

Data Quality Report

Arman Kanooni

101001010100111101000010010111010010 1101010101011101000010001001001001
0010000101001010010010100001011010010101000011110100101010011110100001001011010010
110101010101110100001000010100101001010000101101001010100001111010010101

Table of Contents

- *Definition*
 - *Data*
 - *METADATA*
 - *Data Quality*
 - *Data Quality System*
 - *Data Quality Program*
- *Data Quality Approached*
- *What is the Cost of Non-Quality Data?*
- *How to Improve Data Quality?*
- *TQdM Methodology Overview*
- *What are Data Quality Standards?*
- *Data Quality Management Maturity Grid*
- *How to Develop a Data Quality Culture?*
- *Data Quality – A Case Study*
 - *Case 1: Jeppesen*
 - *Case 2: The Royal Bank of Canada*
 - *Case 3: HP*
 - *Case 4: National Association of Securities Dealers Rockville, MD*



Introduction

- *In the movie **War Games**, set in the U.S. Strategic Air Command Center under a mountain near Colorado Springs, there is a critical scene: on a huge map of the globe are arrows indicating large numbers of nuclear missiles approaching the U.S. from the Soviet Union. The military official in charge is trying to decide if he should ask the President for permission to launch a retaliatory attack because the technicians are telling him the threat is real. At this point, the scientist who originally designed the system (and who knows that something is wrong with the system itself) says, “General, what you see on that board is not reality, it is a computer-generated hallucination!”*
- ***It is increasingly vital that the pictures being displayed are correct!***

Definition

Data

- *Data is defined as facts, images, sounds, or signals stored in documents, files, systems, databases, servers, web sites, etc.*

METADATA

- *Data about data. Data which defines and describes other data, e.g., for the physical data element BUDGNUM, potential metadata might be the*

- (1) Business name*
- (2) Business definition*
- (3) Accountable steward*
- (4) Physical location*
- (5) Valid aliases, etc.*



Definition

Data Quality

- Data quality is defined as *“data that is fit for all purposes in the enterprise that require it”*.

Quality data is:

- Consistent
- Complete
- Accurate
- Valid
- Non-redundant (controlled redundancy, single source, single authority)
- Timely
- Aligned with current business model
- Useable (proper representation when used)
- Has an accepted data definition.



Definition

Data Quality System

- *The totality of an organization's efforts to manage, control and improve data quality.*

The system includes

- *Activities aimed at understanding customer needs,*
- *Activities aimed at detection and correction of specific errors,*
- *Activities aimed at preventing future errors, and*
- *Management activities to build organization infrastructure to do so effectively and efficiently.*

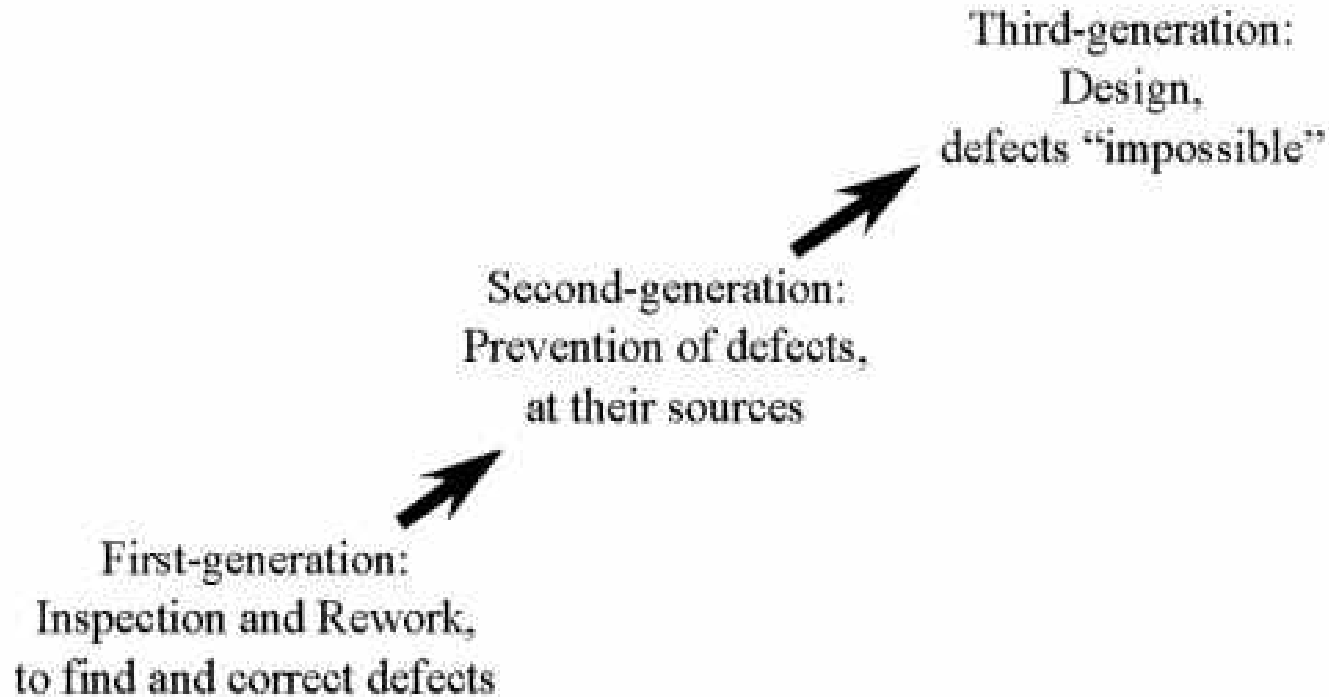
Data Quality Program

- *The organization's plans and implementation for advancing the data quality system.*



Data Quality Approach

Approaches to Data Quality



What is the Cost of Non-Quality Data?

- *Failure costs (scrap & rework)*
 - *Redundant data handling*
 - *Business rework*
 - *Work-around costs*
 - *Data cleanup costs*
 - *Software re-write costs*
- *Assessment cost (inspection)*
- *Prevention cost*

How to Improve Data Quality?

- Improvement does not just happen – someone must take the initiative.
- Data quality improvement is not the responsibility of the data quality department or leader – it is the responsibility of everybody in the enterprise!
- Perfection is not the objective – customer satisfaction is!
- Data quality improvement is not a project, program or initiative – it is a habit!
- Don't be overwhelmed – quality is achieved one step at a time.



Total Quality Data Management Overview

1. Establish information value and cost of non-quality.
2. Analyze definition quality.
3. Analyze content quality.
 - Identify where data quality will be measured
 - Establish what data quality tests are to be performed on a data collection
 - Determine all business processes and applications that create or update data and their relationships to each other
 - Identify the authoritative sources used to validate the accuracy of the data
 - Extract a sample of data for quality assessment that accurately reflects the quality of the total population
 - Measure the quality of a data sample to determine the degree of reliability of a population of data
 - Communicate the reliability level of data and report the types of errors and their significance



Total Quality Data Management Overview

4. Implement data cleanup actions.
5. Implement information quality improvements.
 - Identify a process where improvements can prevent business problems that cause non-quality data
 - Identify the causes of data quality problem and possible actions to eliminate/minimize causes
 - Verify the effectiveness of the improvement actions
 - Make the effective information quality improvements permanent

What are Data Quality Standards?

- ❖ Data quality standards are the real quality requirements.
- ❖ Data Stewards define the required reliability levels of data based upon all knowledge worker requirements.
- ❖ Different data will have different reliability or quality levels (not all data needs to be 100% defect free.)
- ❖ Meaningful levels need to be set and then reset, as needed.
- ❖ The **criteria** used for quality standards are:
 - o **Consequences of non-quality data**; Economic (cost/missed opportunity), Image, Legal/regulatory, Irrevocable damage.
 - o **Cost to prevent versus cost of rework/clean-up.**
 - o **Priority/importance** of business processes, decisions, and objectives.
 - o **Source and nature of the control of the data.**
 - o **Extent of common requirements** (data sharing).



Data Quality Management Maturity Grid

Business Climate

<i>Level</i>	<i>Business / IT Alignment</i>	<i>Business / IT Portfolio</i>
<i>1</i>	<i>Business Driven</i>	<i>Not Defined</i>
<i>2</i>	<i>Emerging Enablement</i>	<i>Defined and Managed</i>
<i>3</i>	<i>Catalyzing Business Innovation</i>	<i>Integrated With The Business</i>

Data Quality Management Maturity Grid

Information Technology Climate

	<i>Business / IT Relationship</i>	<i>Executive Mindset</i>	<i>Organizational Experience</i>
Level 1	<i>Order Taker</i>	<i>Cost</i>	<i>Disappointing</i>
2	<i>Advisor</i>	<i>Investment</i>	<i>Predictable</i>
3	<i>Collaborator</i>	<i>Enabler</i>	<i>Exceptional</i>

Data Quality Management Maturity Grid

Information System Capabilities

	<i>IS Vision</i>	<i>I/S Aspects (People, Process Technology)</i>	<i>IT Governance</i>
Level 1	<i>Non-Existent</i>	<i>Basic Operating Services</i>	<i>IS Responsibility</i>
2	<i>Defined</i>	<i>Solutions Delivery</i>	<i>Business Participates in a Limited Way</i>
3	<i>Pervasive</i>	<i>Innovating</i>	<i>Shared Decision- Making Responsibility</i>

How to develop a data quality culture?

- *Link data quality initiative to some significant business outcome(s).*
- *To get from Level 1 to Level 2, concentrate on cost control and poor quality replacement.*
- *Level 2 establishes the "Solution Provider of Choice", and is focused internally.*
- *To get from Level 2 to Level 3, an external focus is required (i.e., Federal Express knows more about Dell products than Dell does -- and, in fact, sells the information back to Dell.)*

Data Quality – Case Study (Jeppesen)

- Data Quality in every aspect of Data entry and usage
- Same Data Quality in all of Jeppesens Products
- Data integrity for all Jeppesen Products

RTCA/EUROCAE Document DO-200A/ED-76,

Standards for Processing **Aeronautical Data**

- Latitude
- Longitude
- Height
- Scale
- Gravity
- Shoreline
- Orientation



Data Quality – Case Study (Jeppesen)

Verification

- No data entry without “Two Person cross-check”
- “Blind Re-Key”
- Graphical editing

Data Quality – Case Study (Jeppesen)

Validation

- *10,000 Business Rules*
- *Data Quality tools on three levels:*
 - *Entity Level*
 - *Across-Entity Level within Work Area*
 - *Across-Entity Level across complete data set*
- *Validation of affect on future data*
 - *Continuous data validation through data research*
 - *Count validation*
 - *Rule based Data extraction*
 - *Daily Internal Extraction and Quality Control*
 - *Graphical product review*
 - *Independent Quality Control Team*
 - *Automated Tools for Production Area*

Data Quality – Case Study (Jeppesen)

Quality Through Traceability

- *Story about data ready for product use*
- *No data deletion*
- *“Meta data” stored with Aviation Data*
- *Any change “linked” to originating source*
- *“Rework” and “Clean-Up” statistics*
- *On-Time statistics*

Data Quality – Case Study (Jeppesen)

Constant Quality Improvements

- ***Weekly quality and status meeting***
- ***Quality Issues result in:***
 - ***Work Requests for Development***
 - ***Process enhancement requests***
 - ***Workflow enhancement requests***
- ***Internal Quality Audits***
- ***External Quality Audits (ISO & DO-200A)***

Data Quality Case – Royal Bank of Canada

Problem Statement

- *Data problems*
- *IS Development and maintenance costs too high*
- *New applications developed within the old paradigm (i.e., localized, organizational, stove-pipe) which fostered redundancies, rework, interfaces/interpreters, lack of accountability, and dissatisfied down-stream customers, among other consequences of poor data management practices.*
- *Regarding quality of their data, the bank's level of pain was approaching its corresponding pain threshold, so management agreed to sponsor and support efforts to do something about it.*

Data Quality Case – Royal Bank of Canada

Clever Solution

- They leveraged off of other initiatives and utilized outside help as well as automated tools, to get the job done.*
- They began their journey by creating the necessary support infrastructure -- to include*
 - (1) Establishing the company-wide Data Management Group comprised of Data Architects, Modelers, and Data Base Administrators. Application Programmers and Analysts remained decentralized and assigned to their respective customers.*
 - (2) Creating standard practices and processes -- such as, no databases or applications will be approved for production without a data model, which is certified by the Central Data Management Group.*

Clever Solution

(3) Creating the central Information Assets Inventory which addresses:

- (a) What are the Bank's information and application assets?**
- (b) Where are they located?**
- (c) What are their components?**
- (d) Who is responsible for them (the stewards)?**

Data Quality Case – Royal Bank of Canada

The Clever Process

- *The process, by which they collected metadata about their applications and information assets and subsequently populated it into the Platinum repository tool, was somewhat opportunistic on the part of the Data Mgmt Group, as well as clever.*
- *The bank had hired an outside firm to conduct a year 2000 audit of all their applications. In so doing, the audit firm created a metadata file for each and every application (There were several hundred). When the audit firm was done with their work, the Bank's Data Mgmt Group requested and received those metadata files.*

Data Quality Case – Royal Bank of Canada

The Clever Process

- *Next, they hired Platinum Technology to come in and build a bridge between the metadata files and their Platinum repository tool. Once the bridge was in place, it took 2 1/2 months to migrate all the metadata -- to include 635,000 data elements -- to the repository. The Bank's Data Mgmt Group estimates that they saved 7 man-work-years by using this approach.*
- *Next, they had to rationalize the data elements and metadata (i.e., document synonyms, eliminate redundancy, bad definitions, etc.). They already had Integrity -- the data re-engineering tool from Vality Technology -- in-house and asked Vality if the tool would work on metadata as well as data. The answer was yes.*
- *Using the Integrity tool, then, it took them 2 1/2 months to clean up the metadata associated with the 635,000 data elements in the repository. They estimated that they saved another 2 1/2 man-work-years using this automated approach.*

Data Quality Case – Royal Bank of Canada

The Architecture

- *Centralized repository. As data passes thru various phases of its life, it is constantly validated against the repository.*
- *Source file data is defined in the repository and the two (repository and source file) are kept in-sync, over time.*
- *Data warehouses and data marts get their data definition information from the repository and are also kept in sync, over time.*
- *Data engineering and re-engineering efforts get metadata from and supply new metadata to the repository.*
- *Currently, they are operating in an environment where disciplines are in place to ensure consistent, non-redundant, commonly defined, sharable, quality data.*



Data Quality Case – Royal Bank of Canada

The Architecture

- *They are utilizing the web as a better way for customers to get at that data.*
- *Change management is in place via a 4-level process. When level 3 is reached, email goes out to all parties impacted by the change(s).*
- *The results so far are very encouraging -- faster delivery of business information, requires fewer intermediary clerical and programmer resources, quick data discovery is enabled, up-to-date models are trusted and used to enable data engineering and re-engineering, maintenance and workarounds reduced significantly, satisfied down-stream customers, and so on.*
- ***If we don't have quality of definition (metadata), how can we have quality of content!***



Data Quality Case – HP

The HP Business Issue

- *To be able to measure the value of quality data or the cost of bad data.*
- ***Work on this was supported at the HP executive level.***
- *Their Data Quality Program encompassed;*
- *Sponsorship/direction - came from business units (not corporate) and provided vision, support, funding, scope.*
- *Methodology Design - used both internal and external expertise and collaboration will result in a better product.*
- *Methodology Use - developed a case study. Feedback was critical. Obtain commitment to provide feedback up front and use it.*
- *Program Management - planning, communication, coordination and integration.*

Data Quality Case – HP

- *Strong project management principles and disciplines were applied throughout its evolution. This proved vital in getting direction from the DMC, keeping them informed, and getting buy-in at critical steps. They were also lucky during their case study when they ran into a problem and the manager was willing to take the risk and keep going."*
- **Culture -**
 - *Risks: "If we can't talk about the problems, nothing will get fixed."*
 - *Leverage Learning: "Don't reinvent the wheel."*
 - *TQM: Attitude of continuous improvement.*
- **Resources -** *Data quality must be an on-going, sustained effort so be prepared for the long term. Build a pool of people with appropriate skills (begin by identifying what skills exist already in the organization).*

Data Quality Case – HP

- *They also defined a quality improvement process, which consisted of a mandatory handshake between IT and the business community.*
- *Their advice is to articulate a crisp vision, engage your customer, and if what you want lies buried, dig until you find it!*

