DATA QUALITY & THE DMBOK

DONNA BURBANK
VP, INFORMATION MANAGEMENT SERVICES

DONNA.BURBANK@ENTERPRISEARCHITECTS.COM
TWITTER: @DONNABURBANK

DAMA BRASIL
SEPT 2014

www.enterprisearchitects.com
Agenda

› DMBOK Overview

› Data quality management

› Benefits and impacts of data quality

› Activities relevant to data quality management

› Data quality management maturity assessment
Introduction to Donna Burbank

› More than 20 years of experience in the areas of data management, metadata management, and enterprise architecture.
  » Currently VP of Information Management Services at Enterprise Architects
  » Brand Strategy, Product Management, and Product Marketing roles at CA Technologies and Embarcadero Technologies designing several of the leading information management products in the market today
  » Senior consultant for PLATINUM technology’s information management consulting division in both the U.S. and Europe.
  » Worked with dozens of Fortune 500 companies worldwide in the U.S., Latin America, Europe, Asia, and Africa and speaks regularly at industry conferences.
  » President of DAMA Rocky Mountain Chapter
  » Co-author of several books including:
    - Data Modeling for the Business
    - Data Modeling Made Simple with CA ERwin Data Modeler r8

Twitter: @donnaburbank
And When I’m Not Doing Data Management...

Pão de Açúcar, Rio de Janeiro
What Is the DAMA-DMBOK Guide?

› The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK Guide)

› A book published by DAMA-I, 406 pages (also on CD & PDF)

› Available from TechnicsPublications.com or Amazon.com

› Written and edited by DAMA members

› An integrated primer: “definitive introduction”

› Modeled after other BOK documents:
  » PMBOK (Project Management Body of Knowledge)
  » SWEBOK (Software Engineering Body of Knowledge)
  » BABOK (Business Analysis Body of Knowledge)
  » CITBOK (Canadian IT Body of Knowledge)
DAMA-DMBOK Guide Goals

› To develop, build consensus and foster adoption for a generally accepted view of data management.

› To provide standard definitions for data management functions, roles, deliverables and other common terminology.

› To identify “guiding principles”.

› To introduce widely adopted practices, methods and techniques, without references to products and vendors.

› To identify common organisational and cultural issues.

› To guide readers to additional resources.
DAMA Framework

Functions
DMBoK Functions
Brief History of the DMBOK

› First Publication of the DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBoK Guide) 2009

› March 2010 - DAMA-DMBOK hardcopy version

› 2011 – version 2 DAMA Dictionary of Data Management

› 2011 Japanese version

› 2012 Portuguese version

› 2012 Chinese version

› April 2012 – DAMA-DMBOK2 Framework

› Q1 2015 – DMBoK2 Publication expected
Data Quality Management In Context

EA's Information Management Reference Architecture

DQ READINESS & MATURITY

<table>
<thead>
<tr>
<th>INITIAL</th>
<th>LEVEL 2 - REFORMABLE</th>
<th>LEVEL 3 - ENHANCED</th>
<th>LEVEL 4 - ACHIEVED</th>
<th>LEVEL 5 - OPTIMIZED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low awareness of the importance of data quality, and no common understanding of key aspects.</td>
<td>There is some awareness of the importance of data quality, and understanding is spreading across the business.</td>
<td>Data governance is driving the same work, but there is no overarching view.</td>
<td>Everyone in the business recognizes the critical importance of data quality and has the knowledge and data quality improvement.</td>
<td>Everyone in the business understands the critical importance of data quality and takes a proactive approach to ensuring data quality.</td>
</tr>
<tr>
<td>Data quality initiatives are scattered across the business. Data quality metrics are not consistently correlated to business success.</td>
<td>Data quality initiatives are increasingly coordinated and aligned to business success. Data quality metrics are increasingly correlated to business success.</td>
<td>Data Quality Champions are driving initiatives that improve data quality and contribute to the wider business strategy.</td>
<td>A Data Quality Champion is driving initiatives that improve data quality and contribute to the wider business strategy.</td>
<td>A Data Quality Champion drives initiatives that improve data quality and contributes to the wider business strategy and the Community of Interest.</td>
</tr>
<tr>
<td>Data quality as a priority is not aligned with business strategy.</td>
<td>Data quality is viewed as a strategic asset.</td>
<td>Data Quality Champions take on a lead role in ensuring that business strategy is aligned with data quality.</td>
<td>A Data Quality Champion takes on a lead role in ensuring that business strategy is aligned with data quality.</td>
<td>A Data Quality Champion drives initiatives that improve data quality and contribute to the wider business strategy and the Community of Interest.</td>
</tr>
</tbody>
</table>

INFORMATION MANAGEMENT READINESS ASSESSMENT
What is Data Quality Management?

› Poor Data Quality Management is not equate to poor data quality

› But when you don’t have good Data Quality Management...

  » The current level of data quality will be unknown

  » Maintaining a sufficient level of data quality will be a result of lots of hard work and extra effort from staff

  » The risk to the business will increase

› It is infinitely more sensible to ensure good data quality by having good management through a coherent set of policies, standards, processes and supporting technology

“Data errors can cost a company millions of dollars, alienate customers, suppliers and business partners, and make implementing new strategies difficult or even impossible.

The very existence of an organisation can be threatened by poor data”

Joe Peppard – European School of Management and Technology

“Ultimately, poor data quality is like dirt on the windshield. You may be able to drive for a long time with slowly degrading vision, but at some point you either have to stop and clear the windshield or risk everything”

Ken Orr, The Cutter Consortium
10. Data Quality Management

**Definition:** Planning, implementation, and control activities that apply quality management techniques to measure, assess, improve, and ensure the fitness of data for use.

**Goals:**
- To measurably improve the quality of data in relation to defined business expectations.
- To define requirements and specifications for integrating data quality control into the system development lifecycle.
- To provide defined processes for measuring, monitoring, and reporting conformance to acceptable levels of data quality.

**Inputs:**
- Business Requirements
- Data Requirements
- Data Quality Expectations
- Data Policies and Standards
- Business Meta-data
- Technical Meta-data
- Data Sources and Data Stores

**Suppliers:**
- External Sources
- Regulatory Bodies
- Business Subject Matter Experts
- Information Consumers
- Data Producers
- Data Architects
- Data Modelers
- Data Stewards

**Participants:**
- Data Quality Analysts
- Data Analysts
- Database Administrators
- Data Stewards
- Other Data Professionals
- DRM Director
- Data Stewardship Council

**Tools:**
- Data Profiling Tools
- Statistical Analysis Tools
- Data Cleansing Tools
- Data Integration Tools
- Issue and Event Management Tools

** Metrics:**
- Data Value Statistics
- Errors / Requirement Violations
- Conformance to Expectations
- Conformance to Service Levels

**Primary Deliverables:**
- Improved Quality Data
- Data Management
- Operational Analysis
- Data Profiles
- Data Quality Certification Reports
- Data Quality Service Level Agreements

**Activities:**
1. Develop and Promote Data Quality Awareness (O)
2. Define Data Quality Requirements (D)
3. Profile, Analyze, and Assess Data Quality (D)
4. Define Data Quality Metrics (P)
5. Define Data Quality Business Rules (P)
6. Test and Validate Data Quality Requirements (D)
7. Set and Evaluate Data Quality Service Levels (P)
8. Continuously Measure and Monitor Data Quality (C)
9. Manage Data Quality Issues (C)
10. Clean and Correct Data Quality Defects (O)
11. Design and Implement Operational DQM Procedures (D)
12. Monitor Operational DQM Procedures and Performance (C)

**Activities:**
(P) – Planning (C) – Control (D) – Development (O) – Operational
1. Develop & Promote Data Quality Awareness

Part of Your Job is Marketing!

› Promoting and evangelising the importance of data quality as early as possible will improve the chances of success of any Data Quality programme

› This needs to happen at all levels within the organisation, from senior management and key stakeholders down to users and operational staff

› Setting up a Data Quality Community of Interest can help create a common understanding and provide a forum for sharing knowledge and best practice

› Data Quality Management cannot survive without ownership and accountability, so close alignment with the Data Governance programme is essential

Data Governance is Key

› Engage business partners who will work with the data quality team and champion the DQM program

› Identify data ownership roles and responsibilities, including data governance board members and data stewards

› Assign accountability and responsibility for critical data elements and DQM

› Identify key data quality areas to address and directives to the organisation around these key areas

› Synchronise data elements used across the lines of business and provide clear, unambiguous definitions, use of value domains, and data quality rules

› Continuously report on the measured levels of data quality

› Introduce the concepts of data requirements analysis as part of the overall system development life cycle

› Tie high quality data to individual performance objectives
A street in Preston has been branded the most crime-ridden in England and Wales by a government website.

Glovers Court - a quiet street in Preston city centre - has been the location for 150 offences in December, according to the online map.

But police say the crimes were actually committed across the whole of Preston city centre.

Ch Supt James Lee said only three crimes had been reported at Glovers Court.

The figures are part of a new online crime map, set up by the Home Office, which cost £300,000 to develop.

Visitors to the website are able to find out which crimes have taken place on or near their street within the past month and which officers are responsible for their area.

The street with the most recorded crimes is Glovers Court - but Ch Supt Lee said: “The figures don't do it justice, it is actually a safe place to be.

“I don't accept these figures. The postcode relates to the whole of the city.

A new crime-mapping website for England and Wales is experiencing a "temporary problem" as millions of people log on every hour, the Home Office has said.

Hundreds have contacted the BBC website to report problems accessing the site as officials worked to fix the glitch.

The Home Office said www.police.uk was receiving up to five million hits an hour, or some 75,000 a minute.

The site allows you to see the offences reported in your local street by entering a street name or postcode.

Home Secretary Theresa May said the maps would give real facts on crime and anti-social behaviour and make police more accountable.

"Complete farce"

In a message on the microblogging website Twitter, the Home Office said: "Hugely popular street-level crime maps getting 75,000 hits per minute so you might experience delays. Keep trying."

But some users have reported seeing only a blank page when visiting the website, or a message saying "no police area is associated with this address" when entering a postcode or street.

Others have complained of errors in the actual information shown, with some "quiet streets" next to commercial centres, bars and clubs being tarred with their crimes.

A spokesman for charity Victim Support said it was important that victims of crime had consented as to whether information about their incident was released.
Benefit and Impact

**Good data quality benefit**

- Adherence to corporate & Regulatory acts
- Improved confidence in Data
- Reduced “busy work” in data archaeology
- Enriched Customer Satisfaction
- Better decision making
- Effective Marketing and Advertising
- Cost efficiencies
- Improved Operational Efficiency & streamlining

**Poor data quality impact**

- Ineffectual Advertising & Marketing
- Reputational damage
- Diminished Regulatory Compliance
- Decrease in Customer Satisfaction
- Uneconomical Business Processes
- Compromised Health, Safety & Security
- Erratic Business Intelligence
- Amplified Corporate Risk
- Impaired Business Agility
2. Define Data Quality Requirements

- Data Quality can only be considered within the context of the intended use of the data - i.e. fitness for purpose.

- The required level of Data Quality for a particular data component is therefore dependent on the collection of business processes that interact with the component.

- These in turn are driven by the underlying business policies, which are ultimately the source of many Data Quality requirements.

- Determining fitness for purpose requires reporting on meaningful metrics associated with well-defined data quality dimensions.

**Deriving Data Quality Requirements**

- Identify key data components associated with business policies.

- Determine how identified data assertions affect the business.

- Evaluate how data errors are categorized within a set of data quality dimensions.

- Specify the business rules that measure the occurrence of data errors.

- Provide a means for implementing measurement processes that assess conformance to those business rules.
How good does data quality need to be?

Fitness for Purpose

In February 2011, the UK government launched a crime-mapping website for England and Wales (www.police.uk).

Unfortunately, for a number of reasons, the postcode allocated to a specific police incident didn’t always correspond to the precise location of the crime.

The net result was that poor accuracy in the recording of geographical information led many quiet residential streets to be incorrectly identified as crime hotspots.

Data quality can only be considered within the context of the intended use of the data

Data needs to be “fit for purpose”

Data quality needs to be assessed on that basis.
How good does data quality need to be?

Fitness for Purpose

**Bad systems design can cost companies millions.**

One pharmaceutical company had five main UK manufacturing centres, each with its own warehouse of spare parts for the machines in the factories.

In theory, all five sites shared a common system, so spare parts -- 65,000 inventory items in all -- could be ordered from another location.

But in reality, the system was hard to use, so each of the separate sites built up its own inventory of spare parts sufficient for its needs. More than sufficient, in fact:

After a data cleanup, it was discovered that the company had enough spare parts to last 90 years in some cases.

*Data quality can only be considered within the context of the intended use of the data (data needs to be “fit for purpose” and data quality needs to be assessed on that basis)*

---

**In the context of managing the risk of machine downtime this is acceptable.**

---

**However, with the holistic view of the cost of spare parts, this is ridiculous.**

---

Data fit for purpose

Data not fit for purpose
Dimensions of Data Quality

› **Validity**– Conforms to the syntax (format, type, range) of its definition. Database, metadata or documentation rules as to the allowable types (string, integer, floating point etc.), the format (length, number of digits etc.) and range (minimum, maximum or contained within a set of allowable values).

› **Accuracy**– Data correctly describes the "real world" object or event being described. Does it agree with an identified reference of correct information?

› **Reasonableness**– Does the data align with operational context, e.g. birthdate of 01/01/01 is valid, but is it reasonable?

› **Completeness**– Certain attributes always have assigned values. Business rules define what "100% complete" represents.

› **Consistency**– Values in one data set are consistent with values in another data set.

› **Currency**– Data is current and “fresh”. Data lifecycle is important here.

› **Precision**– Level of detail of the data element, e.g. number of significant digits in a number. Rounding, for example, can introduce errors.

› **Privacy**– Need for access control and usage monitoring.

› **Referential Integrity**– Constraints against duplication are in place (e.g. foreign keys in a RDMBS)

› **Timeliness**– The time between when data is expected and when it is available for use.

› **Uniqueness**– No value occurs more than once in the data set.
Six Dimensions of Data Quality

Many use a subset of these dimensions. DAMA UK suggests six.

- **Completeness**—The proportion of stored data against the potential of "100% complete". Business rules define what "100% complete" represents.
- **Uniqueness**—No thing will be recorded more than once based upon how that thing is identified. The Data item measured against itself or its counterpart in another data set or database.
- **Timeliness**—The degree to which data represent reality from the required point in time. The time the real world event being recorded occurred.
- **Validity**—Data are valid if it conforms to the syntax (format, type, range) of its definition. Database, metadata or documentation rules as to the allowable types (string, integer, floating point etc.), the format (length, number of digits etc.) and range (minimum, maximum or contained within a set of allowable values).
- **Accuracy**—The degree to which data correctly describes the "real world" object or event being described. The degree to which data correctly describes the "real world" object or event being described.
- **Consistency**—The absence of difference, when comparing two or more representations of a thing against a definition. The absence of difference, when comparing two or more representations of a thing against a definition.

Source: DAMA UK
The Information Lifecycle

Information Lifecycle (DAMA)

- **PLAN**
  - IM strategy
  - Governance
  - Define policies and procedures for quality, retention, security etc

- **SPECIFY**
  - Architecture
    - Conceptual, logical and physical modelling
  - Install or provision servers, networks, storage, DBMSs
  - Access controls

- **ENABLE**
  - Data created, acquired (external), extracted, imported, migrated, organised

- **CREATE & ACQUIRE**
  - Data validated, edited, cleansed, converted, reviewed, reported, analysed

- **MAINTAIN & USE**
  - Data archived, retained and retrieved

- **ARCHIVE & RETRIEVE**
  - Data deleted

- **PURGE**

(Source: DAMA)
Why Data Quality Profiling?

› Reviewing and refining business policies provides a “top down” view of Data Quality requirements, but a “bottom up” view is crucial to identify existing issues within the data

› This is achieved through an activity known as Data Quality Profiling

› To conduct Data Quality Profiling as efficiently and repeatedly as possible, a specialist DQ tool is normally employed

› The result is an invaluable insight into the real operational data, revealing hidden characteristics, patterns and anomalies
Typical Outputs of Data Quality Profiling

- **COLUMN PROFILING**
  - Record count, unique count, null count, blank count, pattern count
  - Minimum, maximum, mean, mode, median, standard deviation, standard error
  - Completeness (% of non-null records)
  - Data type (defined vs actual)
  - Primary key candidates

- **FREQUENCY ANALYSIS**
  - Count/percentage each distinct value
  - Count/percentage each distinct character pattern

- **PRIMARY/FOREIGN KEY ANALYSIS**
  - Candidate primary/foreign key relationships
  - Referential integrity checks between tables

- **DUPLICATE ANALYSIS**
  - Identification of potential duplicate records (with variable sensitivity)

- **BUSINESS RULES CONFORMANCE**
  - Using a preliminary set of business rules

- **OUTLIER ANALYSIS**
  - Identification of possible out of range values or anomalous records
Metrics provide a common baseline

› Without common metrics, it’s difficult to define “how good is data quality”
4. Define Data Quality Metrics

“You cannot manage what you cannot measure.”

Defined metrics should be used to assess data quality using data quality indicators (DQI):

- **Measurability** – Can be measured and quantified within a discrete range
- **Business Relevance** – Measures something of importance to the business
- **Acceptability** – make sure it’s possible to define what “good” looks like
- **Accountability/Stewardship** – Links to the Data Governance structure with roles and accountability for action
- **Controllability** – Remedial actions are defined
- **Trackability** – Monitored over time to track progress

**Guidelines for Data Quality Indicators**

- Assign a unique identifier to each DQI
- Use a consistent naming convention such as DQI-XNN where NN are two digits and X indicates the associated Data Quality Measure (e.g. V = Validity, I = Integrity, etc.)
- Wherever possible, define each DQI as a percentage, with the numerator/denominator clearly identified in the derivation
- Set the polarity of each DQI such that the minimum value in the permitted range (e.g. 0%) represents the lowest level of quality and the maximum value (e.g. 100%) represents the highest level of quality
- Ensure each DQI definition is complete and includes a full description, rationale, the unit of measurement and permitted range
Data Quality Metrics

Example Measurement Framework
5. Define Data Quality Business Rules

It's important to make sure that data quality aligns with the rules of the business. For example:

› **Data Rules** that define the precise characteristics that data needs to adhere to
  
  » e.g. valid values/ranges for particular fields, relationships between fields/records, etc

› **Target Rules** that define the thresholds for Data Quality Indicators
  
  » e.g. red-amber-green status

› **Notification Rules** that define alerts that should be fired under particular circumstances
  
  » e.g. notifying a data steward if a record fails a validation check, alerting a data owner if data quality falls below a defined threshold, etc

› **Transformation Rules** that define operations that should be applied to data
  
  » e.g. automated correction of common data entry errors, standardisation of fields, etc

**Types of Data Quality Business Rules**

- Value domain membership
- Definitional conformance
- Range conformance
- Format compliance
- Mapping conformance
- Value presence & record completeness
- Consistency rules
- Accuracy verification
- Credibility verification
- Uniqueness verification
- Timeliness validation

**TIP:** Most DQ tools provide a rules repository so that rules can be created, managed, shared and re-used consistently across the business.
6. Test & Validate Data Quality Requirements

› It’s essential that the Data Quality business rules are validated to ensure they accurately reflect the underlying Data Quality requirements

› There are two complementary techniques:
  
  » **Top down** – formal review with business representatives to verify alignment with business expectations and ensure a common understanding
  
  » **Bottom up** – inspection of exceptions occurring on real data to verify correct rule implementation

› Once the Data Quality business rules have been validated, they can then be used to assess the baseline level of data quality
7. Set & Evaluate Data Quality Service Levels

› The Data Quality Indicators and the business rules upon which they are built are used to measure and monitor data quality

› However, in order to ensure timely resolution when thresholds are breached or non-conformant records are identified it’s important to establish a Data Quality Service Level Agreement (SLA)

› This will set out business expectations for response and remediation and provide a starting point for more proactive data quality improvement

**A typical Data Quality SLA should specify:**

- The data elements covered by the agreement
- The business impacts associated with data flaws
- The data quality dimensions associated with each data element
- The expectations for quality for each data element for each of the identified dimensions in each application or system in the value chain
- The methods for measuring against those expectations
- The acceptability threshold for each measurement
- The individual(s) to be notified in case the acceptability threshold is not met
- The timelines and deadlines for expected resolution or remediation of the issue
- The escalation strategy and possible rewards or penalties when deadlines are met/not met
Effective Data Quality Monitoring is one of the most important aspects of DQM – a best practice capability will:

› Support a variety of feedback mechanisms, including interactive dashboards displaying up to date information on the level of data quality for critical data assets

› Facilitate more detailed analysis to pinpoint the underlying problem areas and support root cause analysis

› Track changes in data quality over time to drive improvement and inform longer term data quality strategy

› Empower business users to take responsibility for data quality through the definition of rules and metrics

› Transform existing ad-hoc data quality profiling and measurement activities into “business as usual”

4 Key DQ Feedback Mechanisms

• **Exception Reports** provide timely feedback to data stewards on the quality of data under their stewardship

• **Operational Dashboards** provide data stewards, data owners and senior management with an interactive view of data quality within their area of responsibility

• **Subject Area Summaries** are published on a regular basis to highlight the level of data quality within a particular domain

• An **Annual Data Quality Report** brings together all of the data quality activities to provide a holistic assessment of data quality across the enterprise
Continuously Monitor & Measure Data Quality

Monitoring & Reporting helps create awareness across the organisation.

Remember Activity #1: Develop & Promote Awareness
9. Manage Data Quality Issues

› In order to expedite the resolution of data issues, a means of recording and tracking those issues is required

› A good Data Quality Incident Reporting System (IRS) provides this capability by:

  » Allowing users to log, classify and assign incidents as they are identified

  » Alerting Data Stewards to new incidents

  » Recording subsequent actions and outcomes from initial diagnosis through to final resolution

  » Handling incident escalation where SLAs have been breached

  » Providing management information such as statistics on issue frequency, common patterns, root causes, time to fix and historical trends

› The IRS helps assess performance against the SLA, supports data quality improvement initiatives and informs future Data Quality Strategy
10. Clean & Correct Data Quality Defects

› Detailed analysis of each data quality incident is vital to ensure that the root cause is identified and, wherever possible, eliminated so that repeats of the incident will not occur.

› In addition to this, the existing data quality defect(s) need to be resolved through one of the following mechanisms:

  » **Automated Correction** – obvious defects which are well understood can often be identified and fixed by triggering an automated data cleansing routine, with no manual intervention (e.g. address standardisation or field substitution).

  » **Directed Correction** – less obvious defects can often be identified automatically but may require manual intervention to determine if the suggested fix is appropriate (e.g. identity resolution and de-duplication).

  » **Manual Correction** – in some cases, even though a defect can be identified automatically, the only way of resolving it is through manual inspection and correction (e.g. an invalid combination of fields where it’s not clear which field is at fault).

› Data Quality tools use a scoring system to reflect the level of confidence in applying a correction – this can then be used to decide which defects should be corrected automatically (the cheapest and often the preferred option) and which should be flagged for directed or manual correction.
Data Cleansing Demystified

The quick fox jumps over the lazy dog
Data Cleansing Demystified

The quick fox jumps over the lazy dog
Data Cleansing Demystified

STANDARDISATION

SUBSTITUTION

The quick fox jumps over the lazy dog
Data Cleansing Demystified

The quick fox jumps over the the lazy dog
Data Cleansing Demystified

The quick brown fox jumps over the lazy dog
11. Design & Implement Operational DQM Procedures

› The Data Management Body of Knowledge identifies 4 key activities necessary for operationalising DQM:

» Inspection and monitoring - finding data quality issues)

» Diagnosis and evaluation of remediation alternatives (i.e. investigating possible fixes)

» Resolving the issue - applying an appropriate remedy

» Reporting - monitoring ongoing performance
12. Monitor Operational DQM Procedures & Performance

- Data Quality Management sets out good practice for ensuring data is fit for purpose – however, to succeed, any DQM programme needs to demonstrate tangible long-term benefits.

- This can only be done through ongoing monitoring and evolution of the approach as the organisation matures in its management of Data Quality – this includes:
  - Routine checking that SLAs are being met.
  - Introducing new Data Quality Indicators as previously undiscovered DQ issues are identified.
  - Extending the Measurement Framework to include further dimensions or measures (e.g. the Secondary Dimensions of Data Quality).
  - Developing new feedback mechanisms to satisfy the changing needs to users.
  - Increasing the scope to include new datasets.
  - Building in data quality at source, by improving the design of processes and systems.

- Just as good data will naturally degrade over time, even the best DQM approach will need ongoing refinement to ensure it continues to serve the business as effectively as possible.
# EA’s DQ Readiness & Maturity (based upon CMMI)

<table>
<thead>
<tr>
<th>LEVEL 1 – INITIAL</th>
<th>LEVEL 2 – REPEATABLE</th>
<th>LEVEL 3 – DEFINED</th>
<th>LEVEL 4 – MANAGED</th>
<th>LEVEL 5 – OPTIMISED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DQ CULTURE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There is limited awareness of the importance of data quality or the need for a consistent approach.</td>
<td>There is some awareness of the importance of data quality, but there is no common understanding across the business.</td>
<td>There is good awareness of the importance of data quality across the business and a common understanding about the key aspects.</td>
<td>Everyone in the business recognises the crucial importance of data quality and backs the drive for data quality improvement.</td>
<td>Everyone in the business recognises the crucial importance of data quality and takes a proactive approach to driving data quality improvement.</td>
</tr>
<tr>
<td><strong>RESPONSIBILITY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data quality activities are handled in a reactive manner with no assigned responsibility for resolving issues.</td>
<td>Data quality activities tend to be handled by the same individuals, but this isn’t a formal requirement of their role.</td>
<td>Responsibility for data quality activities is formally assigned through the creation of Data Stewards.</td>
<td>A Data Quality Champion takes a lead role in ensuring each business area adopts good practice with regards to data quality.</td>
<td>A Data Quality Champion ensures local adherence to the DQ standards and contributes to the wider Data Quality Community of Interest.</td>
</tr>
<tr>
<td><strong>MEASUREMENT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Few, if any, measurements of data quality are made on a routine basis and there is no clear understanding about the current level of data quality.</td>
<td>Some basic measurements of validity and completeness are applied to certain datasets, but these aren’t always applied consistently.</td>
<td>There is a standard set of business rules defined for key datasets and these are applied whenever data is received.</td>
<td>A comprehensive and consistent set of business rules and data quality indicators covering all datasets is stored in a local repository.</td>
<td>A comprehensive and consistent set of business rules and data quality indicators covering all datasets is stored centrally in a shared repository.</td>
</tr>
<tr>
<td><strong>REPORTING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No feedback is supplied regarding specific issues or the general level of data quality.</td>
<td>Feedback on data quality tends to be handled on an ad-hoc basis with no routine reporting.</td>
<td>Data Stewards are supplied with prompt feedback when business rules aren’t satisfied.</td>
<td>Data Stewards are supplied with prompt feedback when exceptions occur and on a regular schedule.</td>
<td>Data Stewards are supplied with prompt feedback when exceptions occur, on a regular schedule and on a self-service basis.</td>
</tr>
</tbody>
</table>
DATA QUALITY
Seven DQ Mistakes

1. Failing to consider the intended use of the data
   Data has to be fit for purpose, no more, no less

2. Confusing validity with accuracy
   Validity is only the first step towards accuracy

3. Treating data quality management as a one-time activity
   Quality data can only be ensured through a continuous cycle

4. Fixing data in a data warehouse rather than at source
   Clean data for reporting doesn’t solve the operational issues of poor DQ

5. Applying software quality principles to data quality
   Data is infinitely more volatile than SW and demands a different approach

6. Blaming systems for bad data
   People and processes lie at the heart of most DQ problems

7. Believing that good quality data is the end goal
   Deriving genuine value through information exploitation is the ultimate aim
If You’re Interested in Learning More

www.enterprisearchitects.com/learning

› EA offers a wide variety of training & consulting offerings including:
  » DAMA CDMP Certification Preparation
  » Applied Information Architecture Courses
  » Information Executive Overview
  » Business Architecture
  » TOGAF and Enterprise Architecture