



Successful Information Quality Management

C. Lwanga Yonke

DAMA Phoenix Chapter Meeting

August 16, 2007

Bio

C. Lwanga Yonke is a seasoned information quality and information management expert and leader. He has successfully designed and implemented projects in multiple areas, including data governance, information quality, business intelligence, data warehousing and data architecture. His initial experience is in petroleum engineering and operations.

An ASQ Certified Quality Engineer, Lwanga holds an MBA from California State University and a BS degree in petroleum engineering from the University of California at Berkeley.

Lwanga is a founding member of the International Association for Information and Data Quality (IAIDQ) and currently serves as an Advisor to the IAIDQ Board. He is also a member of the Society of Petroleum Engineers (SPE) and a senior member of the American Society for Quality (ASQ).

Abstract

Successfully tackling the tough challenges caused by poor data quality often seems like an overwhelming and thankless task. Moreover, as awareness about the importance of information quality grows, information quality (IQ) practitioners are increasingly called to tackle a myriad of complex IQ problems.

To be successful in the short and long terms, the IQ practitioner must be equipped with a robust foundation deeply rooted in proven best practices and applicable to various IQ efforts such as from CDI, MDM, compliance, governance, data integration, business intelligence, etc.

Drawing from lessons learned at the frontline, this presentation describes the fundamental components of successful IQ functions and provides practical guidelines on getting started and remaining successful

My Aims

- Rational aim
 - Share IQ ideas, thoughts and best practices that will outlive fads
- Experiential aim
 - We feel energized by the collective thinking of the community of IQ leaders in the room today
 - We feel excited about the possibilities and promises of the gift we will take back to our organizations

Outline

- **Part I - Introduction and overview**
 - IQ trends
 - Data properties and their impact
 - Issues and opportunities for the IQ leader
- **Part II – Understanding information quality management**
 - IQ methodologies & frameworks
 - IQ fundamentals
 - Measuring IQ value and cost
- **Part III – Radical transformations**
 - Managing data as an asset
 - Information stewardship
 - Managing information as a product
 - IQ in system development life cycle (SDLC)
 - Definitions and architecture
- **Part IV – Wrap-up**
 - The road traveled
 - Q&A
 - Check-out & evaluations

Part I - Introduction and Overview

IQ Trends

Information Quality Trends

- Top 3 issues Gartner clients are seeking technology and business research advice about
 - 1) managing information as a strategic asset
 - 2) addressing data and information quality issues
 - 3) information governance

Gartner Inc., "Researching the Research Agenda Survey 2007: Implications for Business Intelligence and Information Management" by Betsy Burton and Tom Austin, 11 May 2007

- In a November 2005 survey
 - 49% of respondents indicated that in their organization, data quality is worse than others realize
 - 48% indicated that their organization had no current plans for a DQ initiative
 - 16% indicated that they are already deployed a DQ solution.

TDWI, "Taking Data Quality to the Enterprise through Data Governance" by Philip Russom, March 2006

Information Quality Trends

- Between YE06 and YE12, Global 1000 organizations will experience a three-fold increase in data, content and application quality issues (0.7 probability)

Gartner Inc., "Findings: BI Survey Shows Accessibility and Quality of Information is a Management Issue" by Betsy Burton and Ted Friedman, 29 March 2007

- The data quality tools market will grow to \$677 million by 2011. This will reflect a strong five-year compound annual growth rate of 17.6% as organizations continue to invest in technologies critical to managing data assets

Gartner Inc., "Forecast: Data Quality Tools, Worldwide, 2006-2011" by Colleen Graham, 8 June 2007

Information Quality Trends

- Strategic Planning Assumptions
 - Through 2010, more than 75% of organizations will not get beyond Levels 1 and 2 in their data quality maturity (0.8 probability).
 - Through 2012, less than 10% of organizations will achieve Level 5 data quality maturity (0.8 probability).

Gartner Inc., "Gartner's Data Quality Maturity Model" by Andreas Bitterer, 7 February 2007

Information Quality Trends

- Regulations
 - US Data Quality Act of 2001
 - US Sarbanes-Oxley Act of 2002
 - US HIIPA
 - Basel II
 - California privacy laws
 - California air quality regulations
 - And many more
- News headlines
 - *“High-School Graduation Rates are Misleading”* [Bakersfield Californian, March 2007]
 - Mars Climate Orbiter crash (Aug 1999)
 - Medical errors
 - And many more

Information Quality Trends

- Professional association
 - International Association for Information and Data Quality (IAIDQ) (www.iaidq.org)
- Two new IQ journals
 - International Journal of Information Quality (IJIQ)
 - Journal of Data and Information Quality (JDIQ)
- Graduate programs in IQ
 - University of Arkansas at Little Rock
 - U. of South Australia, Adelaide (2008)
 - U. of Westminster, London, UK

Emerging Best Practices

- Establish a dedicated IQ function
 - Put someone in charge of the IQ process (at levels senior enough to deliver meaningful results)
 - Develop internal IQ management skills and culture
- Measure and report data quality
- Focus on process improvement (IT and non-IT)
- Adopt enterprise perspective
 - Many subject areas
 - Data flows across departmental and organizational boundaries
- Focus on the revenue side of IQ
 - Anchor IQ activities on strategic business priorities

Emerging New Profession

- Information quality (content) analyst
- Information quality process analyst
- Information quality manager
- For organizations where the CIO functions as a Chief Information Technology Officer:
 - Chief Data Officer (CDO) [Redman, January 2007]
 - Chief Information Quality Officer (CIQO) [English, September 2007]

Information Quality (Content) Analyst

- **Roles & Responsibilities**

- Design and execute data quality assessment and data quality monitoring plans for transaction systems and for a data warehouse
- Play lead role in major data migration projects
- Use DQ assessment results to design and implement data remediation plans
- Map source models to target models
- Analyze findings in order to design, develop and test data extraction, transformation and loading (ETL) processes
- Identify opportunities to improve the relevant business processes

Information Quality (Content) Analyst

- **Essential Skills/Knowledge**

- Excellent data quality and data analysis skills
- Experience with structured data quality assessment/data profiling methods
- Extensive experience with data extraction, data mapping and data cleansing/transformation
- Strong SQL, PL or Transact SQL, programming and relational database design skills
- Data warehousing
- Experience with a leading ETL tool is a plus
- Familiarity with the various types of data quality tools
- Experience with statistical analysis tools and methods is a plus.
- Familiarity with the work of Larry English, Tom Redman or Richard Wang
- Broad exposure to the Zachman Framework and to the concepts of Enterprise Architecture
- Solid verbal and written communication skills

Information Quality Process Analyst

- **Roles & Responsibilities**

- Serve as information quality resource to internal teams
- Lead or participate in data quality improvement projects
- Analyze current work processes to identify opportunities to improve data quality
- Develop & implement improvement plans. Establish processes to measure and sustain the gains
- Facilitate the development of information customer-supplier service level agreements
- Assist in the identification, sharing and adoption of IQ best practices
- Provide just-in-time training on data quality principles, best practices and tools
- Develop and engage a network of internal data quality advocates, in order to increase data quality skills throughout the company

Information Quality Process Analyst

• Essential Skills/Knowledge

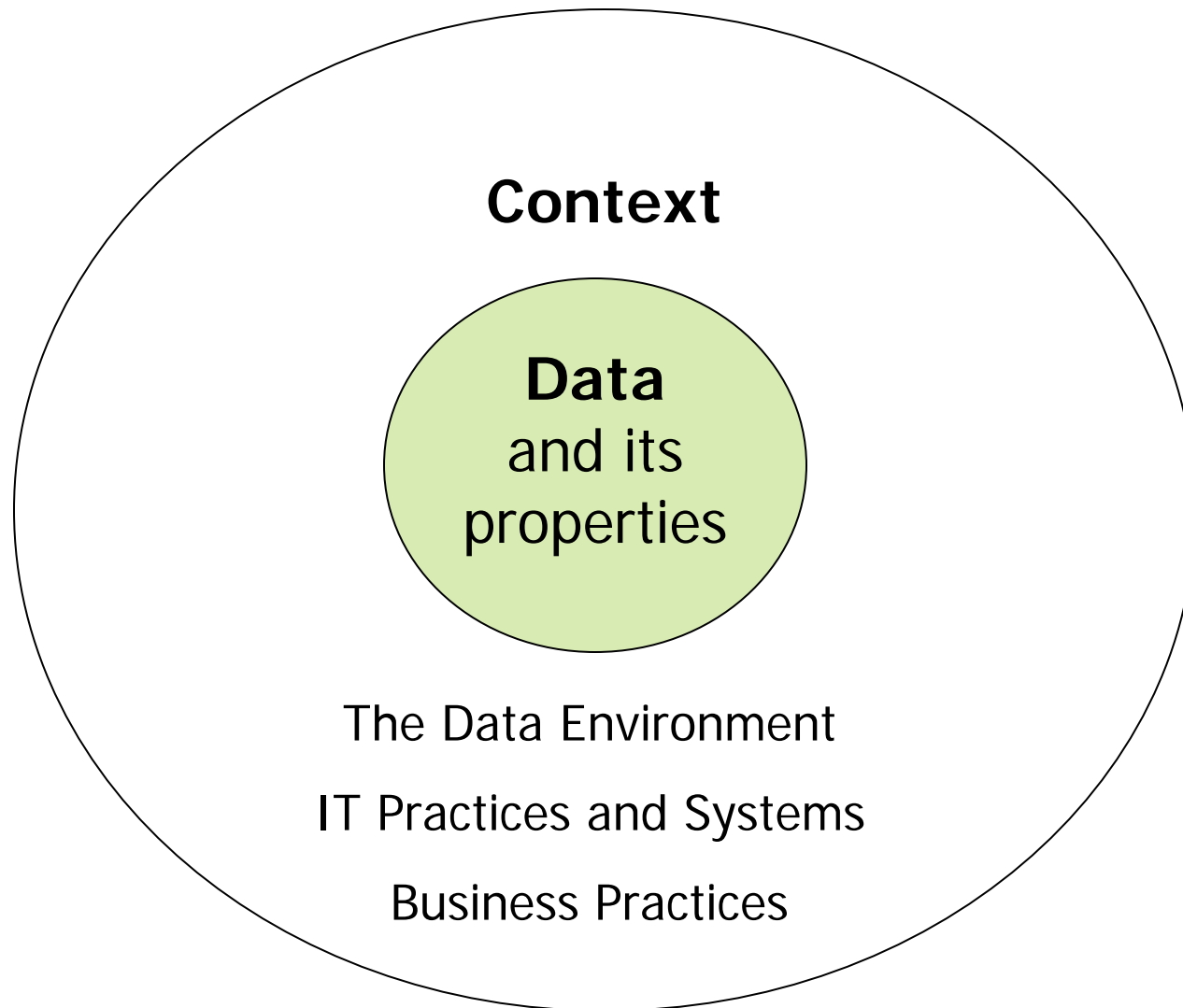
- Proven data quality and continuous improvement skills. Practical experience with data quality principles and methodologies
- Systems thinker. Solid understanding of the interplay of People, Process and Technology in the effectiveness of IT solutions
- Must be able to act interdependently and take a proactive approach to process improvement
- Must enjoy and be very strong in the following areas: analysis, root-cause problem-solving, process mapping skills, dealing with ambiguity
- Excellent communication and listening skills. Strong organizational and team skills. Proven ability to lead and influence others
- Familiarity with the work of Larry English, Tom Redman or Richard Wang
- Broad exposure to the Zachman Framework and to the concepts of Enterprise Architecture
- Great technical writing skills
- Past auditing or business analyst experience a plus
- Lean or six sigma experience a plus
- Commitment to quality

Part I - Introduction and Overview

Data Properties and their Impact

Why Is IQ Improvement So Difficult?

Two Levels of Perspective for Data As Resource/Asset



Why Is IQ Improvement So Difficult?

Data as a Resource

Properties

- Data is not tangible
- Not consumable
- Shareable, but also easily copyable
- Fragile
- Nonfungible
 - one unit of data cannot be substituted with another unit
- Versatile
- No standard valuation rules
- Data values are renewable
 - as pertinent features of the real world change
- Can be stored in a computer
 - and often is

Implications

- Out of sight, out of mind
- Perceived as “free”, not as a scarce resource whose use must be controlled, and life expectancy managed
- Uncontrolled redundancy
- Data inconsistencies
- High costs of poor initial capture
- Unexpected uses
- Disparate DQ requirements
- Hard to justify need for enterprise care
- Pressure to maintain and update local copies meeting local needs
- Hard to separate from IT

Source: Levitin and Redman, 1998. © 1998 Thomas C. Redman. All Rights Reserved.

Copyright © 2006-2007 C. Lwanga Yonke. All rights reserved.

Why Is IQ Improvement So Difficult?

The Data Context

- IQ not taught in colleges and universities (with a few rare pioneering exceptions)
- No common language for IQ and IQ management
 - Data vs. information debate
 - Many definitions of “accuracy”
 - List of DQ dimensions: long and still evolving
 - Etc.
- IQ dimensions encompass several domains (IQ hard to nail down):
 - Data content: e.g. accuracy, precision
 - Data acquisition and delivery process: e.g. entry timeliness, update frequency
 - Data architecture and definition: attribute completeness and granularity
 - Presentation quality: data meaning, metadata availability
 - Security: e.g. integrity, accessibility
 - Etc.

Why Is IQ Improvement So Difficult?

The Data Context

- Most still perceive data quality as a clerical activity (legacy of “data processing” and “key punching” days)
- Data creators and data customers live in different parts of the organization
 - Problem-solving requires an appreciation for the “horizontal” processes
 - No reconciliation between data acquisition costs, borne by the provider unit, and data use benefits, enjoyed by the customer unit
- Data not administered centrally, like other enterprise resources
- Data quality often seen as an IT concern. However
 - IT rarely feels the pain
 - Historical IT development practices have left their non-OT customers with a bitter taste
 - Non-IT teams are reluctant to or do not know how to address DQ problems
- IQ career ladder still poorly defined

Why Is IQ Improvement So Difficult?

End Results

- People often feel powerless when it comes to improving information quality and delivery
- Residual emotional scars from past experiences with bad data quality
- Most fail to recognize that information management can help solve some of their business pains
- Attempts to improve information often perceived as irritants, quicksand, with little chance of success
- Typically, little marketplace glory for leading information improvement projects (compare to six sigma, lean etc.)
- IQ assignments often seen as temporary assignments

Why Is IQ Improvement So Difficult?

The Dilemma (Heresy?) of IQ Improvement

Our prescriptions often fly in the face of deeply entrenched patterns of behaviors:

- People value freedom and independence:
 - My data
 - My department
 - My location
 - My business rules
 - My needs
 - My happiness
 - My local simplicity or complexity
- We offer control and interdependence:
 - Our data
 - Enterprise thinking
 - All our locations
 - Enterprise business rules
 - Needs of all the data customers
 - Customer satisfaction
 - Enterprise governance/complexity

Widely accepted for all the other assets of the organization,
but not yet for the Data/Information asset

Part I - Introduction and Overview

Issues and Opportunities for the IQ Leader

Issues and Opportunities

- Prioritizing the needs
- Choosing a focus
- Choosing a deployment model
- Staffing
- Convincing management
- Institutionalizing IQ

Prioritizing the Needs

- Saying no so that you can say yes
- Consider
 - Risk, probability of success
 - Visibility
 - Leverage

Choosing a Focus

- IT focus
- Non-IT focus
- Organizational chart implications
 - The best home on the org chart
 - You are wanted there
 - Your boss values information quality and what you do
 - Your boss has enough power and resources to remove obstacles and influence business priorities
 - The IQ leader has great networking, influence and coalition-building skills
 - If the above is true, it does not matter much if your position is in IT or in another group

Choosing a Deployment Model

- Approach
 - Structured project-by-project
 - Teach & preach
- Specific IQ identity vs. integration with
 - Existing IT system development methodology (waterfall, agile etc.)
 - Existing business process improvement methodologies: six sigma, lean, business process re-engineering etc.
- Work driven and done by IQ function vs. work driven and done by others but guided by IQ function

Staffing

- Desired skills and competencies
 - Analytical
 - Business
 - IT
- Staff count
- Mix of IT and non-IT professionals
- External hiring or internal development
- Dedicated (full-time) vs part-time resources
- Career ladder

Convincing Management

- Building the case
- Story telling, anecdotes or hard dollars
- The dreaded ROI
- Approaching the CxOs
- Defining commitment (chicken or pig?)

Institutionalizing IQ

- From pilot to “the way we do things around here”
- Integration and partnerships (data architecture, continuous improvement, six-sigma, information stewardship, etc..)
- Inclusion in corporate business planning process
- Staff ladder, HR recognition

Outline

- Part I - Introduction and overview
 - IQ trends
 - Data properties and their impact
 - Issues and opportunities for the IQ leader
- **Part II – Understanding information quality management**
 - IQ methodologies & frameworks
 - IQ fundamentals
 - Measuring IQ value and cost
- Part III – Radical transformations
 - Managing data as an asset
 - Information stewardship
 - Managing information as a product
 - IQ in system development life cycle (SDLC)
 - Definitions and architecture
- Part IV – Wrap-up
 - The road traveled
 - Q&A
 - Check-out & evaluations

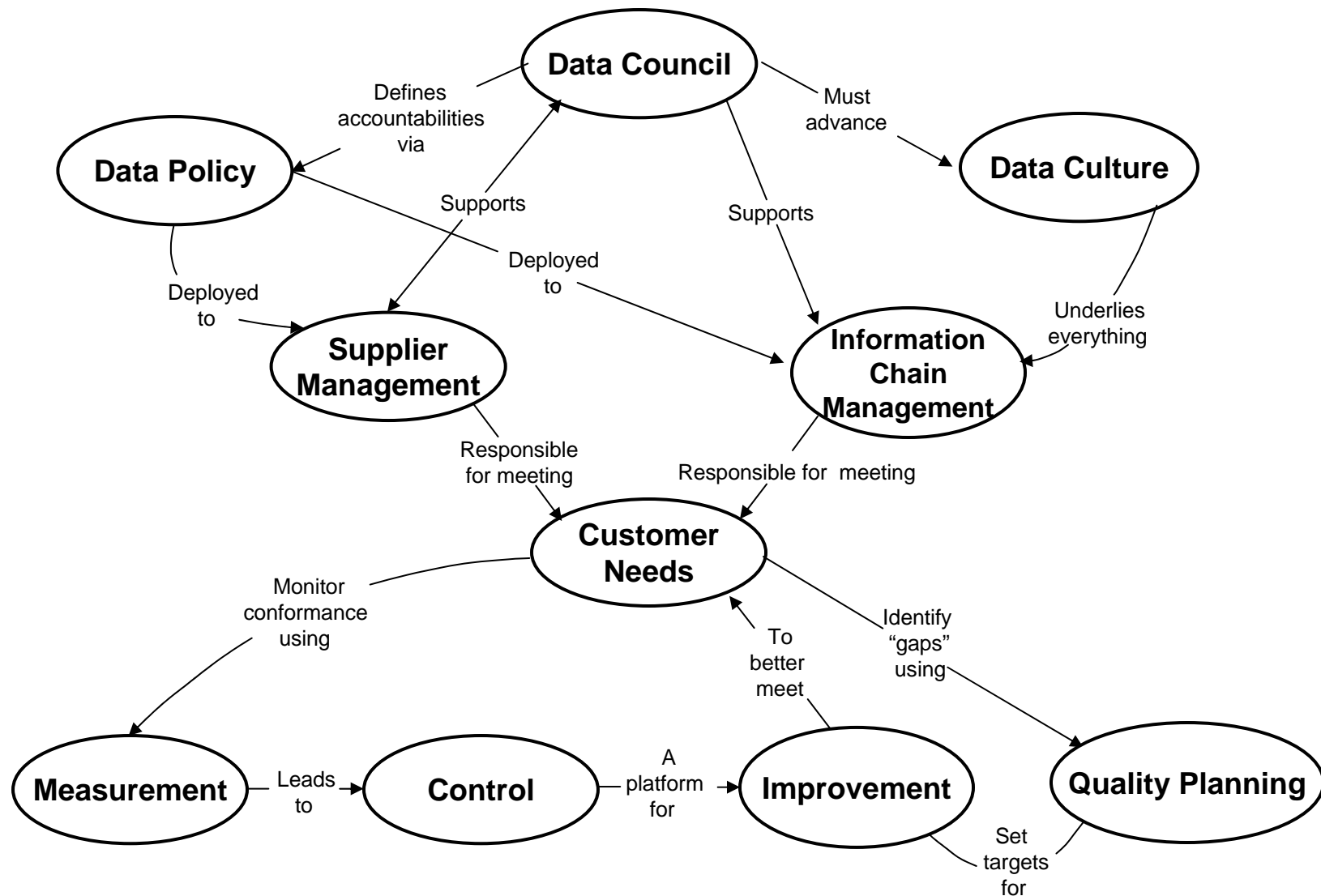
IQ Methodologies & Frameworks

IQ Methodologies & Frameworks

- IQ-specific
 - Redman
 - English
 - Wang
 - McGilvray
- General improvement methodologies
 - Six Sigma
 - Lean
- IT-centric (audit and control)
 - ISO/IEC 17799
 - COBIT 4.1
- Other methodologies

Second-Generation Data Quality Systems

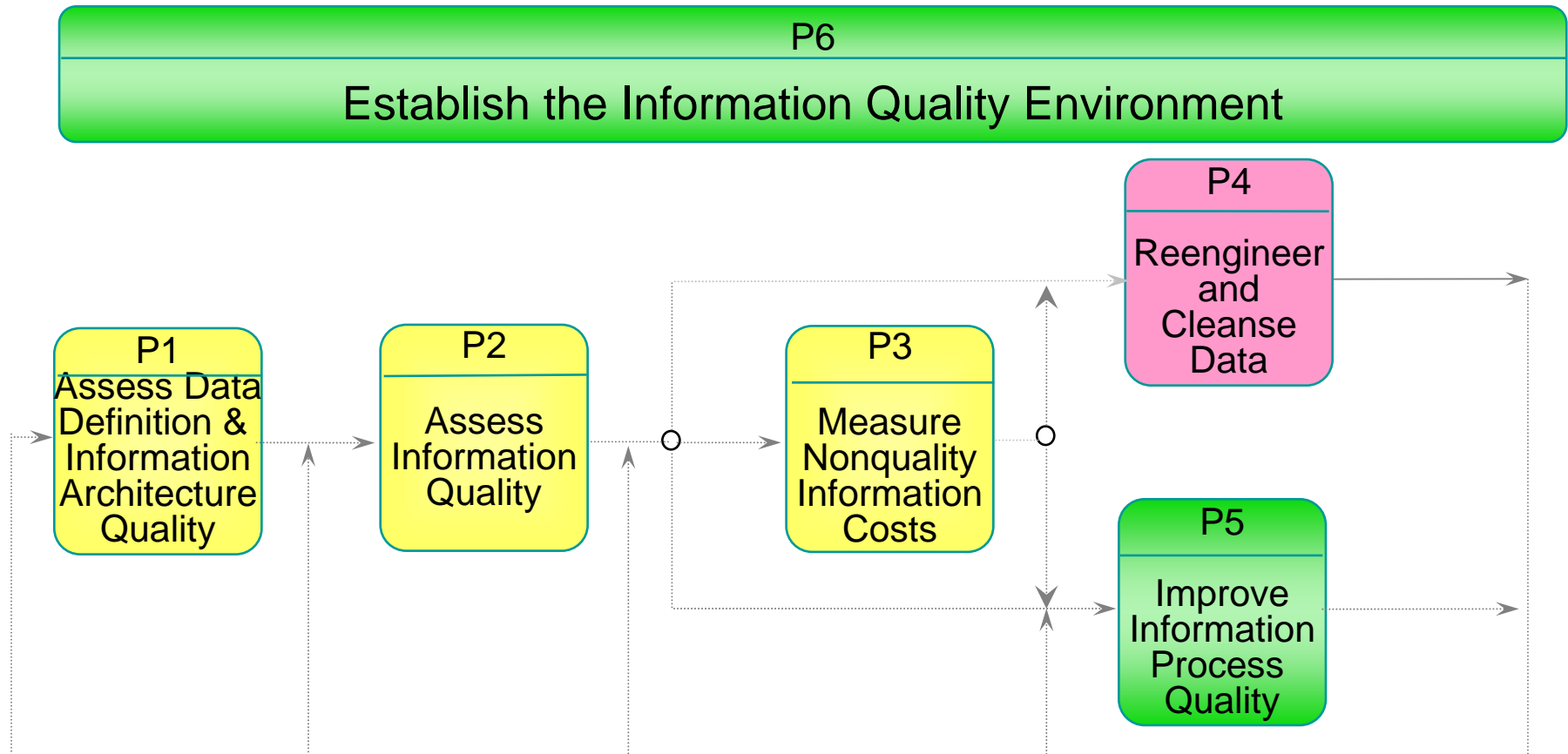
Tom Redman



Source: Redman, 2001. © 2001 Thomas C. Redman. All rights reserved
Copyright © 2006-2007 C. Lwanga Yonke. All rights reserved.

Total Information Quality Management (TIQM)

Larry English



Source: English, 1999. © 1995 – 2007 INFORMATION IMPACT International, Inc. All rights reserved.

Total Data Quality Management (TDQM)

Richard Wang

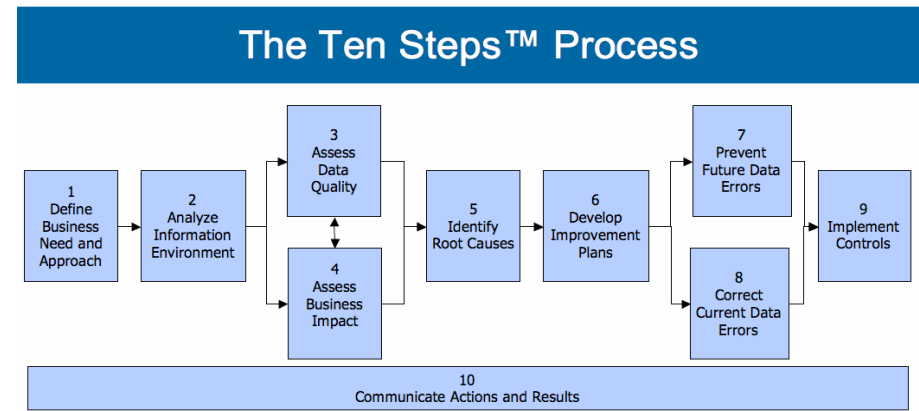
- Define
- Measure
- Analyze
- Improve

The Information and Data Quality Framework and the Ten Steps Process

Danette McGilvray

Information and Data Quality Framework™ (IDQF™)

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



Source: McGilvray, 2007. © 2005-2007 Danette McGilvray, Granite Falls Consulting, Inc. All rights reserved.

Information and Data Quality Framework™ (IDQF™)



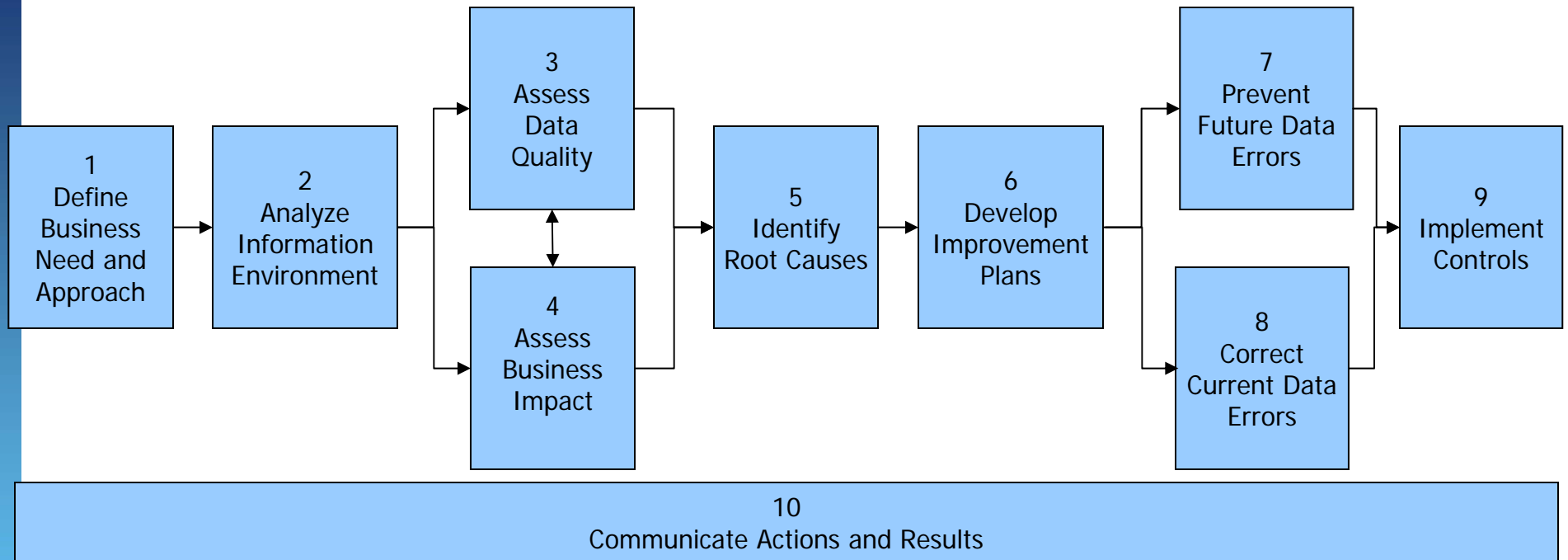
1.01

Source: McGilvray, 2007. © 2005-2007 Danette McGilvray, Granite Falls Consulting, Inc. All rights reserved.

Copyright © 2006-2007 C. Lwanga Yonke. All rights reserved.

The Ten Steps™ Process

Danette McGilvray



Source: McGilvray, 2007. © 2005-2007 Danette McGilvray, Granite Falls Consulting, Inc. All rights reserved.

General Improvement Methodologies

- Six Sigma
- Lean

IT-centric Frameworks

ISO/IEC 17799

7. Asset management

7.1 Responsibility for assets

7.1.1 Inventory of assets

7.1.2 Ownership of asset

7.2 Information classification

11. Access control

11.2 User access management

11.2.1 User registration

11.2.2 Privilege management

11.2.4 Review of user access rights

11.3 User responsibilities

12. Information systems acquisition, development and maintenance

12.2 Correct processing in application

12.2.1 Input data validation

12.2.2 Control of internal processing

12.2.4 Output data validation

Source: ISO/IEC 17999 © ISO/IEC 2005. All right reserved.

Copyright © 2006-2007 C. Lwanga Yonke. All rights reserved.

IT-centric Frameworks

COBIT 4.1

PO2. Define the information architecture

PO2.1 Enterprise architecture model

PO2.2 Enterprise data dictionary and data syntax rules

PO2.3 Data classification scheme

PO2.3 Integrity management

DS5. Ensure system security

DS5.4 User account management

DS11. Manage data

DS11.1 Business requirements for data management

Source: COBIT 4.1 © 1996 – 2007 IT Governance Institute. All rights reserved.

Other Methodologies

- Vendor or tool-specific
- In-house
 - Customized from published methodologies
 - Original

IQ Fundamentals

Fundamental IQ Activities

- Establish accountabilities
- Specify data quality requirements
- Control data acquisition, entry, transfers and delivery
- Measure data quality
- Improve data quality
- Build the IQ environment and culture
- Manage the IQ process

Fundamental IQ Activities

- Establish accountabilities
 - Information stewardship
 - Information governance
- Specify data quality requirements
 - DQ dimensions
 - DQ rules
- Control data acquisition, entry, transfers and delivery
 - Data collection and entry standard operating procedures
 - Data entry validation rules
 - Data transfers and transformation controls
 - Report certifications

Fundamental IQ Activities

- Measure data quality
 - Data model and definition quality
 - Baseline data quality assessment (data profiling, DQ rules conformance, accuracy)
 - Data certification
 - Data quality monitoring
 - Data tracking
 - Customer satisfaction
 - IQ costs and value
 - Measurement system analysis
 - Measurement process
 - Gauge accuracy, repeatability and reproducibility
- Improve data quality
 - Process improvement (IT and non-IT)
 - Data correction/remediation/cleansing
 - Data enrichment

Fundamental IQ Activities

- Build the IQ Environment and Culture
 - Systems and structure
 - Promotion and Communication

Fundamental IQ Activities

Build the IQ Environment and Culture

- Systems and structures
 - Training & development
 - Executives, middle managers and first-line supervisors
 - Staff
 - Suppliers
 - IQ network
 - Policies
 - Standards
 - Auditing
 - Job design, descriptions and expectations
 - Rewards and recognition systems
 - Business planning process
 - Information directory (metadata repository)

The Four Pillars

Build the IQ Environment and Culture

- Promotion and Communication
 - Logo, motto, stories
 - E-mail, newsletter
 - Intranet site
 - Screen-savers
 - Posters
 - Pocket cards
 - Forums, Lunch & Learn
 - Fun events
 - Promotion toys (coffee mugs, shirts, mouse pads, etc.)
 - Road-shows
 - Rewards & recognition
 - Share of mind

The Various Types of Data

Financial statements

Financial reporting data

Operational reporting data

Transaction & operations data

Master, reference and structure data

A bottom-up approach often provides the greatest chance for sustained IQ success

Measuring IQ Value and Costs

IQ Costs Classification

Redman

- Strategy
 - Setting strategy
 - Executing strategy
 - Ability to derive value from data
 - Ability to align organizations
 - Management attention
- Operations
 - Operating costs
 - Customer satisfaction
 - Employee morale
- Tactics/Decision Making
 - Lost sales
 - Risk
 - Decision making capability
 - New technology
 - Reengineering
 - Trust between organizations

Source: Redman, 2001. © 2001 Thomas C. Redman. All rights reserved

IQ Costs Classification

English

- Non-quality information costs (**eliminate**)
 - Internal and external process failure costs
 - Information scrap and rework
 - Lost and missed opportunity costs
- Assessment or inspection costs (**minimize**)
- Process improvement and defect prevention costs (**invest here**)

Source: English, 1999. © 1999 Larry P. English. All rights reserved

Measuring IQ Costs and Value

- Financial metrics
 - Return on Investment (ROI)
 - Net Present Value (NPV)
- Rule of thumb
 - 5 to 20% of annual revenue or budget
 - 30 to 50% of people's time
- Anecdotes and stories (survey knowledge workers)
- Pictures and pithy quotes
- Tri-lingual skill
 - Money
 - Process
 - Data quality

Outline

- Part I - Introduction and overview
 - IQ trends
 - Data properties and their impact
 - Issues and opportunities for the IQ leader
- Part II – Understanding information quality management
 - IQ methodologies & frameworks
 - IQ fundamentals
 - Measuring IQ value and cost
- **Part III – Radical transformations**
 - Managing data as an asset
 - Information stewardship
 - Managing information as a product
 - IQ in system development life cycle (SDLC)
 - Definitions and architecture
- Part IV – Wrap-up
 - The road traveled
 - Q&A
 - Check-out

Case For Action

Understanding Data Quality Problems

- Problems with *data content* (*data values*)
- Problems with *data architecture* (*structure and definitions*)
- Problems with *data delivery process* and *data presentation*
- Persistent data quality problems are symptoms of deeper, more insidious, less tangible and much harder to fix *organizational problems*
- **Organizational problems**
 - Data creators and data customers in different parts of an organization
 - No reconciliation between data acquisition costs, borne by the provider unit, and data use benefits, enjoyed by the customer unit
 - Data not administered centrally, like other enterprise resources
 - Historical IT system development practices
 - Etc.

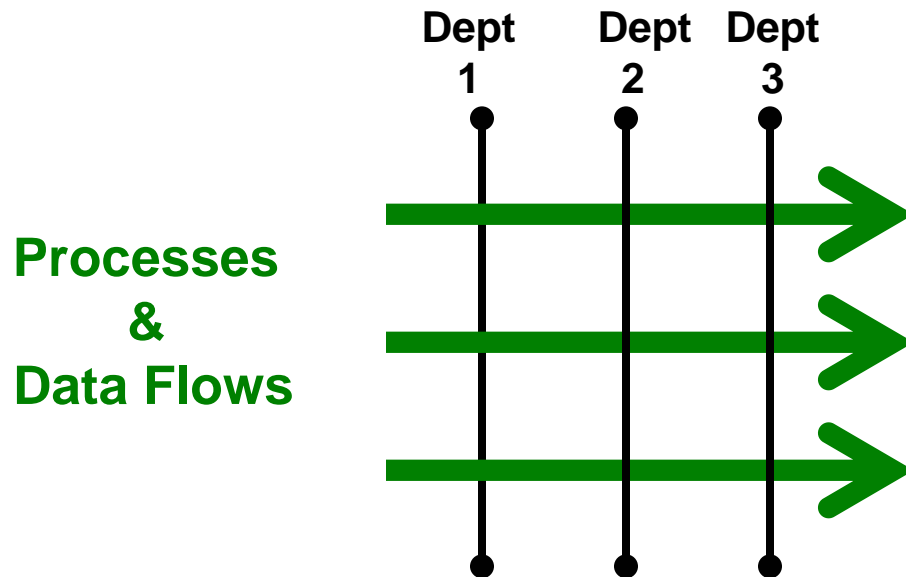
Radical transformations are required to eliminate these organizational problems

Part III – Radical Transformations

Managing Data as an Asset

How Is Your Company Organized?

Functional Organization



- Structure influences behaviors
- The System is perfectly designed to get the results it is getting

Managing Data as an Asset

Learning from Other Assets

	People	Money	Equipment and Facilities	Data
"Model"	Organization chart	Chart of accounts	Facilities drawings Plant and instrumentation diagrams	Logical data model (entity-relationship diagram)
Rules	Hiring Practices Compensation Plans	GAAP FASBE	Maintenance management best practices Asset Accounting	Business and data quality rules Data standards
Tracking	Staff Counts	Budget Monthly financial results	Equipment and facilities inventories	Data inventory

Managing the Data Asset

Redman's 3 Tests

- Care and feeding test
 - Make sure your data has a high quality
 - Protect it from loss or theft
- Unique and significant contribution test
 - Learn how you will use data to make money
- Special properties test
 - Understand the special properties of data
 - Adjust management structures as appropriate

Source: Redman, July 2007. © 2007 Thomas C. Redman. All rights reserved

Copyright © 2006-2007 C. Lwanga Yonke. All rights reserved.

Part III – Radical Transformations

Information Stewardship

Information Stewardship Roles and Functions

Larry English's Perspective

Business Roles

Information Function

Knowledge worker (KW).....	Uses/applies information
Information producer.....	Creates information content
Information transcriber.....	Enters data from Producers
Information translator.....	Interprets and enters information
Information agent.....	Prepares information for KW
Manager (process execution).....	Accountable for information
Process (definition) owner*	Defines standard process
Executive leader.....	Assures IQ in their authority
Business information steward*	Assures definition quality
Executive Leadership Team.....	Resolves issues; deploys policies
Business information Stewardship team.....	First line of definition governance

* *Appointed roles*

Source: English, June 2007 © 2007 Larry P. English. All rights reserved

Information Stewardship

Other Perspectives

- Most common
 - Stewardship program manager
 - Subject area data stewards
 - Primarily roles
 - Assure data quality
 - Assure quality of definitions
- Other
 - Information product manager (Rich Wang)

Part III – Radical Transformations

Managing Information as a Product

Managing Information as a Product

Wang's Four Principles

- Understand information consumers' needs
- Manage information as the product of a well-defined information production process
- Manage the life cycle of information products
 - Creation, growth, maturity, decline
- Appoint an information product manager to manage information processes and products

Source: Fisher *et al*, 2006. © 2006 MIT Information Quality Program. All rights reserved.

Copyright © 2006-2007 C. Lwanga Yonke. All rights reserved.

Part III – Radical Transformations

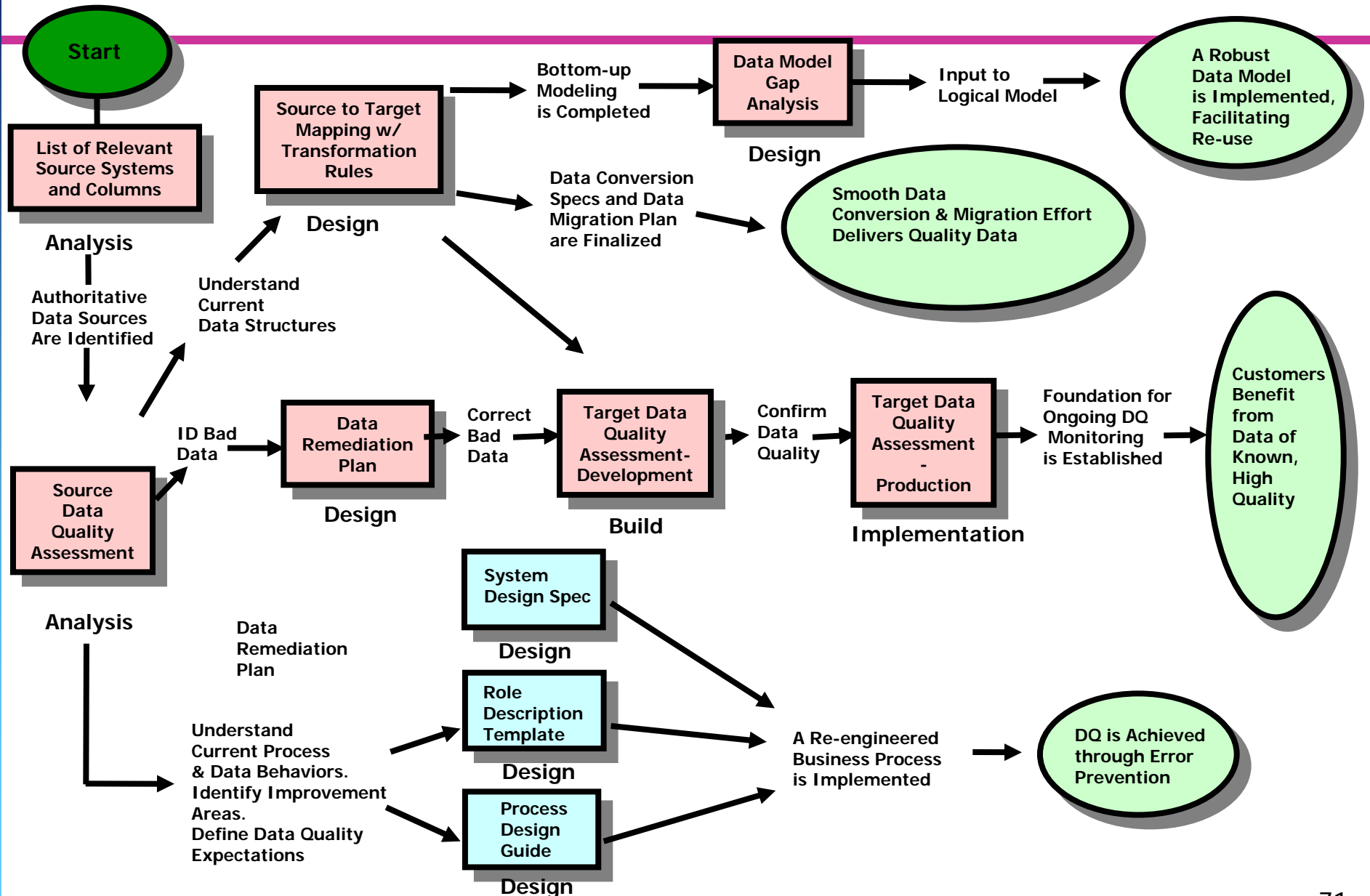
Information Quality Activities in the System Development Life Cycle

System Development Life Cycle

SDLC Tracks	SDLC Phases						
	Feasibility	Planning	Analysis	Design	Build	Implementation	Run and Maintain
Project Management							
Business Process Re-design							
Change Management							
Data Modeling							
Data Quality Assessment & Conversion							
Application							
Technology							
Quality Assurance & Testing							

SDLC Data Quality Assessment and Conversion

Flow of Deliverables



Data Quality Assessment and Conversion Deliverables

- 1) List of relevant source systems and columns
 - 2) Source data quality assessment, including process improvement recommendations
 - 3) Data remediation plan
 - 4) Source-to-target mapping w/ transformation rules
 - 5) Data model gap analysis
 - 6) Target data quality assessment - development
 - 7) Target data quality assessment – production
- Note: not always a linear one-way process

List of Relevant Source Systems & Columns

- Purpose
 - To find the best possible sources of data for the new system
- Canvas business customers to surface all the shadow databases: they may contain data of higher quality than the official corporate system, and success may mean removing the need for them
- Requires good understanding of project's scope
- Provides early understanding of the project's data conversion complexity

Source Data Quality Assessment Report

- Purpose
 - A comprehensive assessment of data quality levels in the source systems
- Includes:
 - a discussion of potential root causes
 - rough estimates of the business impacts (costs of the poor quality identified)
 - recommendations for IM&T and business process improvement

Data Quality Assessment Methodology

- Domain analysis (patterns, frequency distribution or histogram, etc.)
- Completeness and validity
- Referential integrity
- Data quality rules conformance
- *(Many good tools do the above in very different ways. Therefore*
 - a) understand your options,*
 - b) define your needs,*
 - c) develop your work process, then*
 - d) buy the right tool)*
- And **accuracy** !!!!
- Include in the Target assessments, an evaluation of the compliance to transformation rules

Why Perform a Source Data Quality Assessment?

- Understand spoken and unspoken definitions, business rules, data structures etc., in order to:
 - develop comprehensive conversion and transformation specifications
 - validate and enhance the target data model
- Identify and correct the bad data
- Understand the current data and process behaviors, so that process improvement opportunities can be identified and realized (business and IT processes)

Data Remediation Plan

- Purpose
 - A detailed plan that clearly specifies how all the errors identified in the source data quality assessment report will be corrected. It is the data correction design specification
- It identifies and specifies:
 - data remediation in the source system (class 1)
 - data remediation during the initial data ETL (class 2)
 - data remediation in the target system (class 3)
 - costs, effort, timing, funding and staffing resources

Source-To-Target Mapping, with Transformation Rules

- Purpose
 - To describe the requirements necessary to convert the source data into the target system.
- Serves as a design specification that will be used to code the ETL programs
- Requires frequent consultation with logical and physical data modelers
- To identify domain schizophrenia and overloaded facts, map columns and values (including codes)

Data Model Gap Analysis

- Purpose
 - to identify potential needs to extend the target data model to include source facts that do not have a home in the target model
 - to identify and resolve target facts that do not have a parent in any of the available sources.
- Must be started after the completion of the domain analysis and completeness/validity data quality assessments.
- Must be used to update the target data model
- Must also address columns and values (including codes)

Target Data Quality Assessment Report - Development

- Purpose
 - A comprehensive assessment of data quality levels in the target system, in the Development environment
 - Used to provide a snapshot view of the data quality levels in the target database.
 - Serves as validation of the data remediation and data conversion (ETL) activities
- Conducted after the initial data conversion into the new system
- Requires the use of the target's business rules and definitions
- Usually easier to perform than the source assessment, since analyst has "perfect" knowledge of the new system

Target Data Quality Assessment Report - Production

- Purpose
 - A comprehensive assessment of data quality levels in the target system, in the production environment
 - Used to quantify production data quality “as delivered”
 - Serves as the foundation for ongoing data quality monitoring activities
- One of the last activities before mass release to customers and hand-off to Run & Maintain
- Re-uses all the queries and tools from the Target Data Quality Assessment - Development

IT Process Improvement Recommendations

- Possible elements
 - Data capture and validation practices
 - Business rule implementation
 - Application design standards
 - GUI design
 - Data modeling
 - Requirements capture and traceability

Business Process Improvement Recommendations

- Possible elements
 - Business process redesign
 - Information stewardship implementation
 - Data quality specifications
 - Reward and recognition system
 - Information quality policy

Staffing for Data Quality Assessment and Conversion

- Information quality (content) analyst
- Business subject matter expert
- Business information steward
- Data administrator and data modeler
- Business analyst
- IT lead
- Business process sponsor
- Change management coordinator
- Training staff

Features of the Methodology

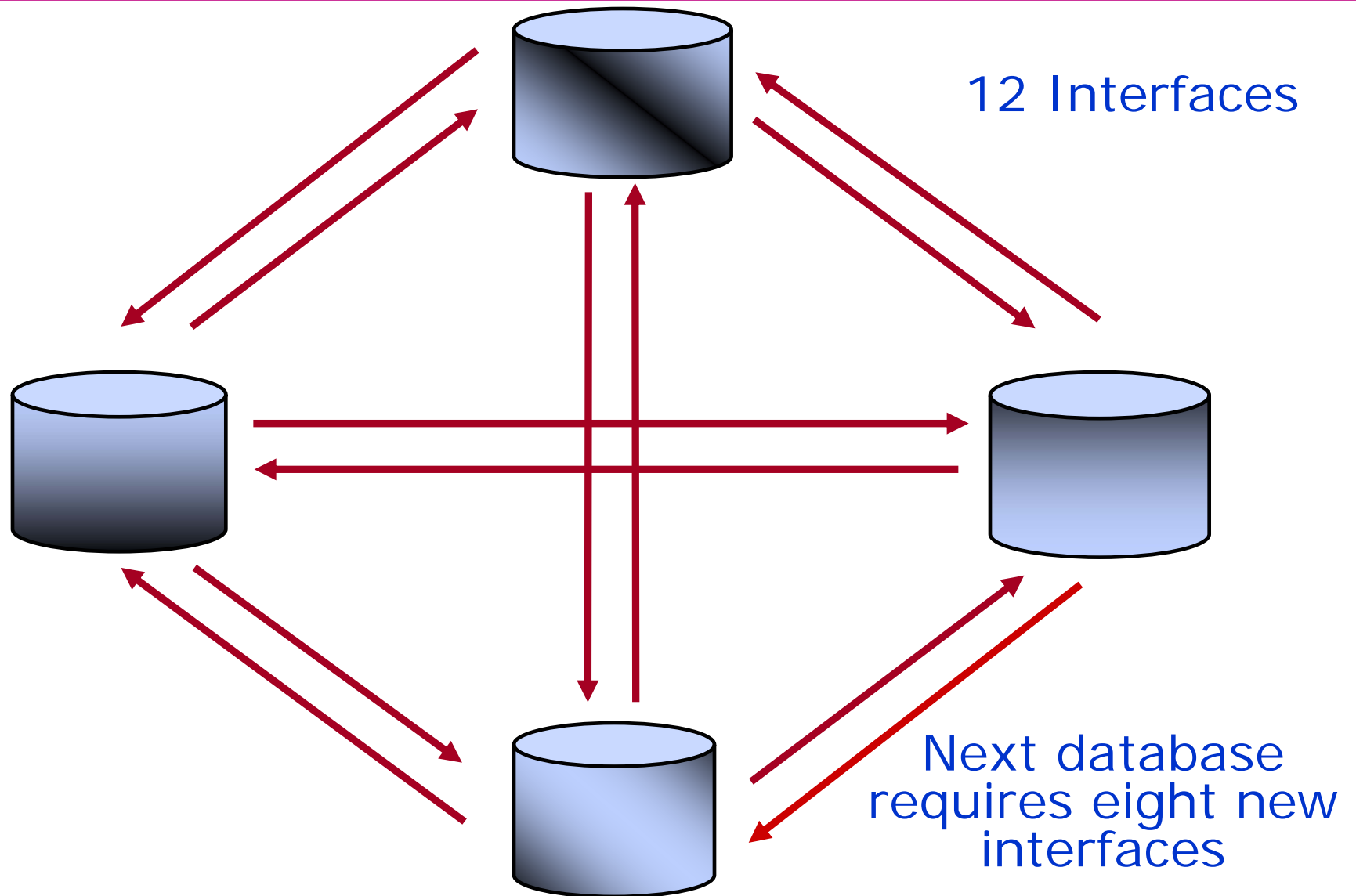
- Includes all aspects of data quality (definition and architecture, content, presentation, etc..)
- Data quality work done as part of the traditional IT development work, not conducted as a separate, parallel activity
- Data quality staff included in project teams, not positioned as external resources
- Leverages interdependencies between IT development functions (business process, data, software, hardware, etc.)
- Integrates people, process and technology to deliver a complete solution

Part III – Radical Transformations

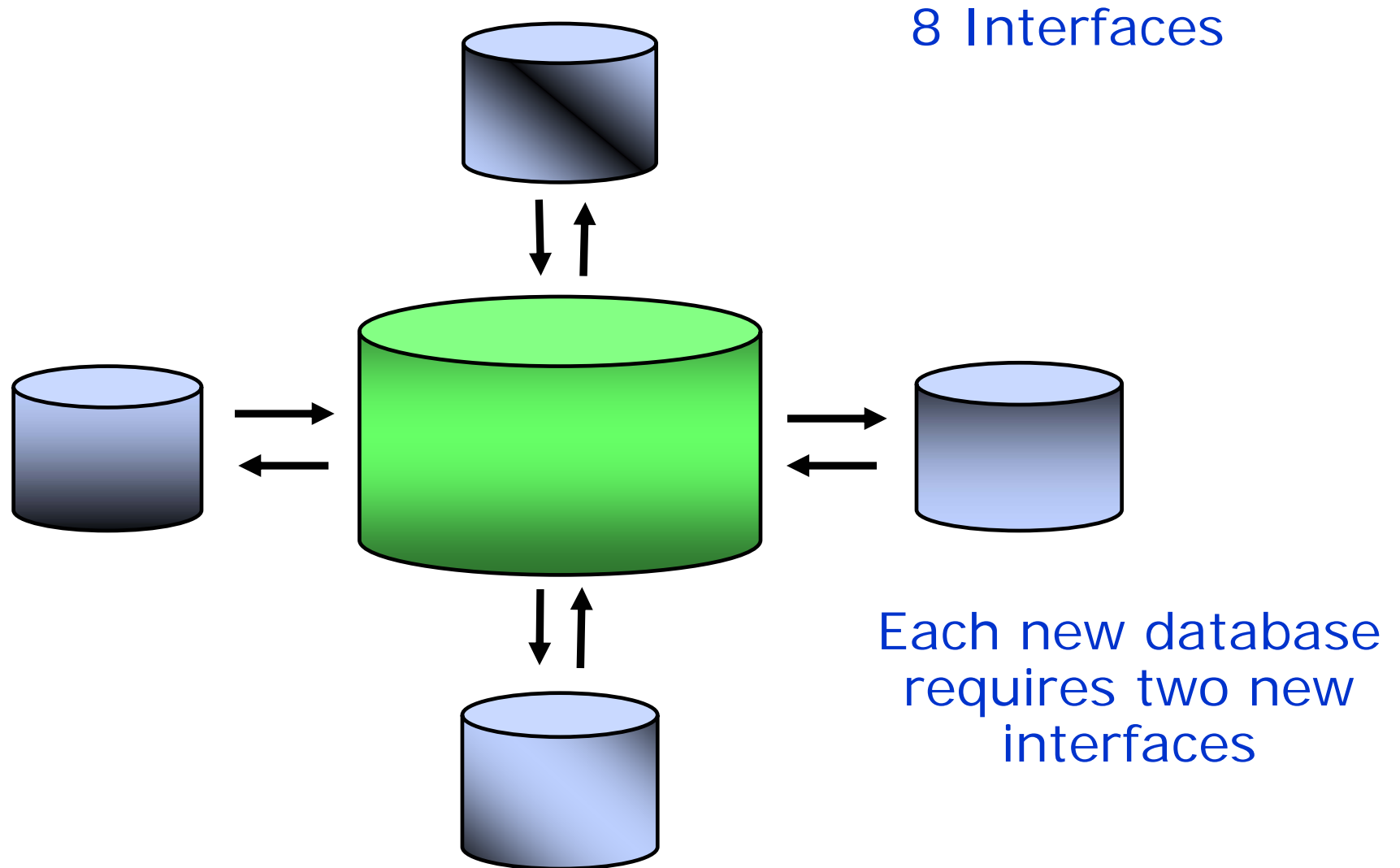
Definitions and Architectures



Information Sharing: Non - Architected



Information Sharing: Architected

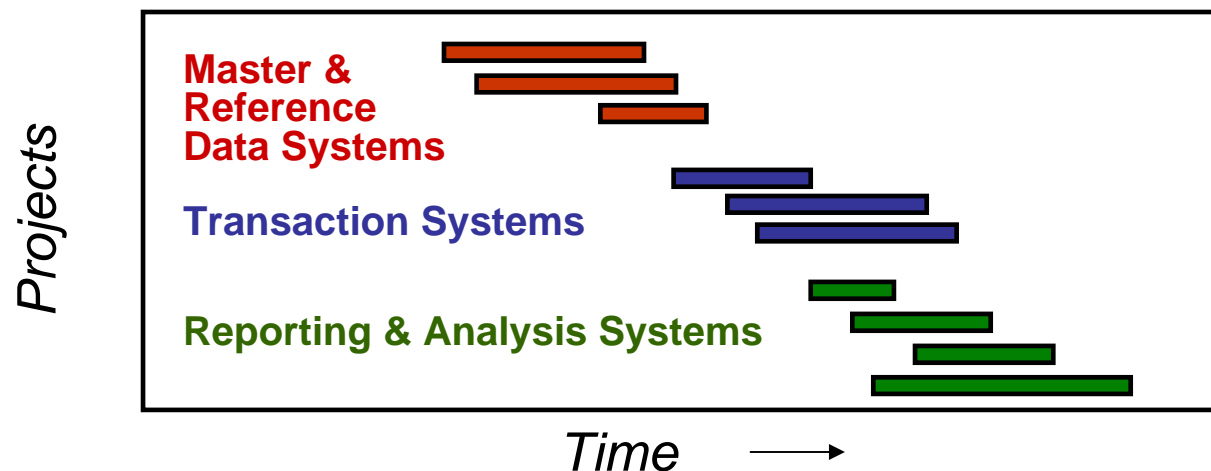


Enterprise Architecture (EA)

- An EA plan has four typical components
 - Business architecture
 - Data architecture
 - Application architecture
 - Technology architecture
- Develop the plan *in the order above*
- then *implement* it

Data-driven Implementation

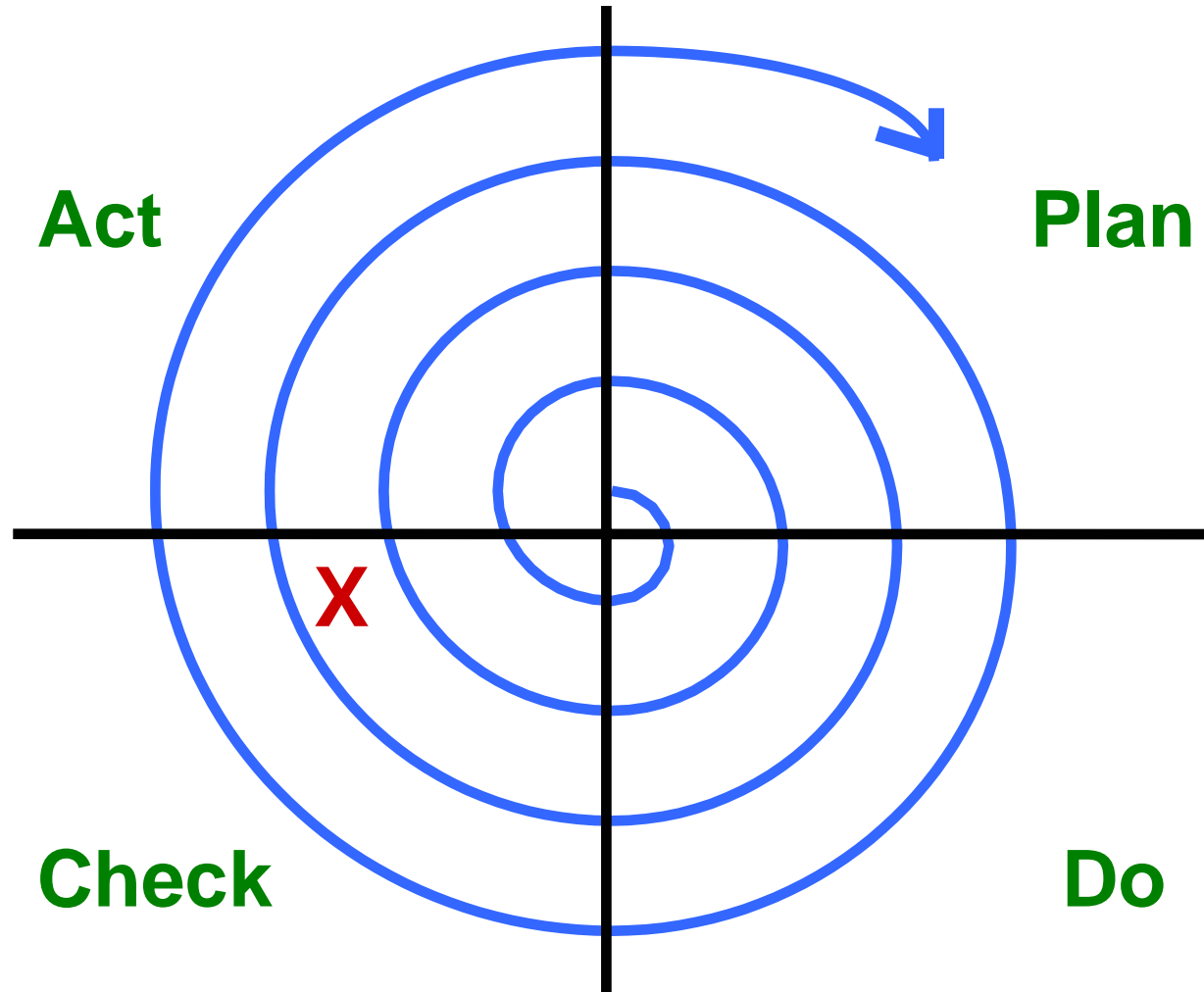
- Data dependency principle
 - First develop the applications that *create* data, then build those that *use* that data
- Implications
 - Foundation-building preempts instant gratification
 - This investment thinking requires tremendous business patience



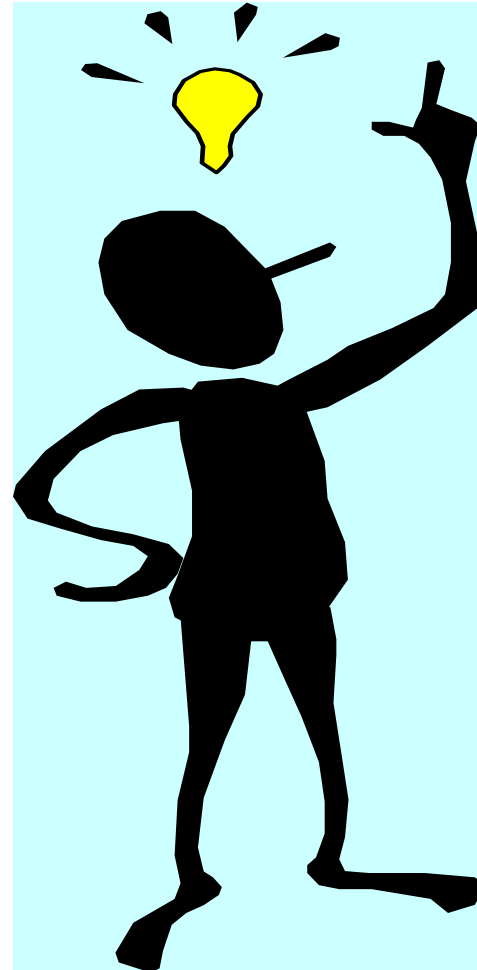
How EA Planning & Implementation Supports DQ

- Reduces the number of data capture systems and databases
 - Multiple disparate systems in different parts of the organization capturing the same fact adds **significant complexity** to data quality improvement efforts
 - **Single source data entry** improves the ability to assure data quality
- Reduces **inter-system data transfers** and potential for data handling errors
- Supports **data quality and business rule enforcement**
 - Database level
 - Application level
- Simplifies **data quality training** for new and transferred workers

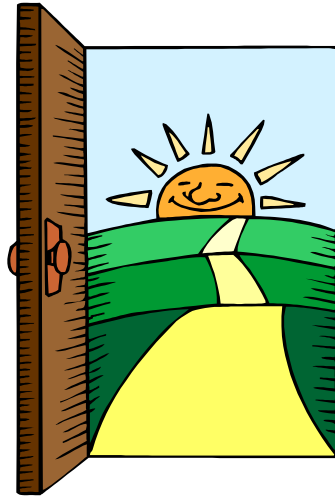
The End...of the Beginning....



QUESTIONS?



Building and Growing a Successful IQ Function



How do you do it? You just have to start.

"The journey of a thousand miles begins with one step"

Asian proverb

The Road Traveled

- Part I - Introduction and overview
 - IQ trends
 - Data properties and their impact
 - Issues and opportunities for the IQ leader
- Part II – Understanding information quality management
 - IQ methodologies & frameworks
 - IQ fundamentals
 - Measuring IQ value and cost
- Part III – Radical transformations
 - Managing data as an asset
 - Information stewardship
 - Managing information as a product
 - IQ in system development life cycle (SDLC)
 - Definitions and architecture
- Part IV – Wrap-up
 - The road traveled
 - Q&A
 - Check-out & evaluations

Check-out

- What stood out for you in this seminar?
- Which topics did you find most useful?
- What changes can you make the next two weeks to enhance the success of your IQ effort?

Acronyms

BI – Business Intelligence
BPR - Business Process Reengineering
CDI – Customer Data Integration
DBMS – Data Base Management System
DMT – Data Management Team
DQ – Data Quality
DQM – Data Quality Monitoring
ERP – Enterprise Resource planning
EHS – Environmental Health & Safety
ETL – Extract Transform Load
IQ – Information Quality
IM&T – Information Management and Technology
IT – Information Technology
PIM – Product Information management
RDW – Reporting Data Warehouse
SQL – Structured Query Language
TQM – Total Quality Management

References (1 of 3)

- English, L. P., *Improving Data Warehouse and Business Information Quality, Methods for Reducing Costs and Increasing Profits*, John Wiley & Sons, 1999
- English, L. P., "Information Governance and Stewardship: Implementing Accountability for the information Resources", *Proceedings of the Data Governance Conference*, San Francisco, CA, June 2007
- English, L. P., "Information Governance: Controlling Information as a Strategic Enterprise Resource", *IAIDQ Information and Data Quality Newsletter*, September 2007
- Fisher, C., Lauria, E., Chengalur-Smith, S., Wang, R., *Introduction to Information Quality*, MIT Information Quality Program, 2006
- Gartner Inc., "*Gartner's Data Quality Maturity Model*" by Andreas Bitterer, 7 February 2007
- Gartner Inc., "*Findings: BI Survey Shows Accessibility and Quality of Information is a Management Issue*" by Betsy Burton and Ted Friedman, 29 March 2007

References (2 of 3)

Gartner Inc., "*Researching the Research Agenda Survey 2007: Implications for Business Intelligence and Information Management*" by Betsy Burton and Tom Austin, 11 May 2007

Gartner Inc., "*Forecast: Data Quality Tools, Worldwide, 2006-2011*" by Colleen Graham, 8 June 2007

International Organization for Standardization, *ISO/IEC 1779 Information Technology – Security Techniques – Code of Practice for Information Security Management*, ISO, Geneva, 2005

IT Governance Institute, *COBIT 4.1*, IT Governance Institute, 2007

Levitin, A. V., Redman, T. C., "Data as a Resource: Properties, Implications and Prescriptions", *Sloan Management Review*, MIT Fall 1998

McGilvray, D., "Ten Steps to Quality Data and Trusted Information", *Proceedings of the 2007 Information and Data Quality Conference*, Las Vegas, September 2007

References (3 of 3)

TDWI, "*Taking Data Quality to the Enterprise through Data Governance*" by Philip Russom, March 2006

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

Redman T. C., "Process Management: The Engine That Drives Data Quality", *Proceedings of the Information and Data Quality Conference*, San Francisco, CA, October 2006

Redman, T. C., "The Body Has a Heart and Soul: Roles and Responsibilities of the Chief Data Officer", *IAIDQ Information and Data Quality Newsletter*, January 2007

Redman, T. C., "What Does it Mean to 'Manage Your Information and Data Assets?'" , *IAIDQ Information and Data Quality Newsletter*, July 2007

Web sites

www.iaidq.org; International Association for Information and Data Quality (IAIDQ)

www.iqtrainwrecks.com

www.isixsigma.com