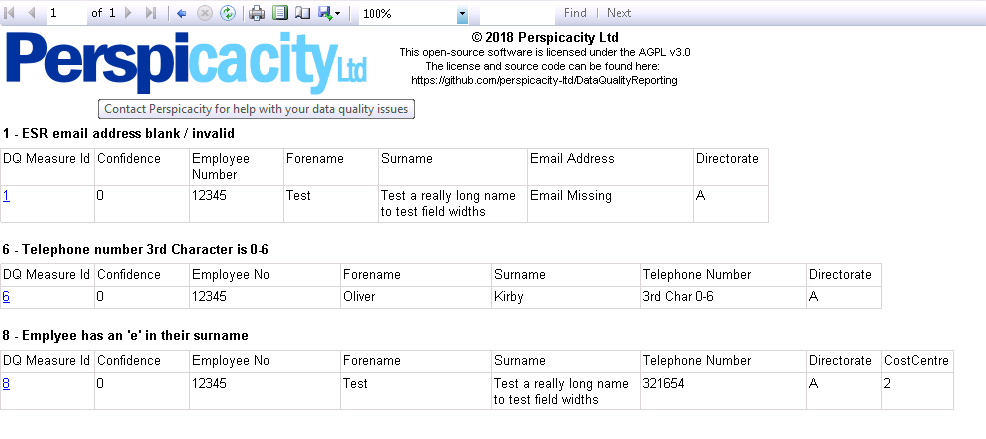
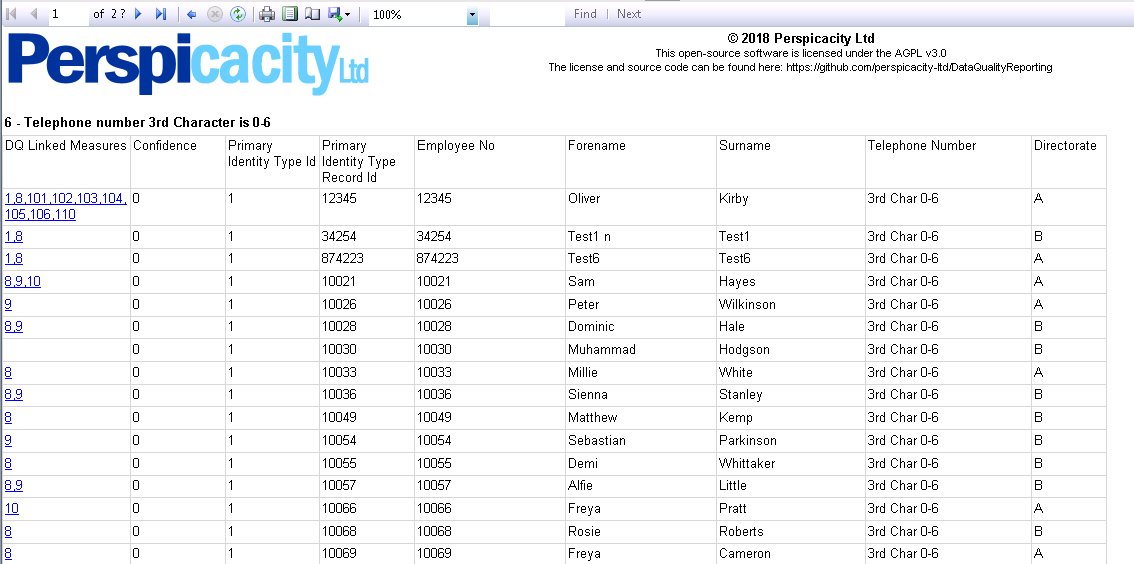
The organisation will work to deliver the critical reports that will surface and ubiquitously highlight the areas of data quality strength and weakness - the real-time assurance measurement system in the assurance review framework. The screenshots below demonstrate an example real-time assurance measurement system, freely available as open source software, ready for the data assurance team to support the organisation in its use and maintenance.





*The summary shows all data quality issues associated with the KPI selected in the Organisation’s Board report. Clicking on a data quality issue will take you to the details of all records with that issue*

*Clicking on the DQ standard mark in the Organisation’s Board report will take you to the Data Quality summary for that KPI*



*At first, all the records for a single issue are displayed. The detailed report shows all the necessary details to correct the record. If a record has multiple issues, the user can click on the record to see all issues for that record in one place. At the end of correction, the user can go back to the summary*

*Finally, all the issues for a single record are displayed. This ensures that a record is corrected once for all the issues, rather than multiple times by multiple people. The user can click to go back to the records for a single issue once correction is complete*

There is a need for a clearly defined formal feedback mechanism to highlight data capture errors discovered during the validation performed during this process.   
As the Trust expands towards a sufficient degree of measurement, and alongside the effective delivery of data quality education and training, this real-time assurance review will drive comprehensive board assurance and provide everyone with confidence in our data.

# Strategic Objectives

There are 7 strategic objectives to be delivered, each designed to address a part of the case for change.

The strategic objectives will be delivered according to the data quality assurance implementation plan which accompanies this document. The objectives will be delivered by 4 work streams focused on the assurance review, the data quality team, the education and training, and the technical development.

## Establish a clear Data Quality policy

The data quality policy will clearly define the principles of good data quality and the roles, responsibilities and accountabilities to achieve them. This policy will be owned by a board subcommittee and they will be responsible for reviewing and ratifying it. A draft data quality policy accompanies this document in readiness for review by the subcommittee and sign off by the board directors

## Establish a KPI Assurance Review Framework

The KPI assurance review will give information on the underlying data confidence for the KPIs which measure the Trusts performance. In the first instance this will be a manual low-tech solution which, although time consuming and less likely to be comprehensive and timely, will ensure the expedient delivery of the first reviews. Over the period of the strategy, this will be transitioned to a real-time assurance measurement system – a higher tech mechanism that will provide an up to date view of data quality for any KPI over a comprehensive range of potential data quality issues.

Details of how the KPI assurance review framework will be delivered can be found in the accompanying data quality assurance implementation plan.

### Complete and maintain an assurance review of DQ in all Board KPIs

This review will be led by the Executive Owner for data quality, who will be responsible for its delivery. It should be undertaken at least annually, but it may more frequent for particular KPIs whose underlying data quality may be changing – the KPIs that require this will be determined by the director of performance.  
The review will use a manual low-tech solution and an example spreadsheet solution accompanies this document.

As a part of the feedback mechanism for the assurance review framework, it is expected the output of this review will be put into the Organisation’s Board report.

### Real-Time Data Assurance Measurement for Board KPIs and Performance Reporting

The real-time data assurance measurement will provide a more frequent up to date view of data quality for any KPI over a much more comprehensive range of potential data quality issues than the manual review does.   
Initial delivery of the real-time assurance review is anticipated to follow a transition from the manual low-tech solution. Whilst the responsibility for this will remain with the Executive Owner, the delivery will be largely automated and the work will involve configuration of the mechanism.

## Establish a Systematic Assurance Review Framework

The systematic assurance review will give information on the underlying data confidence across all the Organisation’s information systems and data sources. Like the KPI assurance review, will be a manual low-tech solution which will be transitioned to a real-time assurance measurement system over the period of the strategy.

Details of how the systematic assurance review framework will be delivered can be found in the accompanying data quality assurance implementation plan.

### Complete and maintain an assurance review of DQ in every Trust information system

This review will be led by the CIO, who will be responsible for its delivery. It should be undertaken at least annually, but it may be more frequent for particular systems whose underlying data quality may be changing – the systems that require this will be determined by the CIO.  
The review will use a manual low-tech solution and it is anticipated that the example KPI spreadsheet solution from the internal audit will be used as a basis for the starting point.

As a part of the feedback mechanism for the assurance review framework, it is expected the output of this review will be used to determine which information systems require more robust training and processes.

### Real-Time Data Assurance Measurement for every Trust information system

Like with the KPIs, the real-time data assurance measurement will provide a more frequent up to date view of data quality for information systems over a much more comprehensive range of potential data quality issues than the manual review does.   
Initial delivery of the real-time assurance review is anticipated to follow a transition from the manual low-tech solution. Whilst the responsibility for this will remain with the Executive Owner, the delivery will be largely automated and the work will involve configuration of the mechanism.

## Ensure there is a feedback mechanism for the Assurance Review Framework

The feedback mechanism for the assurance review framework ensures that all information relating to data quality if found in, or linked to, a single point where the Trust can record and monitor assurance over all aspects of its data quality.

Details of how the feedback mechanism for the assurance review framework will be delivered can be found in the accompanying data quality assurance implementation plan.

### Implement data quality reporting in the Trust Board and Performance Reports

To support the decision making at Board, a RAG rating or a dial for data confidence should be placed next to each indicator in the performance report.  
Initially this will relate to the data confidence determined in the KPI assurance review. Over the period of the strategy this will be transitioned to a link to the real-time data assurance measurement so that clicking on the RAG rating will seamlessly direct the user to the details of records requiring remediation.

### Establish Highlight Reporting for the Board subcommittee

The highlight report will provide the subcommittee with sufficient information to allow accountability for delivery of the data quality strategy. It will provide an overview of the progress of delivery against the framework, how widely the data assurance team are representing and embedding data quality in key forums, and the efficacy of delivery against the data assurance team responsibilities in the data quality policy.

### Establish Data Quality Reporting for System Owners

This central reporting will provide the CIO with an overview of systems related data quality issues and system owners with the details of records requiring remediation. Whilst system owners will not be responsible for the remediation of data quality issues, as this is the responsibility of the data issue owners, the detailed reporting will highlight where information systems require more robust training and standard operation procedures or where modifications to the system are required to prohibit poor quality data capture.

### Establish Data Quality Reporting for User / Team Responsible for Validation

This central reporting will provide Board directors, accountable for remediation of data quality issues, with an overview of KPI related data quality issues. It will also provide data issue owners with the details of records requiring remediation. Data issue owners will be responsible for the validation and / or correcting of records and this reporting will constitute their remediation work list.

### Establish a Validation System and mechanism to capture amendments to data collected through manual systems

There is a need for a clearly defined formal feedback mechanism to highlight errors discovered during reporting and validation processes. This includes validation performed in response to the real-time data assurance measurement, validation performed as part of reporting processes, audit of records, and validation of manually corrected data.

This will require a system that can facilitate the requirements, ideally with a high-tech approach, taking its output from the real-time assurance measurement. In order to facilitate the capture of amendments to manually collected data, this will require the manual data to be loaded with the validation system capturing amended values.  
The dependency on the real-time assurance measurement constrains this to a subsequent part the strategy delivery timeline. The CIO will be accountable for the overall delivery, with the data assurance team responsible for the specification and the IM&T team responsible for delivery of the system.

## Establish the Data Assurance Team

The purpose of the data assurance team is to represent data quality within the organisation.   
This core team won’t cover all of the necessary validation and correction resources for remediation of data quality issues, many of which will require operational or corporate resource and some of which will require a specific business case for additional remediation resource.   
The core team will, however, provide the assurance function to assist the Executive Owner and CIO with delivery of the assurance review framework, start development of the real-time data assurance measurement and work with operational and corporate teams to deliver remediation of smaller data quality issues.

Details of how the data assurance team will be established can be found in the accompanying data quality assurance implementation plan.

### Represent and embed the policy in key forums

So that there is a structured approach to embed the policy into the way the organisation sees their information, the data assurance team will have targets to ensure appropriate representation across committees and performance reviews and for maintaining a physical presence within operational and corporate teams.  
This will ensure that decisions are constantly promoted to consider the context of the quality of the underlying information.

### Deliver the policy

The data assurance team will be principally responsible for delivering the data quality strategy into the way the organisation runs by supporting the organisation to use and maintain the tools provided in the assurance review framework. This will ensure the organisation is able to meet the principles, roles & responsibilities of the policy.

## Ensure training and standard operating procedures for information systems is fit for purpose

The IM System Owners are responsible for ensuring that their systems have the appropriate level of training materials to cover the DQ risks for the system as well as delivering training for the system. This will ensure a systems-focused approach where it is required.

Details of how the training and SOPs for information systems will be made fit for purpose can be found in the accompanying data quality assurance implementation plan.

### Review or Create Information Systems Training & SOPs

Every information system identified by the systematic assurance review should have a training package which is reviewed on a minimum annual frequency. Training should be provided by the IM system owner before access to the system is provided and at least 95% of users should have received training. A set of standard operating procedures, AKA “How-to guides”, should accompany the training.

### Training refreshers

Based upon the risk highlighted in the systematic assurance review, IM system owners will initially decide upon the frequency of periodic “refresher” re-training. Users identified as a “repeat offender” can be escalated for formal re-training by their line manager or by the data assurance team. As the real-time data assurance measurement system makes it possible to identify users who are making errors the IM system will move to error focused retraining, targeting only users who are making mistakes in order to make training effort more efficient.

## Ensure education for data quality is fit for purpose

The data assurance team are responsible for ensuring that all staff have an understanding of the principles of data quality that is appropriate to their role within the organisation.

### Training at induction of core principles

The induction training will cover the characteristics of data quality and data capture principles from the data quality policy. This should be delivered to all staff that use an information system to provide a basic understanding of data quality and how this understanding can help reduce the incidence of data quality issues.  
This training will cover a wide staff base who use different information systems as a part of their role, so versions will need to be created for different staff groups that contextualise the examples used in their training.

### Create Data Quality Training (Validation & Correction)

In addition to the core principles delivered in the induction training, the data quality training will also cover the DQ measurement principles and DQ correction principles to provide sufficient understanding for staff groups who are validating and correcting records with data quality issues.

### Create Data Assurance Team Training (Review, Assessment & Remediation)

To provide a full understanding of the data quality principles, the data assurance team training will also cover the DQ feedback principles, data assurance embedding and representation principles, and DQ in IM systems principles. This training will also cover the roles and responsibilities of the data assurance team from the data quality policy to ensure all staff are aware of their responsibilities to the organisation.

### Training refreshers

All staff will need a periodic “refresher” re-training of the induction training, with additional training made available for “repeat offenders” that have been escalated for formal re-training by their line manager or by the data assurance team. As the real-time data assurance measurement system makes it possible to identify users who are making errors, the periodic induction training will move to error focused retraining, targeting only users who are making mistakes in order to make training effort more efficient.

Training refreshers for the data quality training will also be delivered periodically or for “repeat offenders”, moving to error focused retraining as this becomes possible.

# Governance

## Governance structure

The structure of governance for Data Quality Assurance will place principal accountability for delivery of the data quality strategy with the assigned board subcommittee. Accountability for remediation of data quality will sit with individual directors, each taking ownership for their associated data issues. A core of the responsibilities will sit with the Executive Owner, the CIO, data issue owners, and the data assurance team.

The key roles within the structure are shown in the diagram below:

My Organisation’s Governance Structure

## Data Quality Policy – Core Accountabilities

### Board Directors

Individual Board directors will be accountable for remediation of Data Quality issues associated with their KPIs, as identified by the KPI assurance review.

The board will be responsible for sign-off of the data quality policy when it is presented by the Health Informatics Oversight Group.

### Executive Subcommittee

The subcommittee will be accountable for progress against the data quality strategy, as identified by the highlight report from the data assurance team. This report will detail progress in delivery of the assurance review framework, representation and embedding of the DQ policy, and delivery of the DQ policy.

The subcommittee will be responsible for review and ratification of the DQ policy, which will be presented to the Organisation’s Board for approval and sign-off.

These accountabilities and responsibilities are to be added to the terms of reference of the group.

## Data Quality Policy – Core Responsibilities

### Executive Owner

The Executive Owner will be responsible for delivery of the KPI assurance review. Progress against this will be reported in the highlight report from the data assurance team which is presented to the Health Informatics Oversight Group.

### CIO

The CIO will be responsible for delivery of systematic assurance review of data quality in all information systems and will also be responsible for ensuring the list of systems is kept up to date in order to ensure that new systems are assessed within the systematic assurance review of data quality. Progress against this will be reported in the highlight report from the data assurance team which is presented to the Health Informatics Oversight Group.

Where there is an overlap of accountability or responsibility with the requirements of the GDPR, the CIO will be responsible for aligning the appropriate policies and resolving any conflicts.

The CIO will be accountable for ensuring the DQ in IM Systems principles are upheld by the IM system owners and accountable for delivery of validation system.

### Data Quality Improvement Group

The monthly data quality improvement group meeting will be responsible for administering governance for data assurance at a local level. It will be a minuted meeting, chaired by the data assurance team. Terms of reference will be drafted in advance of the first meeting, where they will be reviewed and agreed. It will be closed to the public and the quorum will include the data assurance team, IM system owners and data issue owners. Invites will be extended by the data assurance team to all board directors that have measures in the process of investigation, quantification or remediation or for whom a business case or DQ improvement project is underway; this will allow for the stakeholders to represent their case for prioritisation of assurance resources.

### Data Issue Owners

Data issue owners are responsible for remediation of the data quality issues that have been assigned to them. There will be a data issue owner of these for each data quality issue that is discovered and the head of data assurance will be responsible for deciding who the appropriate data issue owners are, which may be an individual, a team or a group of people.   
This responsibility will not determined by delegation, but by which group of people have the power, expertise and resource to most appropriately remedy the data quality issue.

### Data Assurance Team

The data assurance team and the head of data assurance will hold the most wide ranging responsibilities for data quality in the organisation. These responsibilities will be detailed in the data quality policy.

## Data Quality Policy – Core Principles

### 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.   
Drawing upon expertise from the Audit Commission, the organisation has chosen to adopt their six characteristics of data quality.

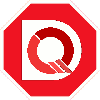
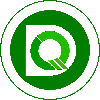


By making our core focus to achieve the 6 core characteristics, we are seeking to ensure our data is recorded in sufficient detail; correctly recorded to reflect what it designed to record; consistent over time; recorded close to the event it represents, to avoid the degradation of validity; designed to serve a relevant purpose; and available when we need to use it.

### Data quality feedback principles

Feedback is critical in data quality to inform us where errors have occurred, where they impact upon our information for decision making, and allow us to report back errors we have discovered. The following principles will facilitate the necessary mechanisms to facilitate effective feedback. A description of each principle is detailed in the appendices.

* Integrate data quality into the Trust Board and performance reporting using the DQ standard marks below, rather than having separate DQ KPIs



* Ensure there is a summary report of data quality issues by size, board KPI and Data Issue Owner
* Ensure there is a dynamic analytical report for discovery of where issues may be happening more frequently
* Ensure there is detailed reporting to show all the records for each DQ issue and all the issues for each DQ record
* Ensure there is a mechanism for users to capture the outcomes of data validation for a record

## Data Quality Policy – Other Responsibilities and Principles

There are a further 5 principles of data quality and numerous other roles and responsibilities. These are detailed in the accompanying data quality policy.

# Challenges

As noted in the strategic objectives, there will be limited capacity within the data assurance team for correction activity. This amount of resource will be insufficient to tackle some of the larger data quality issues within the organisation and it is likely there will always be a requirement for operational units to shoulder the resource implications of data correction within their own resource or to succeed a business case to deliver the resource needs.  
Whilst there will be a large, and at times daunting, amount of work to improve the organisation’s data quality, maintaining the structured and stable approach in this strategy will lead the organisation towards meeting the 6 core principles of data quality and transforming the quality of our data to ensure that the intelligence used to help shape our service is of a standard suitable for a world class service.