DATA QUALITY FRAMEWORK (DQF)

GRIFFITH UNIVERSITY

Prepared by Catherine Delahunty and Wendy Marchment, QPS
# Table of Contents

1 Introduction ........................................................................................................................................... 3  
2 Organisational Structure ..................................................................................................................... 4  
3 Data Quality Roles ............................................................................................................................... 4  
4 Raising Issues ..................................................................................................................................... 5  
5 Rectifying Issues .................................................................................................................................. 5  
6 Communication Tools .......................................................................................................................... 5  
7 The Way Forward .................................................................................................................................. 5  

Attachment 1 – Roles and Definitions ...................................................................................................... 6  
  Data Quality Programme Sponsor: ........................................................................................................ 6  
  Data Owner / Custodian: ...................................................................................................................... 6  
  Data Steward: ....................................................................................................................................... 6  
  Other Roles ........................................................................................................................................... 6  
  Data Quality Analyst: ........................................................................................................................... 6  
  Data Consumers: ................................................................................................................................. 7  
  Data provider: ....................................................................................................................................... 7  

Attachment 2 – Rectification Process ....................................................................................................... 8  

Attachment 3 – Communication Tools ..................................................................................................... 10  

INFORMATION SERVICES
1 Introduction

The objectives of business intelligence applications, such as a Data Warehouse, are to provide useful, accurate, relevant information for management decision making processes by integrating raw, unconnected data from many operational sources. As the Data Warehouse (DW) continues to expand and add products to its suite, attention will inevitably be drawn to the quality of the data. Recognising the critical role data quality plays in the University, the challenge now is to identify the best practices for managing data quality issues.

It is acknowledged that a variety of data quality issues are already addressed at the University, but that this is generally done in relation to external compliance reporting, mostly with uncoordinated approach, with varying degrees of attention by areas and without an overview of data quality issues that affect more than one area or that may affect management reporting.

The notion of data quality refers to the data's ability to conform to requirements or be fit for use. Thus, data might be considered of high quality for one purpose, but of very poor quality for another. Typically, data quality attributes and objectives include:

- **Accuracy** – Data items may be valid but not necessarily accurate. Often it’s necessary to cross-check against other data items to ensure the data is accurate.
- **Timely** – Data items are available for reporting at (agreed) critical times during the processing cycle or at agreed snapshot dates.
- **Relevant** – Data items add some agreed value to the understanding of the aspect of the business activity being reported.
- **Standardised** – Where data items are measuring or categorising the same attributes of real world objects, the same values are used to code or measure them, across all systems in the organisation.
- **Comparability** – The same codes have the same meanings in all systems in which they are used.

Ideally, data quality should be viewed as part of a Business Intelligence strategy which establishes disciplined processes for managing information assets. Such processes should be part of a broader programme for Enterprise Information Management. Data quality processes have organisational implication which much be articulated, understood and practiced on a day to day basis.

Research undertaken by Gartner indicates:

- Virtually every organisation has data quality issues.
- Most organisations do not demonstrate a cultural pattern of managing data as an asset.
- Data quality is often overlooked until problems arise.
- Lack of data quality is often described as an IT problem – not a business problem.
- Data quality should never be considered an IT issue, the owner of the problem is the business.

In short, a methodology for data quality improvement must address the following components:
1. **Strategy and Vision** - identify how data quality improvements support University goals and objectives.

2. **People and Skills** - promote cultural and organisational change to build awareness, understanding, ownership and engagement of key stakeholders.


4. **Technology** - implement data quality analysis, monitoring, controls and cleansing.

In order to achieve these aims, it is proposed that the following framework be endorsed in principle, for further development and implementation as part of a Business Intelligence strategy. Key to the success of the Data Quality Framework is the notion that it will be supported by a structure which encompasses staff from each relevant area within the University.

## 2 Organisational Structure

Data quality needs to be supported by an organisational structure that encompasses people from virtually every major area in the university who are considered subject matter experts in their field. The primary vehicle for rectifying data quality issues will be through the Data Steward group who will meet on a monthly basis.

## 3 Data Quality Roles

The data quality initiative will require new job roles to be defined. Primary roles will be:

1. Data Quality Programme Sponsor - the champion of the DQF and ideally part of the university executive and the Business Intelligence Governance group.

2. Data Owner / Custodian - individuals or groups within the university in the position to obtain, recreate and have significant control over the content of corporate data. Typically they will belong to a business area.

3. Data Steward - the interface between the DQF, the IT department, the BI team and their original functional area. Data stewards work as the subject matter experts.

For a complete list of roles and definitions, see Attachments 1.
4 Raising Issues

Initially, the Data Warehouse team will be the primary group raising data quality issues as they investigate source systems for Data Warehouse segment implementation. However as use of the Data Warehouse extends further into the University community and there is a change to managing data as an asset, additional data quality issues will be identified and raised. When a data quality issues is identified, it will be logged via the Data Quality Web Site which will reside in Quality, Planning & Statistics.

5 Rectifying Issues

Once an issue is identified, it will be assigned to the appropriate Data Steward who will conduct preliminary investigation and present the issue at the next monthly DQ meeting if it requires discussion, or simpler issues may be resolved via phone/email.

For a more detailed description of the process, refer to Attachment 2.

6 Communication Tools

To assist in the identification and rectification of data quality issues, tools will be developed which will communicate the processes and status of data quality issues to the university community. These will include:

- Data Quality Web Site
- Data Quality Issue Rectification Process
- Data Quality Issues List
- Data Steward Group Contact List

Refer to Attachment 3 for a description of the tools.

7 The Way Forward

Once an in principal agreement is reached with respect to the proposed model of the Data Quality Framework, a more detailed document will be produced which will address the issues of:

- Data Governance
- Principles and Policy
- Processes and Procedures
Attachment 1 – Roles and Definitions

Data Quality Programme Sponsor:

Nature: The champion. Ideally part of the executive.

Role: Identifies the processes, priorities and business requirements and is also responsible for data quality communication of issues and standards to managers.

Data Owner / Custodian:

Nature: Individuals or groups within the university in the position to obtain, create and have significant control over the content of corporate data. Typically belong to a business area.

Role: Accountable for data quality. Sets data standards and rules as well as monitoring implementation. The owner develops consistent organisational structures and is responsible for the integration of business systems.

Data Steward:

Nature: The interface between the data quality programme, the IT department, the Business Intelligence team and their original functional area. Data Stewards work as the subject matter experts.

Role: Includes creating and maintaining business rules, identifying data quality issues through data profiling, monitoring quality levels and collaborating with INS to enhance appropriate architecture components to improve data quality.

In order to carry out this responsibility, Data Stewards need strong authority to make decisions with ratification from the Exec Group. While the Stewardship group carry out the work and retain responsibility for data quality improvement, the Information Governance Group / Exec retain oversight, ensuring equity of policy application and overall responsibility for success of the data Quality Framework as a process.

Other Roles

Data Quality Analyst:

Primarily an IT function. Their main role is to use profiling, cleansing and de-duplication tools in collaboration with the data stewards to make sure the stored information adheres to the rules in the respective business context.
Data Consumers:

An authorised user that has been granted access rights to some data within the enterprise. In the case of open information the data consumer is considered to be the general public.

Data provider:

An accepted supplier of information into the system. This includes external providers such as DEST and GCA.
Attachment 2 – Rectification Process

1. A data quality issue is identified during the implementation of a specific Data Warehouse segment or past segments. The Data Warehouse Team will be the usual body that identifies the issue, however, it is not limited to this team. It will be possible for anyone to identify a data quality issue and this process would be used to resolve the issue.

2. These issues will be raised at a monthly meeting of all Data Stewards. The purpose of this meeting is documented further in this document “Data Stewards Group Objective”.

3. Each issue will be assigned to a Data Steward for root cause investigation, evaluation of possible solutions and proposing a solution. Each Data Steward will call upon Data Custodians, Data Experts and other members of the University community (such as INS) to assist in this step.

4. At times, a quality issue identified will be outside the responsibilities of the assigned Steward (university wide issue, multiple computer systems effected, etc). In these instances it may take substantially more resources (time, other department employees, money) to identify the root cause and propose a solution.

5. This quality issue investigation will need to be prioritised with the other tasks of the Data Steward and other staff members' workloads in line with ‘what’s best’ for the University. This priority setting will need to occur at the Data Custodian or Business Intelligence Steering Committee to obtain the required focus, and backing, on investigating and solving the quality issue.

6. If the proposed solution does not require a policy change or senior management approval, the Data Custodian of the effected data can approve the proposed change. These issues will be consolidated into a report presented to the Business Intelligence Steering Committee.

7. If the proposed solution requires broad changes in University policy and procedures or senior management approval, it must be presented to the Business Intelligence Steering Committee for evaluation including resource and cost estimates to implement the proposed solution.

8. If the proposed solution needs approval/consultation by other University committee’s, the Data Custodian will present the proposed solution as needed.

9. If approved by the Business Intelligence Steering Committee or the Data Custodian, the Data Steward will be responsible for the implementation of the solution to resolve the data quality issue.

10. The Data Steward is responsible for reporting back to the Data Stewards Group on the progress of their assigned data quality issues.

11. The Data Quality website will be updated on a monthly basis to add new data quality issues and to reflect any data quality issue progress changes. The Data Stewards Group Convenor will perform this function. The Data Stewards Group Convenor will also communicate all new quality issues to all Data Custodians in a summary report after each monthly meeting.

12. The Data Stewards Group will notify the identifier of the data quality issue, of the issue resolution.
13. The issue identifier will test and sign-off on the resolution performed.
Attachment 3 – Communication Tools

The Data Quality Web Site will provide a single point of reference for all data quality information relating to the Data Warehouse implementation. The web site will be located within the Data Warehouse area of the QPS web site and will provide updated information to the University community on data quality issues identified, rectification procedure taken and any related documents in relation to the process followed.

The Data Quality Issue Rectification Process will be available on the Data Quality web site as a supporting document. It will detail the process followed in the rectification of any data quality issue identified.

Data Quality Issue List will be located on the Data Quality Web site to display the current status of identified data quality issues. The list will detail data regarding each data quality issues (including date raised, person responsible, status, area affected, etc) as well as including progress stages that has occurred in the rectification of the data quality issue. This list will be updated on a monthly basis by the Data Stewards group convenor.

Data Steward Group Contact List will be located on the Data Quality Web site and will list contact details of all members forming the Data Stewards Group. It will also indicate the data area each Data Steward is responsible for.