Developing a Framework to Manage Data Quality in Healthcare

(An actual case study from a major Healthcare Insurance Carrier)

Denis Kosar, Director Data Architecture
Prudential Healthcare, a member company of Aetna US Healthcare
for

MIT Total Data Quality Management Conference October 22-24 1999

1. Purpose of this Paper

The purpose of this paper is to provide the reader with an understanding of the concept of "data quality" and how it can actually impact an organization's ability to perform efficiently and ultimately sustain itself. It will also explore some of the ways that data quality can be dealt with tactically, on a short term basis. In addition, it will identify how we can become more proactive by establishing a Data Quality Program within an organization using a framework consisting of a methodology, tools and standard operating procedures. To accomplish this, a major healthcare insurance carrier will be used. We will look at some of the data quality problems identified and explore what steps were instituted by the healthcare insurance provider to deal with them both tactically and strategically.

What is Data Quality anyway? Some people say Data Quality is "Fitness for intended use". Some vendors describe it in terms of name and address cleansing? Most of us can identify flawed data if we had to, but would you recognize good data if you saw it? What are some of the criteria that makes it good and what makes it bad? How much does poor data quality cost the organization? What can organizations do to deal with poor data quality? Are there tools that can help? Is there a practical way in dealing with it or is it too costly? These are some of the questions that this paper will answer.

2. Case Study Environment

The Health care insurance provider at the time this paper was written was the fifth largest provider of managed care services in America. It provided health insurance service to twelve million American. The products and services that are provided to their clients are built upon information. Client information is one of their most valuable resources. The concept of Information Resource Management is the recognition that information is a corporate asset and must be valued as such. To that end, the mission of their Data Management group was to manage and protect the information assets of the enterprise. To support this mission, the following goals have been defined:

- Promote the importance of information as valuable resources which require the management of the creation, use, storage and delivery.

- Promote data consistency and standardization throughout Prudential by deploying standard operating procedures concerning the definition and use of information.

- Provide for the safe keeping of corporate information and exclusion of un-authorized access through the deployment of security procedures.

- Provide for the data quality of all corporate information through the use of business and standard edit and validation rules.

To better support our clients; provider and client data was sourced from the operational systems and stored in their healthcare data warehouse. In addition, an Underwriter Data Mart was developed to allow underwriters to analyze client

claim data and determine the renewal premium for the next years coverage. This environment can be viewed in more detail in figure 1.

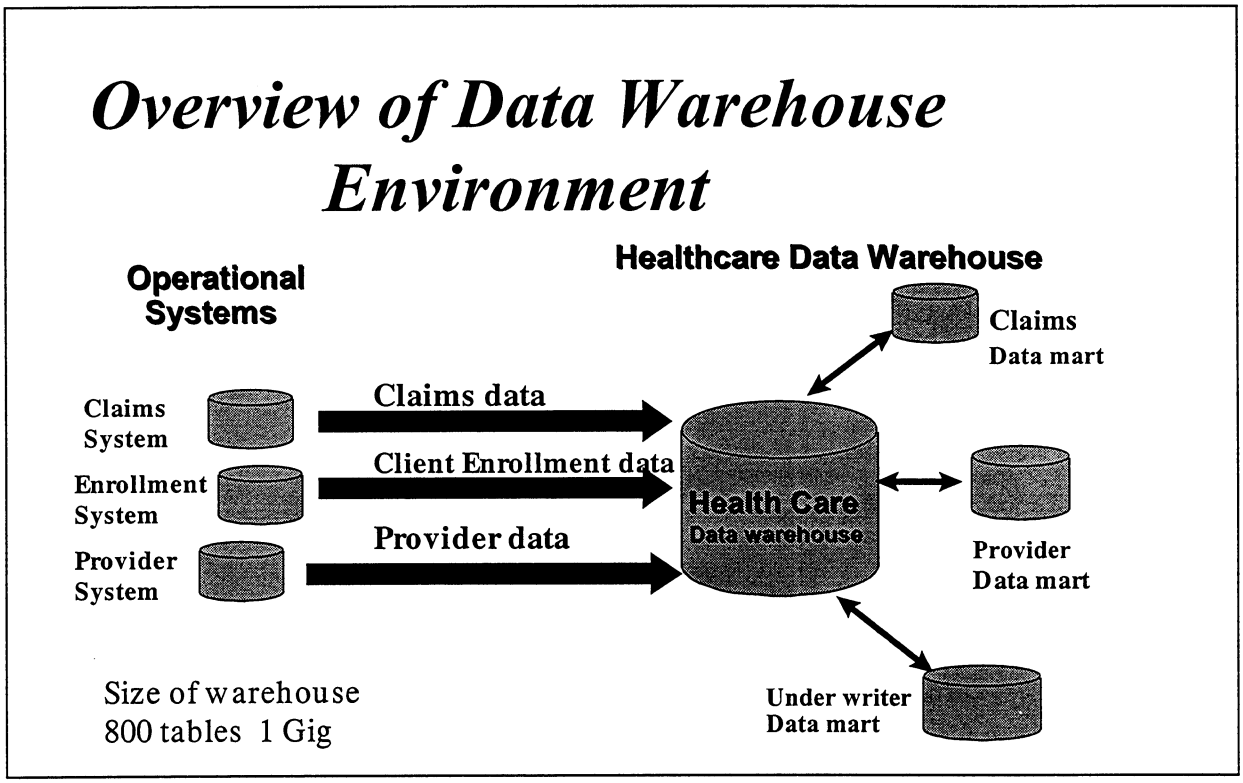

Figure 1

Like most data warehouses built, data quality issues can usually emerged. In this case, the financial data stored in the warehouse was not always consistent and accurate. To correct this, the initial action taken was to go to the source system responsible for the data in question trying to cost justify the resolution as well as the fix. This was done piecemeal as the problems were identified. The major issue was that no single business area currently has the ability to manage the resolution of the data issues since it crosses functional areas.

In addition, no one area had the resources, structure or mandate within the organization to ensure the quality of the enterprise data. As a result, there were persistent and pervasive data creditability, data usage, and access issues that was costing real dollars. The solution was to establish a cross functional data quality management team dedicated to rapidly and effectively eliminating data quality incidents and thus resulting in a continuous positive bottom line impact.

3. Justification of a data quality program

Surveys have revealed that data quality is a leading contributor to data warehouse time and cost overrun. This section will explain how the healthcare insurance provider came to the realization that an enterprise data quality program would be needed. It will explore some of the impacts that it can have on the day to day operations of a business and document some of the symptoms associated with poor data quality. The healthcare insurance carrier found that there was a direct and indirect cost associated with poor or inconsistent data. They found that inconsistent or poor data can

- generate inadequate customer service
- raise operating costs due to extensive adjustments and credits
- provide unreliable information for management decision making
- generate a lack of end-user confidence in Information Technology's ability to support the business
- generate regulatory fines thus raising expenses and lowering public confidence

When the healthcare organization started to research the incidents of invalid data, it found that often they were simply the symptoms of the problems but not the root cause itself. The symptoms by themselves, were not the culprit. It did however strongly suggest a lack of a formal methodology or standard operating procedures for documenting data and process in the enterprise. Some of the symptoms identified were:

- Invalid computations
- Business Rule violations
- Invalid data values and ranges
- Multiple source systems for data
- Invalid data transformation rules

4. Overview of the data quality program

In the case study described above, we learned that the major issue was that data quality problems tended to cross functional boundaries yet no one area possesses the ability or resources to address these issues. The conclusion was to recommend the establishment of a central area that could focus only on these issues and yet could interface with other functional areas across the enterprise. This was the Enterprise Data Quality Program. Figure 2 describes the functional relationships and the following will describe how the program worked. Within this functional diagram, there are five components that interact to facilitate the program:

Data Quality Council, Business Owner, Data Quality Management (DQM), IT Data Management and IT Owners. The following will identify and describe the roles and responsibilities for each area as well as how they interrelate with each other:

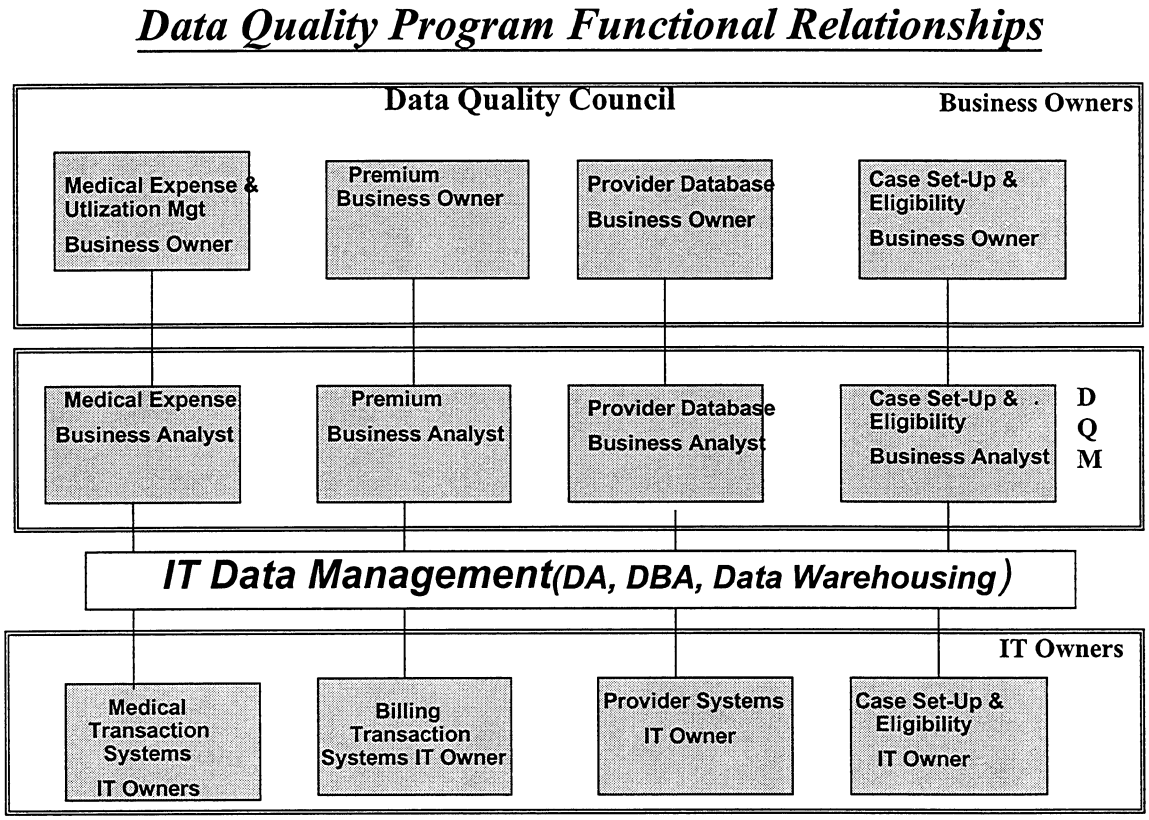

Fig 2

Data Quality Council

This is the overall guidance organization for the Data Quality Program. It consists of one person from each functional area. To ensure that support is forth coming from the IT Owners, the CIO is a member of this Council.

- Prioritize Approved Data Quality Incidents
- Assist Business Owners to define DQM Business Policies and set Strategic Direction
- Work with IT/CIO, DQM and Business Owners to ensure Program Support

Business Owners

The Business Owners, are really the Data Owners within each functional area are the senior management for each functional business area.

- Conduct Daily Operations
- Define Business Policies and Strategic Direction
- Work with DQM team to ensure rigorous Testing

Data Quality Management Team

The Data Quality Management team's primary role is to manage, investigate and resolve data quality issues. You will notice that this team is aligned by functional area.

- Perform root cause identification and resolution
- Identify data elements that directly impact the **Business Drivers**
- Develop data measurement and exception reports
- Coordinate with **IT Data Architecture** data element definitions, retention policies, and usage between functional areas
- Ensure monthly reconciliation's are performed between the transaction systems and the warehouses/general ledger
- Assist Business Owners with developing rigorous **Business Testing** of system releases

- Ensure consistency and efficiency of data warehouses

IT Data Management

This group consists of Data Architects, DBA's and Data Warehousing. They are responsible for all data management related activities in the organization. You will notice that the Data Management team is the conduit with the DQM and IT Owners through the Data Stewardship Role. This role will be described in greater detail in the Framework section to follow.

- Work with DQM team to identify data elements that directly impact the Business Drivers,
- Coordinate with DQM team, data element definitions , retention policies, and usage between functional areas
- Recommend and utilize **Data Quality** tools to assess and when appropriate, cleanse our business data
- Work with IT owners to identify appropriate data sources and business rules for Data Warehouse

IT Owners

This group is the IT management organization responsible for the applications that support the Business Owners.

- Project Management
- Write Technical Specifications
- Application & IT Resource management
- Implement Data Quality Tools and System

5. Overview of Data Quality Framework

I would strongly suggest that an effective data quality program is only as good as the framework it is build upon. The framework consists of the standards, methodology and tools necessary to support the program These framework components are identified below in Figure 3.

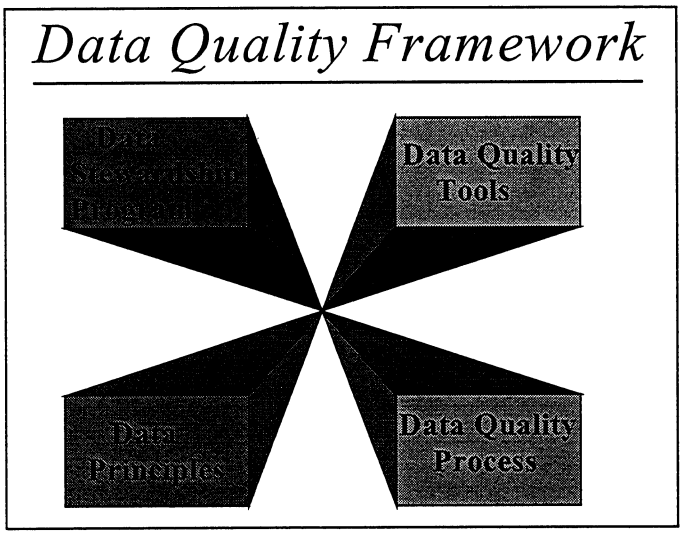

Fig 3

The data quality framework consists of four different components; Data Stewardship program, Data Principles, Data Quality Tools and Data Quality Process.

Data Stewardship Program

Data Stewardship is a program by which staff members are designated as information experts of a particular data subject area. This person will have complete knowledge about the data within this subject area such as Client,

Provider, Claims and Pharmacy. The data steward working with the data administrator as a consultant, will participate in data analysis, data mapping and data modeling. The roles and responsibilities of the data stewards are detailed here.

Data Documentation

The data steward will review and approve all documentation within their subject area to insure completeness and accuracy. After their review, the steward will make sure the business data owner signs off on everything as a final approval. The data steward must also be familiar with the naming standards and insure that all standards are followed within the subject area.

Data Modeling

The data stewards will participate in all data modeling sessions when required. more than one steward may be involved in the same session. This is typically the case when there is an overlap of data subject areas to be analyzed.

Business Rule Specifications

The data steward works as a resource and consultant in the creation and/or

enhancement of business rules within a subject area. The steward then reviews all specifications to insure that they are clear, concise and accurate in order for business users to easily be able to reference and understand them.

Standard Calculation and Transformation Definitions

The data steward will work with data administrators to determine and formulate any calculation or transformation rules needed to populate the data warehouse within the subject area specified.

Data Quality

The data steward reviews documentation to insure program compliance. The data steward works in conjunction with data administrators when dealing with business and/or IT areas to identify data quality issues and determine appropriate solutions. The steward insures appropriate standards are followed, such as the use of approved data cleansing tools, the following of appropriate business rules, and entity and attribute definitions.

Data Sourcing for the Data Warehouse

The data steward is used as a consulting resource in the research of the data warehouse sourcing. They review and approve all documentation pertaining to their subject areas. A system of records will need to be documented, to insure that data is consistently acquired from the same sources. The stewards will insure that this documentation is complete and accurate.

Data Principles

The second component of the framework is Data Principles. They are management directives that identify how an organization will design, develop and deploy its data architecture strategy. They are intended to be used by IT data management, development personnel and business users. To be useful and effective, each individual principle must ensure that the principle...

- is clear
- addresses a significant information technology subject
- is applicable to the enterprise
- is achievable by the enterprise
- presents realistic constraints to choices.

These principles will serve as a standard framework to guide the design, development and usage of the enterprise data architecture and will govern good data quality practices. Here are a few examples that were used by the Healthcare Insurance provider :

- Data must be accurate, consistent and complete.
- Every data element will be managed by a data steward.

- Explanations for each data element will be available in an on-line data repository.
- Every data element used in the information management environment will be obtained from one and only one identified source.
- Time sensitive data will be time stamped.
- All data must be refreshed on a defined basis.
- All data that is deemed of poor quality will be cleansed at its source.

The Data Steward is responsible for the enforcement of the data principles. There are various checkpoints within the system development life cycle.

Data Quality Tools

The third component of the framework is Data Quality Tools. Now that you have established a Data Stewardship Program to identify the proper business definitions and structures for data, and established a standard list of Data Principles to govern their use,. you must identify what tools are necessary to assess and cleanse the data in your organization. Data quality is enhanced by the institution of proper and adequate processes to ensure that data is validated and cleansed at its source. Data quality tools will facilitate the data assessment and cleansing process. These tools provide three basic functions across the enterprise. They analyze and report on data quality, standardize data attributes and merge/match records that reside in enterprise databases.

There are three types of data quality tools; Name and Address Cleansing tools,

Data Quality Cleansing tools and General Data Quality Cleansing tools. The following describes their usage.

Name and Address Cleansing tools

These tools are used to assess name and address data within an organization. The tools have special search engines that can check for correct spelling of client names and street addresses. Most tools will not only identify potential problems but also make an effort to cleanse the data as well. Normally it is the sales or marketing function that has the most to gain from using these tools.

Data Quality Cleansing tools

These tools are very specialized as they are used to cleanse data once it is found to be incorrect or invalid. They are different from the name and address cleansing tools because they are used to cleanse other types of data such as financial data.

General Data Quality Cleansing tools

These tools will provide the ability to conduct a generic assessment of your data that resides in standard files or databases. They look for general patterns and try to associate various values within the same data source or across many data sources. These tools are sometimes called data profiling tools because they are typically used to blindly read files without any pre-conceived notions about the data and its associations. Most of these tools can not access the data directly but must have it extracted into a flat file first.

Data Quality Process

The fourth and final component of the Framework is the Data Quality Process. This is a methodology that consists of nine steps. These steps are identified in. Figure 4.

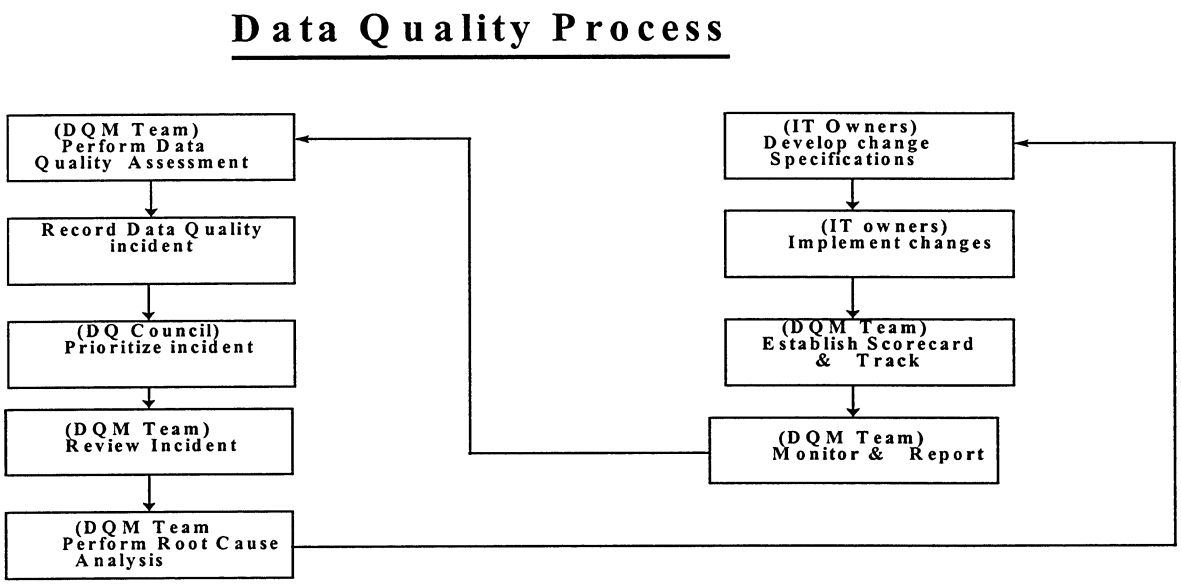

Figure 4

Perform Quality Assessment

The first step in the process is to perform a data quality assessment of the enterprise's most important data. The reason for this is so you will have a baseline to associate the quality going forward. This is normally done by a member of the Data Quality Management Team (DQM) on a periodic basis.

Record Data Quality incident

Whenever the integrity of the quality of the data is determined to be suspect it

should be documented in a Data Quality Incident database. Information such as date/time along with the reporter of the incident and the description is of the problem is documented. This is normally done by someone in the Business Owner's area however it can also be done by a member of the data quality team or the IT development team.

Prioritize incident

The Data Quality Council review the incidents reported with the Business Owners to determine the impact to the business. Incidents are normally prioritized based on business impact, the total dollar amount and the frequency or scope. They are then prioritized and assigned to the Data Quality Management Team for review.

Review Incident

The DQM Team will review the incidents in order of priority. In addition the Data Steward will attempt to validate that the problem truly exists by doing scans or queries against the production database.

Perform Root Cause Analysis

Once the incident has been validated, it is then assigned to a member of the DQM team to perform Root Cause Analysis. This is simply a way of identifying where the problems exists and why.

Develop change Specifications

The Root Cause results are reported back to the DQM team and the IT Owners and a specification identifying the solution is documented it required. Not are changes are made due to their potential impact to the business or expense associated with their testing and implementation.

Implement changes

If it is deemed appropriate, the change is tested and made by the IT Owners. This is done with the support and cooperation of the DQM and Data Management Teams.

Establish Scorecard & Track

Now that the change has been made, it is tracked by the DQM Team for future reference using a scorecard.

Monitor & Report

The DQM Team will periodically monitor the change and report back to the Business Owners, and IT Owners as appropriate.

6. Closing Remarks

In conclusion, you have now seen how a major healthcare insurance provider identified the impacts of poor data quality on their business. You should also have an understanding what a data quality program is, how it works and why it is

important to have one. Finally you should understand why it is important to established a framework consisting of Data Stewardship, Data Principles, Data Quality Tools and a process to facilitate the implementation and support of your program.