the information authority
setting data standards for further education

the information authority - Policy Guide

Data Quality Framework

Author:	Adele Dixon Data Quality and Improvement Manager *the information authority*
Date:	14 January 2009
Version:	1.6
Approver:	*information authority* Board

Table of Contents

1. Introduction

Good quality data is crucial for the effective use of performance measures and targets within the FE Sector. Good quality data will help departments to: manage delivery against priorities; assess whether they need to revise policies and programmes; and report reliably on their achievements.

A data quality framework is intended to provide a common objective approach to the assessment of management and improvement of data quality. The approach is aligned and consistent with the business user expectations for quality information and will be accomplished by using a range of methods and techniques that assess the quality of the data and then through the application of consistent embedded work practices. This will also identify the data quality priorities and encourage continuous improvement in data quality.

One of the key deliverables for *the information authority* secretariat is to provide and pursue 'best practice' for the improvement of data quality within the FE sector. The best practice includes providing Information advice and guidance and the documentation, development and implementation of the:

- Data quality governance guidelines
- Data standards framework
- Data quality policies, processes and procedures
- Data quality roles and responsibilities

In order to achieve the assurance of high quality data, the data quality framework will describe the identification, documentation and validation of the relevant data quality measurement criteria. These measurement criteria in turn will develop and be transformed into quality rules, benchmarks and metrics, which are used to:

- assess the business impact of poor quality data
- develop performance models to gauge the severity of data quality issues
- provide ongoing data quality measurement
- monitor the conformance with FE sector expectations
- assess the impact of proposed process changes

2. Document Scope

The scope of this document is to govern the FE sector in establishing a robust data quality framework to underpin the quality of data and information in the FE Systems. This includes but is not exclusive to learner, workforce (staff), financial, information advice and guidance, exam entries and achievement data and information.

3. Document Purpose

The purpose of this document is to act as a framework to improve data quality through the descriptions of components/techniques and processes. The document will also explain how these components are used within the data quality investigation, assessment and management process.

At the outset the framework is an attempt to organise, assess and incorporate all activities related to data quality and therefore by intention is a living document that is subject to change control procedures and a regular review process.

Upon completion of review, the document will be forwarded to the information authority board for formal acceptance and approval.

4. References

4.1 Information Authority Policy guide Data Governance Framework version 2 (work in progress)
Documents can also be provided by *The information authority* project office

5. Goals and Objectives

The aim of the data quality framework is to achieve agreement and acceptance for a common approach to the process of the data quality improvement.

6. Overview

The data quality framework has a number of critical components that collectively form the Best Practice and recommendations for the ongoing development of data quality best practice.
These include:

- Data quality dimensions
- Roles and responsibilities
- Business processes
- Data quality metrics

The components are described in more detail in the remainder of the document.

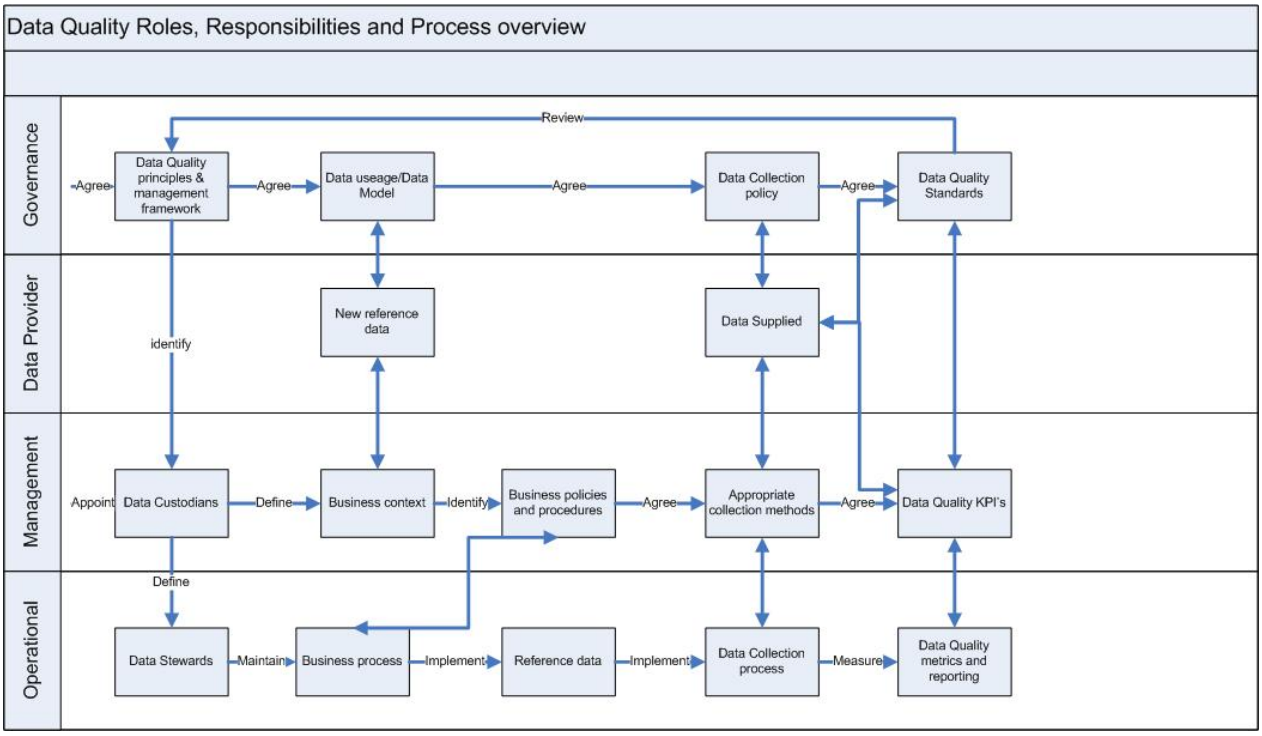
7. Dimensions of Data Quality

As data quality improvement is more about process, people and governance than technology, it is important that a simple model is used which makes sense to both the business and IT teams. The following provides a simple data quality model that is understandable and where used, can trigger significant beneficial change.

Dimension	Principles	Question\Criteria
Accuracy	Data should represent what actually happened in the real world. It should do so sufficiently closely to meet the reasons for which it is collected and maintained.	How vital is the data and is it accurate enough for the intended purpose?
Completeness	Data should be complete – should not contain extra, invalid or missing data	Does this need to be a required/ mandatory data field?
Validity	Data should be recorded and maintained in compliance with relevant rules and definitions. It also includes what in FE is described as credibility.	What rules will be applied to ensure the data conforms to national or local standards?
Reliability	Data should be collected and processed consistently	How can we consistently enforce the same standards for recording each data item?
Timeliness	Data collection should be captured as quickly as necessary to support processes for which it is collected.	Do we need to specify how often the data should be collected?
Relevance	Data captured should be relevant to the purpose for which it is used	Do we still need to collect this data entity/attribute?

8. Data Quality Roles and responsibilities

To ensure that the management of performance is coherent, the following roles and responsibilities are operated:



8.1 Data Quality Governance

The data quality framework is designed to support collaborative working between data custodians, data stewards and managers. In an environment where these groups work collaboratively, it will assist in:

- Achieving clarity in the specification of goals and objectives
- Ensuring value is achieved from the governance process
- Creating a clear mission for the process
- Maintaining scope and focus
- Establishing effective accountability
- Defining criteria for success and subsequent communication

The data quality framework and associated processes describe and illustrate.

- who is responsible for decisions or actions
- what is in scope of data and information
- why governance is necessary
- when and under what circumstances decisions or actions are required
- how remedy and resolution methods are applied

8.2 Data Custodians

Data custodians are a person or a recognisable board from a range of organisational groups who are ultimately responsible for setting the rules and requirements for data, and managing the dataset and information risk associated with its storage and use.

8.3 Data Stewards

Data stewards are those individuals ultimately responsible for the definition, management, control, integrity or maintenance of data as a resource within the FE system. They may have differing roles but share a common set of objectives:

- Accountable for the data under their stewardship
- Tasked with improving the quality and value of the data
- Are a ‘subject matter expert’ for that area of data

8.4 Data Collection

The purpose of the data collection policy and collection methods is to monitor report and improve the quality of data captured, collected, transformed, and disseminated by the Data Service

8.5 Data Provider

In this context, the 'Data Provider' is an organisation within the FE sector that provides post-16 education and/or training, such as:

FE Colleges
School Sixth Forms
Employers providing work based learning
Adult Community learning centres
Performance academies
Specialist colleges
Local Authorities
Ufi

Or, an organisation, governing body or board that maintains or provides new reference data such as:

The Data Service
MIAP

9 Business Processes

Good quality data leads to accurate success rates; correct funding and the ability to track progress accurately against key performance measures and government targets.

Poor quality data leads to inconsistencies in funding, planning and strategic decision-making. The potential consequences of poor quality data could range from the unequal distribution of learning opportunities across the country to individual learners being denied the access to the correct level of funding for training opportunities. This is not, the desirable outcome either for Learners, the FE Sector, Local or Central Government or ultimately the Taxpayer.

9.1 Managing Data Quality

Managing Data Quality is about creating a culture of commitment to improving the quality of data. This commitment includes governance, policy, process, training and monitoring.

The information authority's Data Quality Framework sets out five areas of activity:

1. The governance of data quality in terms of leadership arrangements and standards required to drive quality improvements.

2. Ensure polices and procedures are in place to ensure accurate data recording and reporting.

3. Ensure systems and processes in place to secure data quality within day-to-day business activity

4. Support in the creation of a development plan and material to ensure the people, skills and knowledge allow effective planning to achieve these objectives.

5. Monitor and controls in place to ensure the appropriate and accurate use of data for decision making and reporting.

9.2 Clearly Understood Standards

In order to be able to produce quality data, it is imperative that there is a sound understanding of what is expected by the data provider. Clearly understood standards supported by guidance and assistance will ensure that all staff involved in the data quality process adheres to the appropriate framework.

Formal control procedures will be established based on already identified good practices to ensure that there is a documented set of guidelines, roles and responsibilities. The Data stewards will consult on a regular basis on how best these procedures can be evaluated.

9.3 Creation of a Quality Culture

The key aim is to embed a culture of data quality as part of the business as usual processes across the sector.

There is a need to clearly define roles and responsibilities across the sector for data quality with ‘Data Custodians’ having overall strategic responsibility.

9.4 Skills

One of the main objectives is to identify the skills that all data stewards and data providers need to undertake for their work and ensure they have the knowledge, competencies and training material available to ensure they are able to carry out relevant duties to deliver a high standard of data quality.

9.5 Process Improvement

It is important to encourage mechanisms to continuously improve data quality. A delivery plan associated with the Data Stewards will be used to drive improvement within the FE sector. It will contain clearly identified actions, responsibilities and timescales and be measured and progress reported to the ‘Data Custodians’.

10 Defining Data Quality Requirement Metrics (DQRM)

Data quality standards will not always be 100% achievable in large data sets and there will need to be a level of tolerance. The Data stewards will engage the Data custodians to agree the following objectives:

- Business relevance: The metric must be defined within a business context that explains how the metric score correlates to improve the business performance.
- Measurability: The metric must have a process that quantifies a measurement within a discrete range
- Controllability: The metric must reflect a controllable aspect of the business process; when the measurement is not in a desirable range, some action to improve the data must be triggered
- Reportability: The metric's definition should provide the right level of information to the data steward when the measured value is not acceptable
- Trackability: Documenting a time series of reported measurements must provide insight into the result of improvement efforts over time as well as support statistical process control.

These objectives will continually be assessed and prioritised and act as the benchmark for the improvement in data quality.

10.1 Assessment

There are a number of considerations that should be taken into account in determining the targets for each field, object, record and record set.

- Analysis of historic data quality achievement against new DQRM;
- Quality requirements, risks and priorities from the DQRM;
- Cost benefit;
- Anticipated likelihood of quality achievement;
- Potential constraints to achievement of quality;
- Use of incremental targets;
- Requirement – fit for purpose; and
- Usefulness / validity of the measure.

The capability to monitor and measure results should lead toward the establishment of Return on Investment for the program and methods include the use of KPI dashboards and frequent management reports.

10.2 Measurements

In the current environment, it is not clear what we would use to measure success of data quality. Documented below are a number of alternatives and common approaches used.

10.2.1 Dashboards

Dashboards are a way of displaying measurement results against explicit business-defined thresholds. A dashboard can display individual results as well as measures derived from more than one individual result.

There are four types of analysis which dashboards are used to display:

- Trending analysis – How have the measurement results changed over time?
- Consistency analysis – Is the current quality of data satisfactory?
- Timeliness analysis – When was the data updated?
- Completeness analysis – Is the data available?

Example of the type of data available is shown below.

Monthly Potential Duplicate Resolution of ULN (end of month)

Central Register	Dec	Jan	Feb	Mar	Apr	May	June	July
Resolved In Month	0	546	2149	5885	3224	2393	624	3468
Cumulative Resolved	0	0	1364	8580	11804	14197	14821	18289
Outstanding	10873	12516	12123	7284	7658	7350	9823	8772
Target to be Resolved	0	1479	874	2140	3239	1877	2809	2154
% Target (1)		37%	246%	275%	99%	127%	22%	161%

Dashboards may differ from typical daily status reports in two main ways:

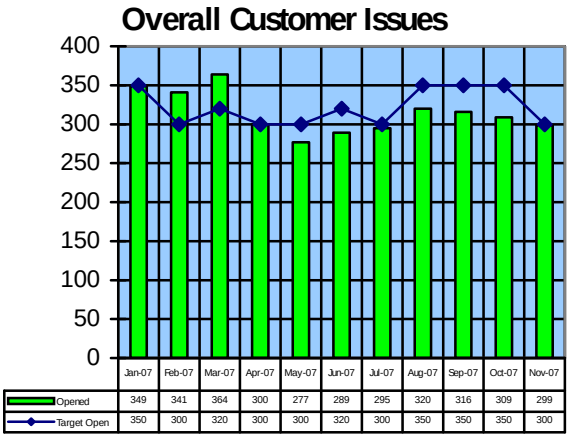
- Includes historic results in order to display trend
- May display a score or category for the result. (Example RAG status)

10.2.2 Key performance indicators

In selecting data quality key performance indicators (KPIs), it is critical to limit them to those factors supporting the Data stewards reach their goals. It is also important to keep the number small to keep everyone's attention focused on achieving the same KPIs.

Once good key performance indicators are defined that reflect the Data stewards goals and objectives that are measurable, this can be used as a powerful management tool. This gives all custodians a clear picture of what is important and what they need to do to make it happen. KPIs can be used to manage performance and also focus individuals on meeting and exceeding expectations. Posting the KPIs via a dashboard will show interested parties what the target for each one and show the progress towards the target. This will also motivate people towards reaching the set targets.

Shown in the diagram below is an example performance graph: 'target against actual' of the number of support calls opened on a busy service desk.



10.2.3 Activity based cost model

Activity Based Costing (ABC) is an internationally recognized system in which staff timings and costs are produced by breaking down each activity into its constituent parts to determine the work effort required.

This enables staff costs to be attributed to all activities within an organisation and provide a measure of the amount of resource used to undertake a specific element of Data Quality work.

This in turn provides invaluable data to support any planning, monitoring and control system, and will enable key custodians to make better and more informed business decisions.

The implementation of ABC can help the understanding of the various data quality costs involved. Although this may be viewed as costly and impact resources in the short term, the long term gains will then enable effective planning of resource, the ability analyse the cost, identify the activities that add value and those which do not add value. Finally, based on this, this will give the ability to implement improvements and the benefits can be realised. This is a continuous improvement in terms of analysing the cost; to reduce or eliminate the non-value added activities, to effectively plan resources and to achieve an overall efficiency in data quality improvement.

11 Conclusions

The Data Quality Framework is more about process, people and governance than the use of technology to investigate and resolve data quality problems. The most critical aspect to the framework is the promotion of collaborative working and the use of the most effective methodologies. The *information authority* will pursue a proactive approach where the secretariat commits to the framework mechanism and communicates the strategy and manages the process of introduction to the FE sector

Equally, the Data Quality Framework must be seen as part of an overall Data and Information Management process that includes sound Governance processes and the ongoing development of Data Standards. The Data Standards Framework will develop the working environment, methods and processes to enable the improvements to be realised. This will include the following components:

- Standards and Data Quality Governance
- Standards Metadata
- Master & Reference Data Management
- Data Quality Framework
- Data Stakeholder, Custodians & Data Stewardship
- Data Modelling

Finally, the Framework is the starting point for developing a data quality process that has continuous improvement at the centre of operation and includes both the changes necessary to modify work behavior and also the technology deployed.

12 Recommendations

The recommendations for the ‘Data Quality Framework’ fall in to the following categories:

- Develop Data Quality Process and Methodology
- Identify and define priorities
- Problem determination, investigation and resolution
- Report and measurement