



MIT 2008 Information Quality
Industry Symposium
July 16-17, 2008
Boston, Massachusetts

Ten Steps to Quality Data and Trusted Information™ - An Overview

Danette McGilvray
Granite Falls Consulting, Inc.
President and Principal
Phone: 510-501-8234
Email: danette@gfalls.com
Web: www.gfalls.com
Fremont, California USA

Copyright Information

These materials, and any part thereof, are protected under copyright law. Course participants are granted the right to "fair use" of these materials. The contents of this document may not be reproduced or transmitted in any form, in whole or in part, or by any means, mechanical or electronic, for any other use, without the express written consent of Danette McGilvray or Elsevier Inc.

Portions of this work are from the forthcoming book available Summer 2008, *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™*, by Danette McGilvray, published by Morgan Kaufmann Publishers, Copyright 2008 Elsevier Inc. All rights reserved.

All uses of Ten Steps to Quality Data and Trusted Information™, The Ten Steps™, or Ten Steps™ throughout these materials are protected by trademark law.

Presentation Description

Data quality situations familiar to many organizations:

- Business has invested heavily in data purchased from external sources, yet cannot depend on the quality to meet the company needs.
- The data warehouse has been in production for over a year. Users from the business intelligence group don't trust the reports, complain about the quality, and are reverting to their own spreadsheets for verification.
- An ERP (Enterprise Resource Planning) application has been implemented. Data previously used by one business function is being used in end-to-end processes – with poor results.
- The organization is starting a data integration project. The project team has a tight schedule, yet already knows there are quality issues with the source data to be moved.

There is help available! This presentation provides an overview of a methodology, Ten Steps to Quality Data and Trusted Information™, which is a systematic approach to improving and creating data and information quality. The methodology combines a conceptual framework for understanding information quality and The Ten Steps™ process which provides instructions, techniques, and best practices for implementing the key concepts.

Background

Danette McGilvray is President and Principal of Granite Falls Consulting, Inc., a firm specializing in information quality management. Projects include enterprise data quality services, data warehousing strategies, data governance, and best practices for large-scale ERP data migrations for Fortune 500 organizations. Her book on data quality, "Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™" (Morgan Kaufmann) will be available Summer 2008.

Danette is an invited speaker at conferences throughout the US and Europe. She is a member of DMReview.com's Ask the Expert panel. Her previous experience as a leader of enterprise data quality within a company and now working with clients in various industries, gives her understanding of the information quality challenges faced daily by organizations. She has been profiled in PC Week and HP Measure Magazine and was an invited delegate to the People's Republic of China to discuss roles and opportunities for women in the computer field.

Dealing with Poor Health

“Doctor, my left arm hurts!”

The doctor puts your arm in a sling, gives you an aspirin and tells you to go home.



But what if you were really having a heart attack?

You would expect the doctor to **diagnose** your condition and take **emergency measures** to save your life.

After you were **stabilized** you would expect the doctor to:

- Run **tests**
- Get to the **root cause** of the heart attack
- Recommend measures to **correct damage** done (if possible) and **prevent** another heart attack.

The doctor would have you come in for **periodic additional tests** and **follow-up to assess** your condition.

MIT2008IQIS

© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.

5

Dealing with Unhealthy Data and Information

When it comes to data and information quality, how often do we:

- Address the immediate problem, then
- Go for the “easy fix” (the aspirin and sling) and
- Expect that to take care of our problems?

- No tests or assessments are run to determine the location or magnitude of our problems
- No root cause analysis is performed
- No preventive measures are put into place

And then we are surprised when the problems appear and reappear!

Just like your own health, you can:

- **Prevent** data quality “health” problems
- **Assess** and take **action** if they appear

MIT2008IQIS

© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.

6

Dealing with Unhealthy Data and Information

This presentation will introduce you to a methodology that will help with your organization's data and information quality health.

Common data quality situations:

- Company has implemented an **ERP** (Enterprise Resource Planning) application. Data previously used by one business function is being used in end-to-end processes – with poor results.
- The company is starting a **data integration** project. The project team has a tight schedule, yet already knows there are quality issues with the source data to be moved.
- The company invests in a major **data clean-up project**, and a few years later starts **another data clean-up project** because data quality declined and is causing issues for the business.
- The **data warehouse** has been in production for over a year. Users from the business intelligence group don't trust the reports, complain about quality, and are reverting to their own spreadsheets for verification.
- Business has invested heavily in **data purchased from external sources**, yet cannot depend on the quality to meet the company needs.
- Data quality is an important part of your **daily responsibilities**.

Data vs. Information

- **Data** – Known facts or other items of interest to the business
- **Information** – facts within context
- **Are there differences between the two?**
- This approach does not generally differentiate between data and information. Some organizations respond to “data quality” and others respond to “information quality.” Use the term most meaningful to your organization and those with whom you are speaking.

Ten Steps to Quality Data and Trusted Information

A systematic approach to **improving and creating data and information quality** within your business.

The methodology combines:

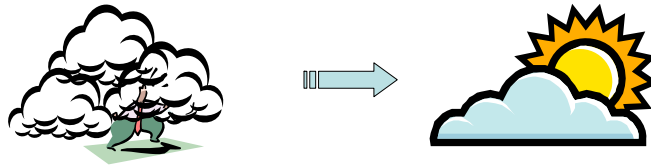
- A **conceptual framework** for understanding information quality with
- The **tools, instructions, and best practices** for improving information quality.

Your company's* "wellness" program for data and information.

* Company includes any organization such as for-profit businesses, non-profit and charitable organizations, government agencies, and educational institutions.



Outline



- **Key Concepts**
- **The Ten Steps Process and Projects**
- **Summary and Best Practices**

Framework – Like the Food Pyramid

Just as the Food Pyramid provides **guidelines and a visual of the components** for healthy eating and physical activity, the Framework for Information Quality provides the components necessary for “healthy” data.



Source: mini poster.pdf available at <http://www.mypyramid.gov/>, accessed July 16, 2007. The Food Pyramid was developed by The Center for Nutrition Policy and Promotion, an organization of the U.S. Department of Agriculture. It was established to improve nutrition and promote dietary guidance for all Americans.

Framework for Information Quality (FIQ)

1 Business Goals / Strategy / Issues / Opportunities (Why)						
	2 Plan	Obtain	Store & Share	Maintain	Apply	Dispose
3 Data (What)						
Processes (How)		4				
People/Orgs (Who)						
Technology (How)						
5 Location (Where) and Time (When and How Long)						
Requirements and Constraints: Business, Technology, Legal, Contractual, Industry, Internal Policies, Privacy, Security, Compliance, Regulatory						
Responsibility: Accountability, Authority, Governance, Stewardship, Ownership, Motivation, Reward						
6 Improvement and Prevention: Root Cause, Continuous Improvement, Monitor, Metrics, Targets						
Structure and Meaning: Definitions, Context, Relationships, Standards, Rules, Architecture, Models, Metadata, Reference Data, Semantics, Taxonomies, Ontologies, Hierarchies						
Communication: Awareness, Out-Reach, Education, Training, Documentation						
Change: Management of Change and Associated Impact, Organizational Change Management, Change Control						
7 Culture and Environment						

FIQ Sections Explained (1)

- ① **Business Goals/Strategy/Issues/Opportunities:** The “why”– Anything done with information should support the business in meeting its goals.
- ② **Information Life Cycle*** Use POSMAD to help remember the information life cycle:
 - **P**lan - Identify objectives, plan information architecture, develop standards and definitions; model and design applications, databases, processes, organizations, etc.
 - **O**btain – Data or information is acquired in some way, e.g. create records, purchase data, load external files, etc.
 - **S**tore and Share – data is stored and made available for use.
 - **M**aintain – Update, change, manipulate data; cleanse and transform data, match and merge records, etc.
 - **A**pply - “Retrieve” data, use information. Includes all information usage such as completing a transaction, writing a report, making a management decision, completing automated processes, etc.
 - **D**ispose – Archiving information; delete the data or records.

* Also known as an Information Chain, Information Value Chain, Information Resource Life Cycle

MIT2008IQIS

© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.

13

FIQ Sections Explained (2)

- ③ **Key Components** affecting information quality
 - **Data (What)** - Known facts or other items of interest to the business.
 - **Process (How)** - Activities, actions, tasks, or procedures that touch the data or information (business processes, data management processes, processes external to the company, etc.).
 - **People/Organizations (Who)** - Organizations, teams, roles/responsibilities or individuals.
 - **Technology (How)** – Forms, applications, databases, files, programs, code, or media that store, share, or manipulate the data, are involved with the processes, or are used by the people and organizations.
- ④ **Interaction Matrix** between information life cycle phases and key components
- ⑤ **Location (Where) and Time (When and How Long)**

Note: Top half of the framework along with the first orange bar answers the interrogatives of who, what, how, why, where, when, and how long

MIT2008IQIS

© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.

14

FIQ Sections Explained (3)

- ⑥ **Broad-Impact Components** – additional factors affecting information quality. Lower your risk by ensuring these factors have been considered and appropriately addressed. If they are *not* addressed, you are still at risk (RRISCC) as far as information quality is concerned.
- **R**equirements and Constraints
 - **R**esponsibility
 - **I**mprovement and Prevention
 - **S**tructure and Meaning
 - **C**ommunication
 - **C**hange
- ⑦ **Culture and Environment**- Your company's attitudes, values, customs, practices and social behavior; conditions that surround people in your company and affect the way they work and act. Take into account to better accomplish your goals.

Using the Framework

Use as a tool for:

- **Diagnosis** – Realize where breakdowns are occurring; assess your practices, determine if all components necessary for information quality are present.
- **Planning** – Design new processes, determine where to invest time, money, and resources.
- **Communication** – Explain the components required for and impacting information quality.

The framework allows us to organize our thinking in a way so we can plan and **take effective action**.

Data Quality Dimensions

Aspects or features of information and a way to classify information and data quality needs. Dimensions are used to define, measure, and manage the quality of the data and information.

- In order to improve information quality, there must be a way to measure it.
- Measure the dimensions that best address your business need.
- There is no industry standard for the types of data quality dimensions.
- The dimensions defined here are derived from experience and are those most feasible and useful within the usual constraints of most businesses.

MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
17

Data Quality Dimensions

<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Data Specifications</div> <p>A measure of the existence, completeness, quality, and documentation of data standards, data models, business rules, metadata, and reference data.</p>	<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Ease-of-Use and Maintainability</div> <p>A measure of the degree to which data can be accessed and used and the degree to which data can be updated, maintained, and managed.</p>
<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Data Integrity Fundamentals</div> <p>A measure of the existence, validity, structure, content and other basic characteristics of data.</p>	<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Data Coverage</div> <p>A measure of the availability and comprehensiveness of data compared to the total data universe or population of interest.</p>
<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Duplication</div> <p>A measure of unwanted duplication existing within or across systems for a particular field, record, or data set.</p>	<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Presentation Quality</div> <p>A measure of how information is presented to and collected from those who utilize the information. Format and appearance support the appropriate use of the information.</p>
<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Accuracy</div> <p>A measure of the correctness of the content of the data (which requires an authoritative source of reference to be identified and accessible).</p>	<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Perception, Relevance, and Trust</div> <p>A measure of the perception of and confidence in the data quality; the importance, value, and relevance of the data to the business needs.</p>
<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Consistency and Synchronization</div> <p>A measure of the equivalence of information stored or used in various data stores, applications, and systems, and the processes for making data equivalent.</p>	<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Data Decay</div> <p>A measure of the rate of negative change to the data.</p>
<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Timeliness and Availability</div> <p>A measure of the degree to which data are current and available for use as specified and in the timeframe in which they are expected.</p>	<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">Transactability</div> <p>A measure of the degree to which data will produce the desired business transaction or outcome.</p>

08-03

MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
18

Assessments and Dimensions of Quality

Different tools, techniques, and processes are used to assess, measure, and manage the various dimensions of quality (with varying levels of time, money, and resource required).

Why differentiate the dimensions of quality?

- Match dimensions against a business need and prioritize which assessments to complete and in what order.
- Understand what you will (and will not) get from assessing each dimension
- Better define and manage the sequence of activities in your project plan within time, money, and resource constraints

Determining What to Assess for Quality

First, understand the business issues driving the data quality assessment and improvement activities. Then ask yourself:

- **Should** I assess the data?
 - Only spend time testing when you expect the results to give you actionable information related to your business needs
- **Can** I assess the data?
 - Is it possible or practical to look at this quality dimension?
 - Sometimes you cannot practically assess the dimension of quality or the cost to do so is prohibitive

Only assess and manage quality for those dimensions where the answer to both of the questions above is “yes.”

Business Impact Techniques

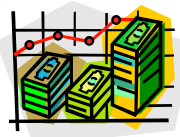
Awareness of data quality issues leads to questions:

- “Why does this matter?”
- “Why should I care?”
- “What impact does this have on the business?”

Answer those questions by using quantitative and qualitative techniques to assess the impact of data quality on the business.

Use results from assessing business impact to:

- Establish the business case for information quality
- Gain support for investing in information quality
- Determine the optimal level of investment



MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
21

Business Impact Techniques

1	Anecdotes	Collect examples or stories of the impact of poor data quality.
2	Usage	Inventory the current and/or future uses of the data.
3	Five “Whys”	Ask “Why” five times to get to real business impact.
4	Benefit vs. Cost Matrix	Analyze and rate the relationship between benefits and costs of issues, recommendations, or improvements.
5	Ranking and Prioritization	Rank impact of missing and incorrect data to specific business processes.
6	Process Impact	Illustrate the effects of poor quality data to business processes.
7	Cost of Low Quality Data	Quantify the costs and revenue impact of poor quality data.
8	Cost-Benefit Analysis	Compare potential benefits of investing in data quality with anticipated costs through an in-depth evaluation. Includes Return on Investment (ROI) – profit from an investment as a percentage of the amount invested.

Less Complex/ Less Time **Relative Time and Effort** More Complex/ More Time

← 1 2 3 4 5 6 7 8 →

08-01

MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
22

Choosing Which Techniques to Use

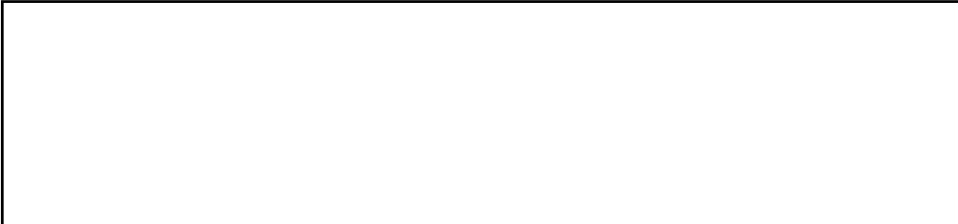
- Use the techniques that best fit your situation, time, and resources available.
 - Many of the techniques work together or can be used alone
- The continuum shows relative effort – not relative results:
 - You can understand business impact even without completing a full cost/benefit analysis
 - Less complicated does not necessarily mean less useful results
 - More complex does not necessarily mean more useful results
 - The best results come from using the techniques most appropriate to your situation

Root Cause Analysis

Root cause analysis is the process of analyzing all possible causes of a problem, issue, or condition to determine the actual cause.

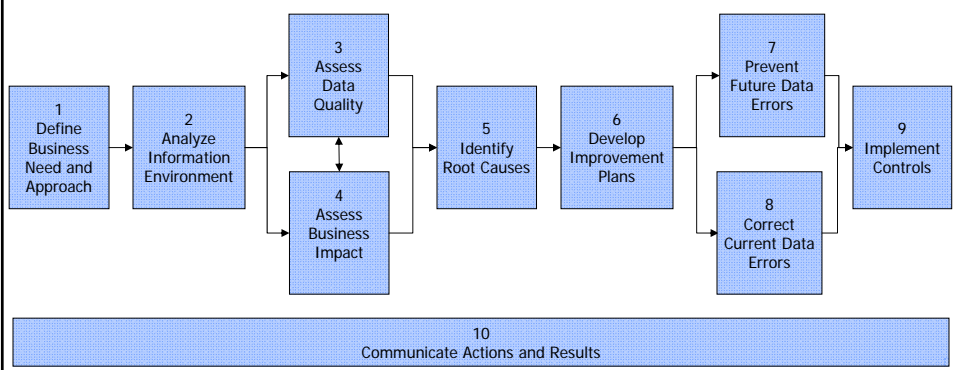
1	Five “Whys” for Root Cause	Leverage a basic quality approach by asking “Why” five times to get to root cause.
2	Track and Trace	Identify the location of the problem by tracking the data through the information life cycle and determining the root cause where the problem first appears.
3	Cause-and-Effect / Fishbone Diagram	Use a standard quality technique to identify, explore, and graphically display all possible causes of an issue.

The Ten Steps Process and Projects



MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
25

The Ten Steps



<p>1. Define Business Need and Approach. Define and agree on the issue, opportunity, or the goal to guide all work done throughout the project. Refer to the Business Need and Approach throughout the other steps in order to keep the goal at the forefront of all activities.</p> <p>2. Analyze Information Environment. Gather, compile, and analyze information about the current situation and information environment. Document and verify the information life cycle, which provides a basis for future steps, ensures relevant data is being assessed, and helps discover root causes. Design data capture and assessment plan.</p> <p>3. Assess Data Quality. Evaluate data quality for the quality dimensions applicable to the issue. The assessment results provide a basis for future steps such as identifying root causes and needed improvements and data corrections.</p> <p>4. Assess Business Impact. Provides a variety of techniques to determine the impact of poor quality data to the business. The impact provides input to establish the business case for improvement, gain support for information quality, and determine appropriate investments in your information resource.</p>	<p>5. Identify Root Causes. Identify and prioritize the true causes of the data quality problems and develop specific recommendations for addressing the root causes.</p> <p>6. Develop Improvement Plans. Develop and execute improvement plans based on recommendations.</p> <p>7. Prevent Future Data Errors. Implement solutions that address root causes of the data quality problems.</p> <p>8. Correct Current Data Errors. Implement steps to make appropriate data corrections.</p> <p>9. Implement Controls. Monitor and verify improvements that were implemented. Maintain improved results by standardizing, documenting, and continuously monitoring successful improvements.</p> <p>10. Communicate Actions and Results. Document and communicate results of quality tests, improvements made, and results of those improvements. Communication is so important that it is part of every step.</p>
--	---

MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
26

Using the Ten Steps Methodology

Pick and choose appropriate steps, activities, and techniques from the methodology:

- For **information-quality focused** projects
- In the course of **daily work** where you have responsibility for managing data quality or the work you do impacts data quality
- To **integrate** specific DQ/IQ activities into **other projects and methodologies** (e.g. ERP migration or building a data warehouse)

Approaches to Data Quality in Projects (1)

Establish Business Case

- Exploratory assessment or quick proof of concept assessing data quality on a very limited set of data. As an individual, you can implement a brief project that will help you make a business case for further data quality improvements. If you already have a specific data quality problem, you may just want to assess the business impact of that problem without further quality assessment.

Establish Data Quality Baseline

- When the business has committed to improving data quality and there is support for a project team and resources.

Determine Root Causes

- Use this approach when you already know the data quality issues and have determined the impact of those issues warrants further investigation into the real cause.

Approaches to Data Quality in Projects (2)

Implement Improvements

- Execute the recommendations developed when the data quality assessment and business impact analysis generate a plan for data quality improvement.

Implement Ongoing Monitoring and Metrics

- Focus on instituting operational processes for monitoring, evaluating, reporting, and acting on results.

Address Data Quality as an Individual

- Use data quality techniques in the course of daily work where you have responsibility for managing data quality or to address a specific data quality issue as an individual.

Integrate Data Quality Activities into Other Projects and Methodologies

- Combine The Ten Steps activities with your company's favored project management and project life cycle and include in your specific project plan.

MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
29

Approaches to Data Quality and The Ten Steps Process

```

    graph LR
      1[1 Define Business Need and Approach] --> 2[2 Analyze Information Environment]
      2 --> 3[3 Assess Data Quality]
      2 --> 4[4 Assess Business Impact]
      3 <--> 4
      3 --> 5[5 Identify Root Causes]
      4 --> 5
      5 --> 6[6 Develop Improvement Plans]
      6 --> 7[7 Prevent Future Data Errors]
      6 --> 8[8 Correct Current Data Errors]
      7 --> 9[9 Implement Controls]
      8 --> 9
      10[10 Communicate Actions and Results]
  
```

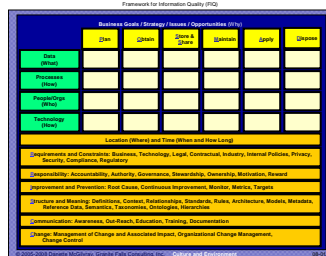
10 Communicate Actions and Results
Establish Business Case: Steps 1, 2, 3, 4, 10
Establish Data Quality Baseline: Steps 1, 2, 3, 4, 5, 6, 10
Determine Root Causes: Steps 1, 2, 3, 4, 5, 6, 10
Implement Improvements: Leverage baseline results plus Steps 7, 8, 10
Implement On-going Monitoring and Metrics: Leverage baseline results and improvements implemented plus Steps 9 and 10
Address Data Quality as an Individual: Steps vary
Integrate Data Quality Activities Into Other Projects and Methodologies: Steps vary

MIT2008IQIS
© 2005-2008 Danette McGilvray, Granite Falls Consulting, Inc.
30

Summary and Best Practices

The Methodology Has Two Main Components

Framework for Information Quality (FIQ) and Other Key Concepts



Ten Steps Process



- Provides the foundation for understanding information and data quality
- Shows the components necessary for information quality
- Concrete instructions for implementing, improving, and creating data quality
- Process for implementing framework and key concepts
- Contains examples, templates, techniques, and advice

Best Practice – Apply to Any Data and Information

Use what you have learned to improve any data:

- Customer
- Order Management
- Sales and Marketing
- Finance
- Procurement
- Manufacturing
- Etc.



Apply to any category of data:

- Master data
- Transactional data
- Reference data
- Metadata

Guidelines for Applying the Methodology

- **Relevant.** Ensure your work is associated with the business issue to be resolved.
- **Pick-and-choose.** Use only those steps applicable to your project.
- **Level of detail.** Start at a high level and go to more detail only if needed.
- **Scale.** Use for one-person few week project to a several-month project with project team. Use in your individual work.
- **Reuse (80/20 rule).** Bring together existing knowledge in such a way that you can understand it better. Supplement existing material with original research only as needed.
- **Tool independent.** Make better use of the tools you have.

Do's and Don'ts

- You don't have to have the CEO's support to get started
 - You DO have to have the appropriate level of management support to start while continuing to obtain management support as high up in the organization as possible
- You don't have to have all the answers
 - You DO need to do your homework, know your company, and be open to many options
- You don't need to do everything all at once
 - You DO need to have a plan of action and get started

Questions??

Thank you for attending and your participation!
Feel free to contact me if you have comments or questions:
danette@gfalls.com

THANK YOU!

Danette McGilvray
Granite Falls Consulting, Inc.
President and Principal Email: danette@gfalls.com
Phone: 510-501-8234 Web: www.gfalls.com
Fax: 510-505-9898 Fremont, California USA

Selected Resources (1)

Brassard, Michael and Diane Ritter. *The Memory Jogger II – A Pocket Guide of Tools for Continuous Improvement and Effective Planning*. GOAL/QPC. Methuen, MA. 1994.

English, Larry. *Improving Data Warehouse and Business Information Quality*. Wiley. New York. 1999. ISBN: 0471253839.

Eppler, Martin. *Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes*, Springer 2003. ISBN: 3540003983.

Hay, David C. *Requirements Analysis: From Business Views to Architecture*. Prentice Hall PTR, New Jersey. 2003.

Huang, Kuan-Tsae, Yang W. Lee and Richard Y. Wang. *Quality Information and Knowledge*. Prentice Hall. New Jersey. 1999.

Loshin, David. *Enterprise Knowledge Management: The Data Quality Approach*. Morgan Kaufmann. San Francisco. 2001. ISBN: 0124558402.

McGilvray, Danette. *Data Quality and the Project Life Cycle*. The Data Administration Newsletter – tdan.com. July 10, 2007. <http://www.tdan.com/view-articles/5092>

Selected Resources (2)

McGilvray, Danette. *Data Governance: A Necessity in an Integrated Information World, Part 1*. DM Review. December 2006.

http://www.dmreview.com/article_sub.cfm?articleId=1069951.

Part 2. DM Review. January 2007.

http://www.dmreview.com/article_sub.cfm?articleId=1072431

Olson, Jack E. *Data Quality: The Accuracy Dimension*. Morgan Kaufmann. San Francisco. 2003.

Redman, Thomas C.: *Data Quality: Management and Technology*, Bantam, New York, NY, 1992.

Redman, Thomas E. *Data Quality for the Information Age*. Artech House. Boston. 1996.

Redman, Thomas C. *Data Quality The Field Guide*. Digital Press. Boston. 2001.

Rummler, Geary A. and Alan P. Brache. *Improving Performance: How to Manage the White Space on the Organization Chart*. Jossey-Bass. San Francisco. 1990.