

Case Study: How Data Quality Has Evolved at MathWorks

ABSTRACT

Getting a Data Quality program off the ground can be a difficult task. Where to start? How to break down the work to a manageable level? How to get people interested in improving data? How to build the case to expand the Data Quality program? The presentation will provide an overview of how Data Quality at MathWorks evolved from an informal program to one with management support and backing. The presentation will provide a look at the data challenges facing MathWorks, summarize two phases of the Data Quality Program, outline key factors for increasing the visibility and importance of Data Quality, and will provide lessons learned during the process.


BIOGRAPHY

Kirk Amidon

Marketing Data Analyst
MathWorks, Inc.




Kirk Amidon has been with MathWorks for 13 years, the last 5 of which he has worked as a Marketing Data Analyst. He works mainly with MathWorks' customer data (accounts and contacts) and became interested in data quality when his job was adversely impacted by poor data quality. Kirk is passionate about the positive impact Data Quality can have on a business, and has worked with two data quality teams over the past 3 years. Prior to working at MathWorks, Kirk worked in the Hotel/Casino industry in the Lake Tahoe area for 5 years. He has a BA in Economics from the University of Massachusetts, Amherst.



Case Study: Obtaining Executive Sponsorship: How Data Quality has evolved at MathWorks.

Kirk Amidon
MathWorks, Inc

© 2011 The MathWorks, Inc.



Agenda

- Introductions
- Summary of Data Quality approach
- 6 Keys to starting a Data Quality program
- Next Steps/Final thoughts

Goal: Present a retrospective of my experiences with starting a Data Quality program. Provide real world examples of things that went well, and highlighting lessons learned.

2



MathWorks Vital Statistics

Developers of MATLAB & Simulink

2,200 staff worldwide

Support staff worldwide

Development staff in Natick, MA

30% of revenue invested in R&D

\$500M annual revenue

*2009 - orders from
23,000 companies
in 128 countries*




Intro

Currently: Sr. Marketing Analyst

Primary Focus: Customer Data

With MathWorks for 13 years


Interest in Data Quality started due to constantly running into data issues during analysis.



Where to Start?

User Perception of Data	Lots of duplicates
“not my job”	Difficult to get data fixed
No DQ advocate	Silo'ed processes

5



First Phase Summary

6



Landscape

In 2007, MathWorks was not ready for **formal** Data Quality

And - a couple past efforts failed

But - several groups were kicking tires

We decided to start small

7



First Phase


2007 - 2009

- Consolidated two small teams
 - Cross-functional (Sales, Mktg, IT, Service)
 - Met weekly
- Initial exploration
 - Is there really an issue? How big is it?
 - Focused on firefighting (Reactive)
 - Took on many smaller issues
- Created a steering team
 - Managers of impacted areas
 - Met 2x annually to discuss progress



Challenge: Team did not have enough authority to mandate real change-many of the issues felt “too big”

8



First Phase

Accomplishments

- ✓ Identified issues
- ✓ Quantified costs
- ✓ Developed fixes for some issues
- ✓ Created short and long term goals/plans
- ✓ Issues boiled up to departmental reviews


+

Good Timing

Lean economy led to company wide waste reduction strategy.

Result: Company decided to formally work on Data Quality.

9



Second Phase Summary

10



Second Phase

2010-ongoing

- Driven by Business Model Team, with executive sponsorship
 - Involved the key people from initial phase
 - Focus on larger issues – root causes

- Took big step back to analyze customer data
 - Business had drastically changed over past 10-15 years
 - Systems and process had not caught up to business

- Leveraged 6-Sigma (DMAIC) resources

11



Second Phase: Initial Findings



- Duplicates are biggest issue

- Data collected, but not used immediately

- Insufficient controls around data lifecycle
 - Volume of new data overwhelms systems
 - 90% of data comes from web forms
 - No end of life strategy

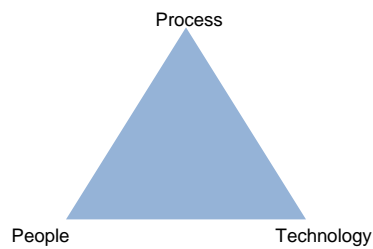
12



Second Phase: Initial Recommendations

Get control Data

- Add controls on web forms
- Identify returning visitors
- Qualify Account data prior to creation
- Identify old data with no value



Focus on Processes

- Root cause of problems
- Tight Economy
- Controllable

13



Key Themes for Obtaining Sponsorship

14



6 Themes for Obtaining Sponsorship

Establish Goals	Limit the scope
Get right people involved	Communication
Metrics	Pick an Approach

15



6 Keys for Obtaining Sponsorship

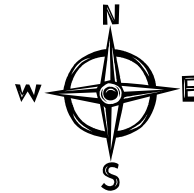
Establish Goals	Limit the scope
Get right people involved	Communication
Metrics	Pick an Approach

16



Establish Goals

- Establish the current state of Data Quality
- Determine the ultimate goal
- Build a roadmap
 - Establish and document 5 year and 3 year goals, with 1 year objectives supporting goals
 - Longer goals should be strategic in nature
 - Ensure all projects you start support the plan
- Be realistic in what can be accomplished
 - Everything takes longer than expected



“if you don’t know where you are going, you will wind up somewhere else”
 -Yogi Berra



Sample Goals

5 Yr Goal	3 Yr Goal	1 Yr Objective
Use Metrics to Guide Data Quality Efforts	Our underlying DQ metrics are relevant to the business	Understand how the noise in the data impacts our ability to measure
Implement waste reduction strategies	Methods are in place at the point of collection to increase accuracy and depth of Data.	Email Address Validation
Raise awareness of Data Quality	Obtain buy in from Senior management that Data Quality is an area of importance critical to MathWorks' success	Deliver a Director level presentation outlining current state of data quality, focusing on costs and waste reduction
Incorporate Industry Best Practices Improve the approach to Data Quality	Team is continually involved in industry organizations	Attend the MIT DQ conference





6 Keys for Obtaining Sponsorship



19



Scope Limitation

Tools used to identify the In-scope Data:

- Core Data Analysis
 - Analyzed all data elements across 28 business processes
 - Meetings with all business areas
- Survey
 - Surveyed 600 users of customer data
 - Used both Quantitative and Qualitative questions
- Data Segmentation
 - Found >90% of our data entered via the web
 - Two channels of web data: >75% of duplicates caused by one channel

Don't try to boil the ocean

20



Scope Limitation

We knew Data Quality was a problem, but where to start?



Decision: focus on Core Customer Data

A core customer attribute is:

- An attribute of a account or contact
- Used by multiple business areas and processes throughout the customer lifecycle.
- Needed to run our business (not just a “nice to have”).

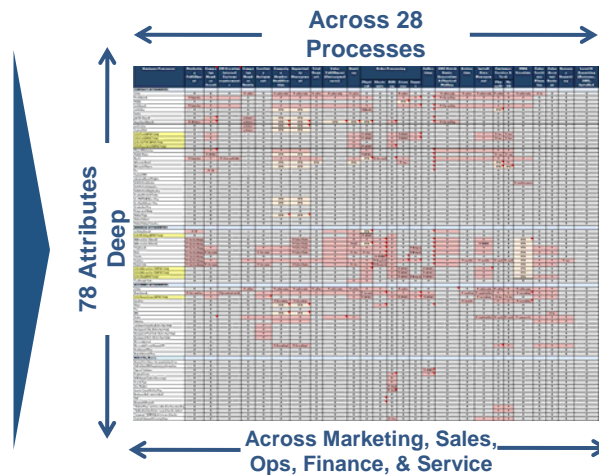
21



Scope limitation: Core Attributes

Approach:

- Each business area defined which attributes are needed at each stage of the customer lifecycle



Result: Limited the scope from 78 attributes down to 13!

22



Scope limitation - Survey

Surveyed 600 users of customer data

Quantitative

Qualitative

Q There are many duplicate records in our systems.

What are the top 3 significant issues?

A 75% indicated "Strongly Agree"

"I tend to try to work around the fragmentation rather than spend time aligning the contacts under one company name."

Qualitative responses can provide tremendous insight!

23



6 Keys for Obtaining Sponsorship

Establish Goals

Limit the scope

Get right people involved

Communication

Metrics

Pick an Approach

24



Involve the right people

- Find people with vested interests in improving Data Quality
 - People who have complained about data getting in their way
 - People looking for a better way
- Empower those who feel the pain
 - Don't have to be part of the core team
 - Test ideas or processes
 - Can advocate in their department
- Beware of “champions of the obscure” – worried about the wrong end of the 80-20 rule

25



Involve the right people

Cross departmental

- Need an approach that crosses boundaries
- All groups need to be represented

Types of people

- Change agents
- Innovators
- Risk takers

Roles Needed

- Sponsor
- Facilitator/Project manager
- Analyst(s)
- Subject Matter Experts
- Business Area Experts

The team should be as autonomous as possible – avoid trap of Data Quality being viewed as belonging to a single business area.

26



6 Keys for Obtaining Sponsorship



27



Communication

- Establish Communication plans
 - Core Team
 - Extended Team
 - Company Wide
- Promote how DQ is the solution
 - Have several “elevator speeches”
 - Build a 15 min presentation. Shop it around
 - Promote your successes
 - Write a blog
- Keep expanding circle of influence



Know your Audience – customize your message to suit the audience!

28



Communication Examples

How to bridge the gap?



- Conversation with VP of Marketing
- Summarize the top 'x' things learned here– send to key people

Created the burning bridge

29



6 Keys for Obtaining Sponsorship

Establish Goals

Limit the scope


Get right people involved

Communication

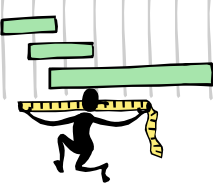
Metrics

Pick an Approach

30



Metrics




- Tie Metrics to Business Issues
 - Highlight inefficiencies
 - Provide context

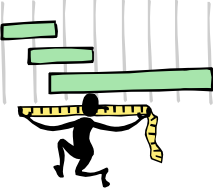
Metric	“sales reps spent 60 hours on territory planning due to duplicate data”	“we have a 20% duplicate rate”
Attributes	<ul style="list-style-type: none"> ➢ Clear ➢ Actionable 	<ul style="list-style-type: none"> ➢ Lacks Context ➢ No clear action
Call to Action	“Do Something!”	“So What?”

Data Quality is a death by 1000 cuts – no one impact is fatal on its own, but in aggregate they are.

31



Metrics



- Have to be believable
 - Document your metrics
 - Be transparent
- Don't try to detail all of the costs
 - Common estimate is 10-20% of revenue lost to poor quality.

Good Enough <u>is</u> Good Enough		
\$5 -10 <u>MILLION</u> in waste due to bad data	VS	Easy to Quantify and Document 500k (0.1%)

Easier to get backing with a believable figure – even if the 'truth' is closer to the 10-20%

32



Overcoming Roadblocks

- Roadblocks will be created by
 - Champions of the Obscure
 - People who resist change
 - Fear of unknown
- Metrics can be used to overcome them
 - Many times the “common belief” is wrong



Proposed Business Rule:

Email Address should be a unique identifier.

Roadblock:

“we can't do it, some customers need to share addresses”

Solution: Initial analysis showed about 3% of contacts shared email addresses. Detailed analysis showed 99.9% of these were probable duplicate records.

33



6 Keys for Obtaining Sponsorship

Establish Goals

Limit the scope

Get right people involved

Communication

Metrics

Pick an Approach

34

Pick an Approach

- All approaches work, if applied
 - DMAIC – PDCA - TIQDM
- Find approach that fits culture
 - Don't force a poor fit
 - Try several on
 - Leverage existing tools where possible
- Good approaches are consistent, repeatable, and formal



35

Why DMAIC

- Software Development Teams already using 6-Sigma tools and processes
- Internal resources available
- “language” embedded in company culture
- Parallels between Software Quality and Data Quality



36



6 Keys for Obtaining Sponsorship

Establish Goals

Limit the scope

Get right people involved

Communication

Metrics

Pick an Approach

37



**So you have Executive Sponsorship
(now what?)**

38



Post Exec Approval

- Not out of the woods yet – but you have permission to play in the sandbox
- Expanding the scope brings you back to square one
 - Opportunity to tackle issues on a much larger scale
- New people will be exposed to DQ for the first time.
 - Important to let them “catch up” to DQ ideas/practices
 - Don’t be discouraged by different reactions to old ideas
- The 6 keys all still apply
 - Many will need to be adjusted or revised

39



Post Exec Approval

- You will run into more “Fear of the Unknown” as projects commence
- There will be times where compromise is needed – the business may not be ready for the ideal solution from a DQ perspective.
- Don’t be afraid to call in the bigger hammer when faced with roadblocks
- As projects are Identified and Running, keep lines of communication open with sponsors.

40



An Incomplete Journey

Much more to do

- Establish long term responsibilities
- Work on next level of priority
- Determine approach of Governance
- Keep the momentum

41



Post Mortem/Lessons Learned

- Always give previews of findings individually prior to a large meeting – avoid landmines
- Ensure all team members are communicating back to their groups.
- Resist urge to go to “solution mode” immediately, spend time exploring the actual problem.
- Ensure Metrics are used to make “data – driven” decisions
- Don't position DQ as a “project”. Managers are used to thinking in project terms (completion dates, duration, etc)

“I thought we addressed Data Quality already” - a Sr. Managers' reaction – recalling purchasing a Matching Tool 2 years prior.

42