Ten Steps to Quality Data and Trusted Information™

ABSTRACT- -

Do these situations sound familiar? Your company is involved in a data integration project such as building a data warehouse or migrating several source systems into an ERP (Enterprise Resource Planning) application. Data quality issues are impacting the project timeline or users are distrustful of the information that is provided. Whether you are just starting the project or are already in production, it is not unusual to find that information quality issues prevent the company from realizing the full benefit of their investment in the new systems. Join us to hear of a methodology that applies to all kinds of organizations (for-profit businesses, education, government, healthcare, and nonprofits) and all types of data (finance, manufacturing, sales, marketing, human resources, etc.). Learn practical approaches and proven techniques to address your information quality challenges:

- Dimensions for measuring and managing data quality
- Techniques for assessing the business impact of information quality
- The Ten Steps to Quality Data and Trusted Information(tm)

BIOGRAPHY- -

Danette McGilvray
President and Principal
Granite Falls Consulting, Inc.

Danette McGilvray is President and Principal of Granite Falls Consulting, Inc., a firm specializing in information quality management and data governance and stewardship. Projects include establishing data quality and governance programs, data warehousing strategies, and large-scale ERP data migration for Fortune 500 organizations. She is the author of Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information(tm) (Morgan Kaufmann, 2008). Her background as a program and project manager, working with people in all levels and functional areas of the organization, gives her a perspective of information quality challenges based on real-life experience. Danette consults with clients and is an invited speaker in the US and Europe. She has contributed articles to various industry journals and newsletters and has been profiled in PC Week and HP Measure Magazine. She was an invited delegate to the People's Republic of China to discuss roles and opportunities for women in the computer field.

Granite Falls Consulting, Inc.

MIT Information **Q**uality **I**ndustry **S**ymposium

Ten Steps to Quality Data and Trusted Information™

MIT 2009 IQIS, July 15, 2009
Cambridge, Massachusetts
Tutorial Workshop
9:00 – 10:30 a.m.

Copyright Information

These materials, and any part thereof, are protected under copyright law. 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 *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™* (Morgan Kaufmann, 2008), by Danette McGilvray, 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

Do these situations sound familiar? Your company is involved in a data integration project such as building a data warehouse or migrating several source systems into an ERP (Enterprise Resource Planning) application. Data quality issues are impacting the project timeline or users are distrustful of the information that is provided. Whether you are just starting the project or are already in production, it is not unusual to find that information quality issues prevent the company from realizing the full benefit of their investment in the new systems. Join us to hear of a methodology that applies to all kinds of organizations (for-profit businesses, education, government, healthcare, and nonprofits) and all types of data (finance, manufacturing, sales, marketing, human resources, etc.). Learn practical approaches and proven techniques to address your information quality challenges:

- Dimensions for measuring and managing data quality
- Techniques for assessing the business impact of information quality
- The Ten Steps to Quality Data and Trusted Information™

Background

Danette McGilvray is President and Principal of Granite Falls Consulting, Inc., a firm specializing in information quality management and data governance and stewardship. Projects include establishing data quality and governance programs, data warehousing strategies, and large-scale ERP data migration for Fortune 500 organizations.

She is the author of *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™* (Morgan Kaufmann, 2008). Her background as a program and project manager, working with people in all levels and functional areas of the organization, gives her a perspective of information quality challenges based on real-life experience. Danette consults with clients and is an invited speaker in the US and Europe. She has contributed articles to various industry journals and newsletters and has been profiled in PC Week and HP Measure Magazine. She was an invited delegate to the People's Republic of China to discuss roles and opportunities for women in the computer field.

Assumptions About This Audience

You already know that:

- Information quality is important.
- We have the responsibility to give equal emphasis to the quality and management of the data and information as we do to the processes, people and organizations, technology, and other resources that support our businesses.

You are interested in:

- How to create and improve information quality within your organization.
- How to show others that data quality is important.

Processes

Data and Information

People and Organizations

Technology
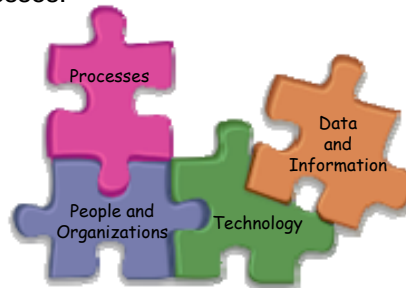
Business Needs …

What you might hear from the business - we need to:

- Have a 360 degree view of our customers
- Ensure that we will pass the next FDA audit
- Have information integrated from various business areas, so I can make informed decisions to better run our organization
- Know how many employees and contractors working at our company
- Trust the inventory numbers in our system
- Ship products on time and track them throughout the process
- Comply with government legislation
- Etc. Etc. Etc.
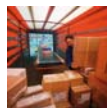
These needs get translated into some type of project or activity …

…Translated into Projects and Activities

Implement an Enterprise Resource Planning (ERP) application

Implement software for specific needs such as Customer Relationship Management (CRM) or Master Data Management (MDM)

Data clean-up project …. again

Build a data warehouse

Dealing with data is part of your daily responsibilities.

Develop new applications

Etc. Etc. Etc.

What they have in common – data:
- Data are migrated and integrated
- New data may have to be created
- Data may have to be purchased from external sources
- Existing data are often used for new purposes
- Data no longer meet requirements of the new situations

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 7

Your Information and Data Quality Challenges

Describe the situation where you are experiencing information and data quality challenges in your job and/or organization.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 8

Ten Steps to Quality Data and Trusted Information™

Your Action

Today's focus → **Ten Steps™ Process**

Key Concepts

Think of this as your "wellness" program for data and information.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 9

The Ten Steps™ Process

Today's focus →

| 1 Define Business Need and Approach | 2 Analyze Information Environment | 3 Assess Data Quality / 4 Assess Business Impact | 5 Identify Root Causes | 6 Develop Improvement Plans | 7 Prevent Future Data Errors / 8 Correct Current Data Errors | 9 Implement Controls |

10 Communicate Actions and Results

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 10

The Ten Steps™ Process

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.
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 are being assessed, and helps discover root causes. Design data capture and assessment plan.
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.
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**.**
5. **Identify Root Causes.** Identify and prioritize the true causes of the data quality problems and develop specific recommendations for addressing the root causes.
6. **Develop Improvement Plans.** Develop and execute improvement plans based on recommendations.
7. **Prevent Future Data Errors.** Implement solutions that address root causes of the data quality problems.
8. **Correct Current Data Errors.** Implement steps to make appropriate data corrections.
9. **Implement Controls.** Monitor and verify improvements that were implemented. Maintain improved results by standardizing, documenting, and continuously monitoring successful improvements.
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.

 11

Data Quality Dimensions

A data quality dimension is an aspect or feature 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.

 12

Data Quality Dimensions

From Step 3 – Assess Data Quality

Data Specifications	A measure of the existence, complete-ness, quality, and documentation of data standards, data models, business rules, metadata, and reference data.	**Ease-of-Use and Maintainability**	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.
Data Integrity Fundamentals	A measure of the existence, validity, structure, content and other basic characteristics of the data.	**Data Coverage**	A measure of the availability and comprehensiveness of data compared to the total data universe or population of interest.
Duplication	A measure of unwanted duplication existing within or across systems for a particular field, record, or data set.	**Presentation Quality**	A measure of how information is presented to and collected from those who utilize it. Format and appearance support appropriate use of information.
Accuracy	A measure of the correctness of the content of the data (which requires an authoritative source of reference to be identified and accessible).	**Perception, Relevance, and Trust**	A measure of the perception of and confidence in the quality of the data; the importance, value, and rele-vance of the data to business needs.
Consistency and Synchronization	A measure of the equivalence of infor-mation stored or used in various data stores, applications, and systems, and the processes for making data equivalent.	**Data Decay**	A measure of the rate of negative change to the data.
Timeliness and Availability	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.	**Transactability**	A measure of the degree to which data will produce the desired business transaction or outcome.

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   13

Assessments and Dimensions of Quality

Why differentiate the 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).
- 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.

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   14

Determining What to Assess

First:
- Understand the business issues driving the data quality assessment, improvement, and creation activities.

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/manage/maintain quality for those dimensions where the answer to both of the questions above is "yes."

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   15

Applying the DQ Dimensions In the Project Life Cycle (1)

Typical Project Life Cycle	Justification	Planning	Requirements and Analysis	Design	Development and Testing	Deployment	Post Production Support
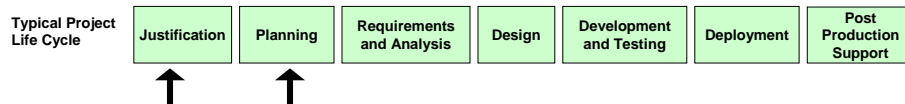
Justification:
- Survey knowledge workers for their perception, relevance, and trust of the data quality and use as input to justify data quality activities in a project.
- Use Business Impact Techniques (discussed in Step 4 – Assess Business Impact) to show how data quality issues are affecting the business and why they should be addressed.

Planning:
- Quick, high-level data profiling provides insight and allows you to take into account data quality issues that can impact the timeline and resources.

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   16

Applying the DQ Dimensions In the Project Life Cycle (2)

Typical Project Life Cycle	Justification	Planning	Requirements and Analysis	Design	Development and Testing	Deployment	Post Production Support
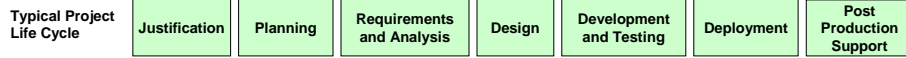
Requirements and Analysis:
- Gather data quality requirements using the dimensions at a high level as a guide.
- Conduct in-depth data profiling to learn the basics about your data. Profiling results will expedite source-to-target mappings, provide input to transformation rules, help when choosing data sources, confirm selection criteria, and help you understand the quality of purchased data.
- Analyze applicable data specifications (data standards, data model, business rules, metadata, and reference data) in both the source and target databases.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 17

Applying the DQ Dimensions In the Project Life Cycle (3)

Typical Project Life Cycle	Justification	Planning	Requirements and Analysis	Design	Development and Testing	Deployment	Post Production Support
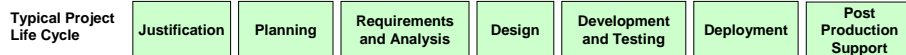
Design:
- Profile existing data to provide input into the creation of new models that support data to be moved to a target system. Profiling also exposes structural differences between an existing target model and the source data to be moved.
- Ensure the design incorporates the data quality requirements.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 18

Applying the DQ Dimensions In the Project Life Cycle (4)

Typical Project Life Cycle	Justification	Planning	Requirements and Analysis	Design	Development and Testing	Deployment	Post Production Support
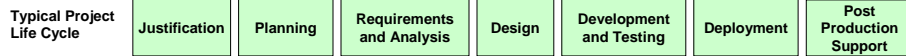
Development and Testing:
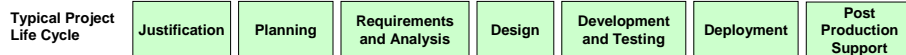
- Continue to check and ensure the quality of all data, including reference data, metadata, created data, and purchased data.
- Continue to correct current data errors based on the dimensions of interest throughout the project cycle.
- Profile and check data prior to and after test loads.
- Update transformation rules, business rules, and other data specifications as necessary based on test results.
- When first testing for software quality assurance, profile the test data and ensure that their content is known. Much time can be spent chasing suspected software problems when the problem is actually in the test data. Controlling the test data well allows the team to focus their efforts on software functionality if issues are found.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 19

Applying the DQ Dimensions In the Project Life Cycle (5)

Typical Project Life Cycle	Justification	Planning	Requirements and Analysis	Design	Development and Testing	Deployment	Post Production Support

Deployment:

- Use quick data quality assessments to confirm that data extracts are correct prior to the final data loads at go-live.
- Conduct quick data quality assessments after the data loads and before the system is released to the users.

Post Production Support:

- Institute appropriate ongoing monitoring and metrics to check data quality and provide the ability to take quick action if needed.
- Make use of any data quality assessments done during the project and modify the processes as needed to test the data quality on a periodic basis.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 20

The Value of Information

What is worth saving?

What is the most difficult/costly to replace? The desk, the chair, or the information in the computer?

We intuitively know the answers to these questions.

So why do we meet resistance when asking for time, money, and human resources to deal with data and information?

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   21

Showing the Value and Impact to Business

What impact do data quality and governance have on the organization? On me? On my responsibilities? Why do they matter?

- These are the right questions!
- But historically they have been difficult to answer.
- There are techniques to help assess business impact (value) of information and data quality.

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   22

Why Assess Business Impact?

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

Consider assessing business impact if you hear:

- "Why are information quality and data governance important?"
- "Why should I spend time cleaning data or preventing data issues?"
- "Why should I put data governance into place?"
- "What's in it for me (my organization, my role, my responsibilities)?"
- Words of support, but see no action.

Business Impact Techniques

From Step 4 – Assess Business Impact

Techniques for assessing the impact of data quality on the business

|		
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** —————►

14

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. All techniques can be effective:
  - You can understand business impact even without completing a full cost/benefit analysis.
  - Less complex 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

Less Time/ Less Complex   More Time/ More Complex

       1 2   3   4   5 6   7   8
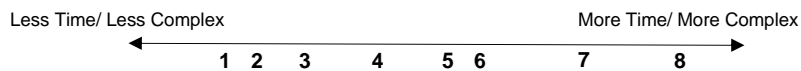
MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   25

When to Assess Business Impact?

The need to show business impact can happen any time:

- During an **information-quality focused** project
- In the course of **daily work** where you have responsibility for managing data quality/governance or the work you do impacts data quality/governance.
- When trying to include data-related activities in **other projects and methodologies** (e.g. ERP migration or building a data warehouse)

Typical Project Life Cycle	Justification	Planning	Requirements and Analysis	Design	Development and Testing	Deployment	Post Production Support
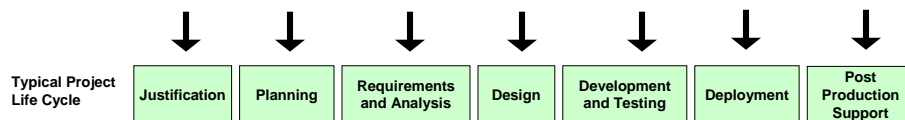
MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   26

Typical Data Quality Project Approaches (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 Cause: 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.
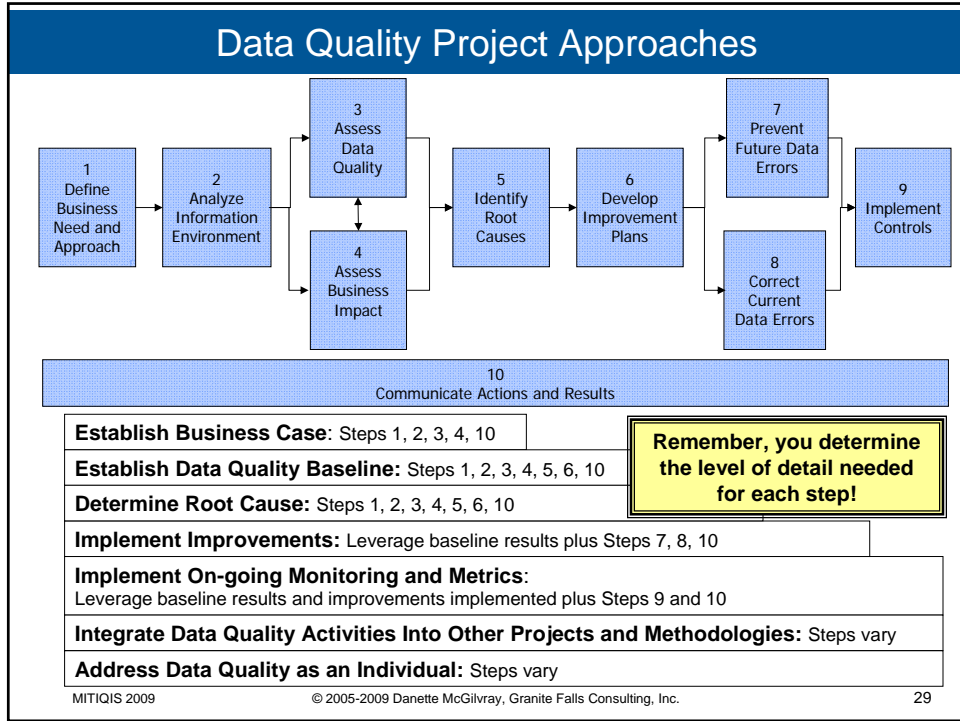
Implement Improvements: Executes the recommendations developed when the data quality assessment and business impact analysis generated a plan for data quality improvement.

Typical Data Quality Project Approaches (2)

Implement On-going Monitoring and Metrics: Focuses on instituting operational processes for monitoring, evaluating, and reporting results. When designing and implementing your control process, include actions for addressing issues found – both to correct current errors and prevent future errors.

Integrate Data Quality Activities Into Other Projects and Methodologies. Combine the Ten Steps techniques with your company's favored project management and project life cycle, and your specific project plan. E.g. data profiling to expedite source-to-target mappings, analyze source and target databases, etc.

Address Data Quality as an Individual. In the course of daily work where you have responsibility for managing data quality. Use when data quality is an important part of or impacts your job responsibilities. Apply various techniques from the Ten Step methodology in your daily activities or to address a specific data quality issue.

Data Quality Project Approaches

Establish Business Case: Steps 1, 2, 3, 4, 10

Establish Data Quality Baseline: Steps 1, 2, 3, 4, 5, 6, 10

Determine Root Cause: Steps 1, 2, 3, 4, 5, 6, 10

Remember, you determine the level of detail needed for each step!

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

Integrate Data Quality Activities Into Other Projects and Methodologies: Steps vary

Address Data Quality as an Individual: Steps vary

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   29

Your Next Steps

How does this apply to you and your situation?

What are *your* next steps?

- This week
- Next Monday
- Next month

MITIQIS 2009   © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc.   30

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.

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 31

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

MITIQIS 2009 © 2005-2009 Danette McGilvray, Granite Falls Consulting, Inc. 32

Want More Information?

- For details on the methodology (concepts and the Ten Steps process, bibliography for additional references) see *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™*, by Danette McGilvray, (Morgan Kaufmann Publishers, Copyright 2008 Elsevier Inc.) Available at Amazon.com or your favorite bookseller.
- For consulting, training, and management presentations contact Danette at danette@gfalls.com.
- For additional resources see http://tensteps.gfalls.com
  - Downloadable pdfs of the Framework for Information Quality, the data quality dimensions, business impact techniques, and other concepts
  - Templates described in the book
  - Additional web sites

Thanks for your time today.
Please feel free to contact me with questions or comments.

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