

Institute of Information Management



University of St.Gallen

Towards a Framework for Corporate Data Quality Management

ACIS-07

Dr. Boris Otto, **Kristin Wende**, Alexander Schmidt, Philipp Osl
Toowoomba, December 6, 2007



Agenda

- Motivation
- (Corporate) Data Quality Management
- The CDQM Framework
- Summary

Business drivers for corporate data quality

Corporate Management/ Business Intelligence

- Poor data quality causes “blurry” management decisions
- No single point of truth
- Manual effort necessary during report creation

Compliance

- Legal and regulatory risks through bad or incomplete corporate data
- Contractual breaches and liability cases likely

Process Integration along the Value Chain

- Common material and partner data as a mandatory pre-requisite for efficient order-to-cash and procure-to-pay processes
- Necessity to establish unique data integration methodologies

Customer-centric Business Models

- One-face-to-the-customer requires consistent and sustainable customer and contract data management
- Data integration necessary on business unit and regional level

Electronic Product Information

- Customers and business partners demand high-quality electronic product information
- Information lifecycle management from F&E to Sales & Distribution

Data quality: Different requirements from different users or “Beauty is in the eye of the beholder”

How many customers may we correctly address?

How good is the quality of my data?

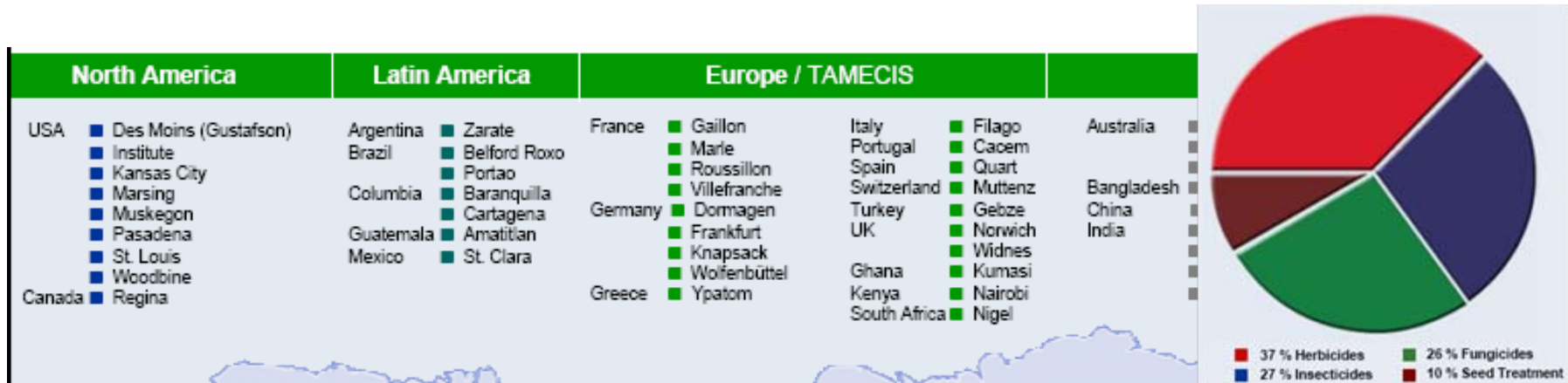
How many NULL values has the attribute CUS_ADR_KEY?

How many product master data have a valid article description?



- ➔ dependence of perceived quality from the user's needs
- ➔ “fitness for use” – the ability of satisfying the requirements of intended use in a specific situation

The focus of corporate data quality management (CDQM) is on companies looking for a corporate-wide approach to managing data quality



Data quality management is quality-oriented data management that comprises all organisational, methodical, conceptual and technical tasks related to managing data as an asset.

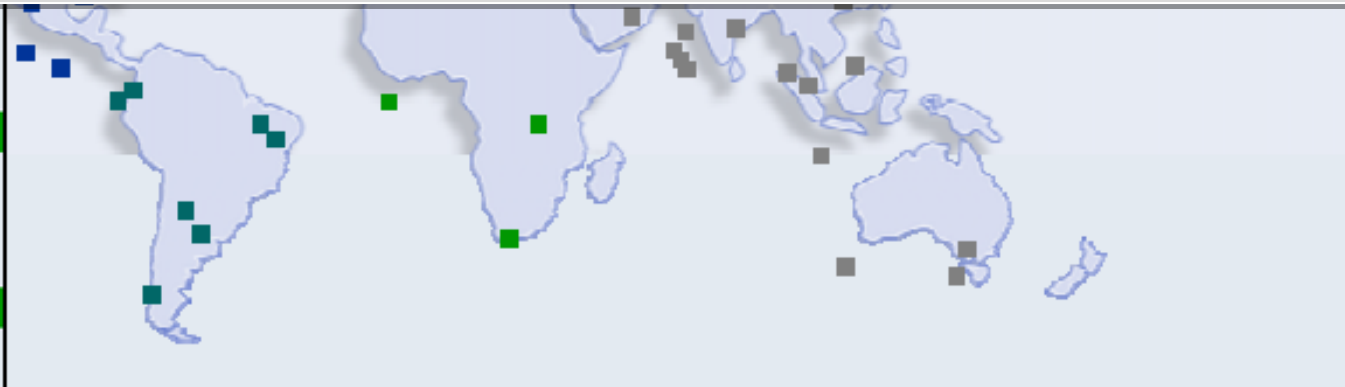
- Fungicides
- Herbicides
- Seed Treatment

Environmental Science

- Professional
- Consumer

BioScience

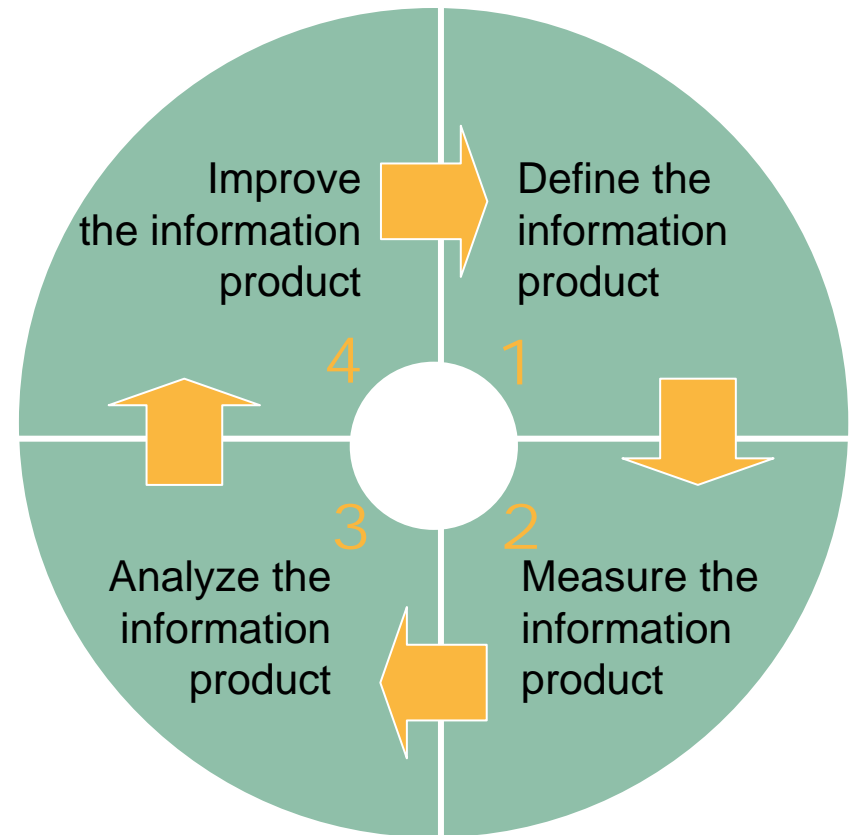
- Vegetables
- Agricultural Crops
- New Business Ventures



The Total Data Quality Management (TDQM) methodology from MIT is the best known approach to DQM

Information Product (IP) approach: “Treat information as a product”

- 1 Understand consumers' information needs.
- 2 Manage information as the product of a well-defined production process.
- 3 Manage information as a product with a life cycle.
- 4 Appoint an information product manager to manage the information processes and resulting products.



[Wang et al. 1998, Wang 1998]

An integrated approach that combines business-driven and technical perspectives on data quality management (DQM) is needed

Existing concepts for DQM fall short of:

- ➔ providing an integrated view incorporating a long-term, strategic orientation of DQM
- ➔ identifying relevant design objects
- ➔ defining organisational responsibilities
- ➔ specifying methods and tools for the realisation

The CDQM framework

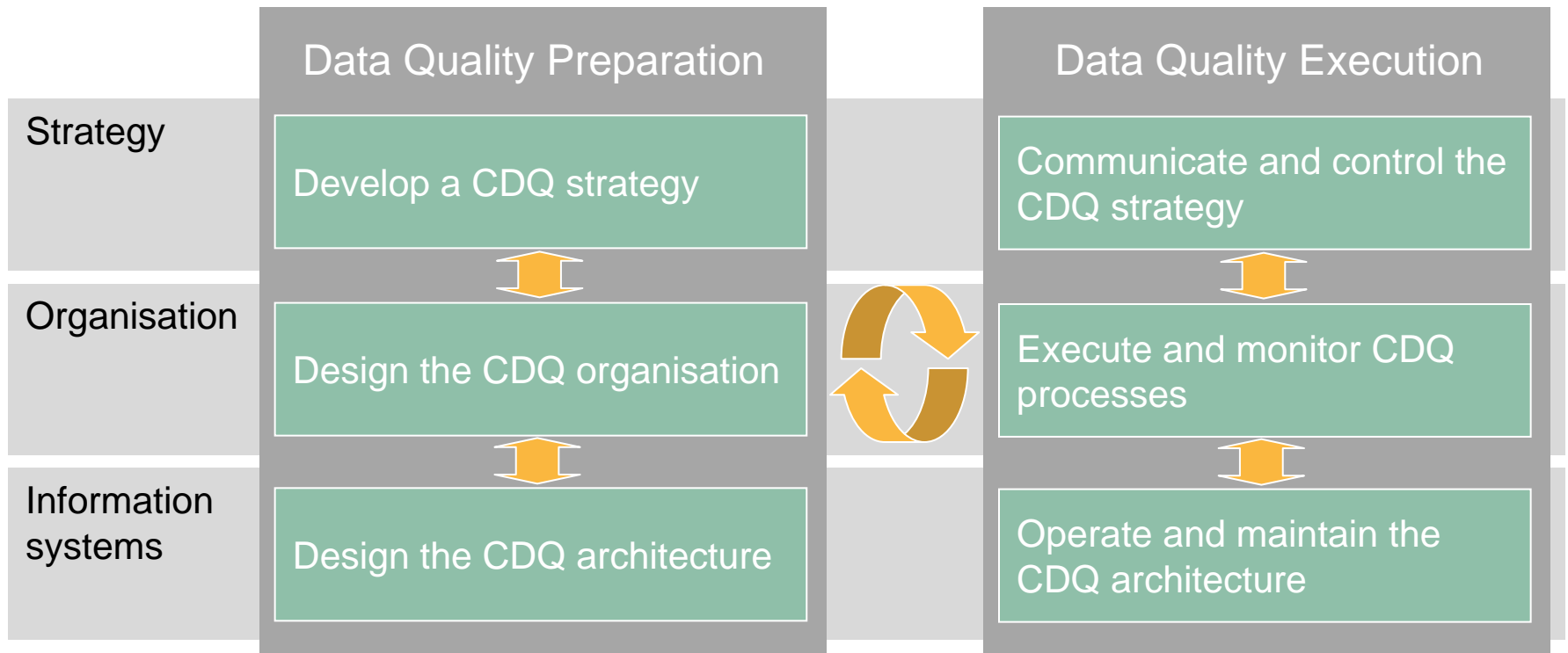


Structures relevant elements (design objects) and interdependencies of DQM at a high level of abstraction



Covers technical or system-related aspects and business-related and organisational issues

The CDQM framework distinguishes three horizontal layers and two columns



The preparation perspective sets the stage by defining what needs to be done, who is involved and how responsibilities are distributed

Develop a CDQ Strategy

- Develop a data quality strategy including strategic objectives
- Define a portfolio of strategic data quality initiatives
- Formulate the business case
- Carry out a status quo assessment and establish a review process

Design the CDQ Organisation

- Determine consumers' information needs
- Define “data manufacturing” processes
- Define roles and responsibilities across divisional boundaries
- Specify data quality metrics and standards
- Establish policies and procedures

Design the CDQ Architecture

- Develop a common information object model
- Create a business data dictionary
- Define information systems support

Example: Business Data Dictionary (BDD) – logical data model and implementation

Version 0.5-4.5 Welcome CDQUser

Business Data Directory

BDD Home Definitions Search Administrator Help

Actions

Modify Delete

Definitions

Definition Home Overview New Definition Search Admin Help

Help Home User Guide

Details View

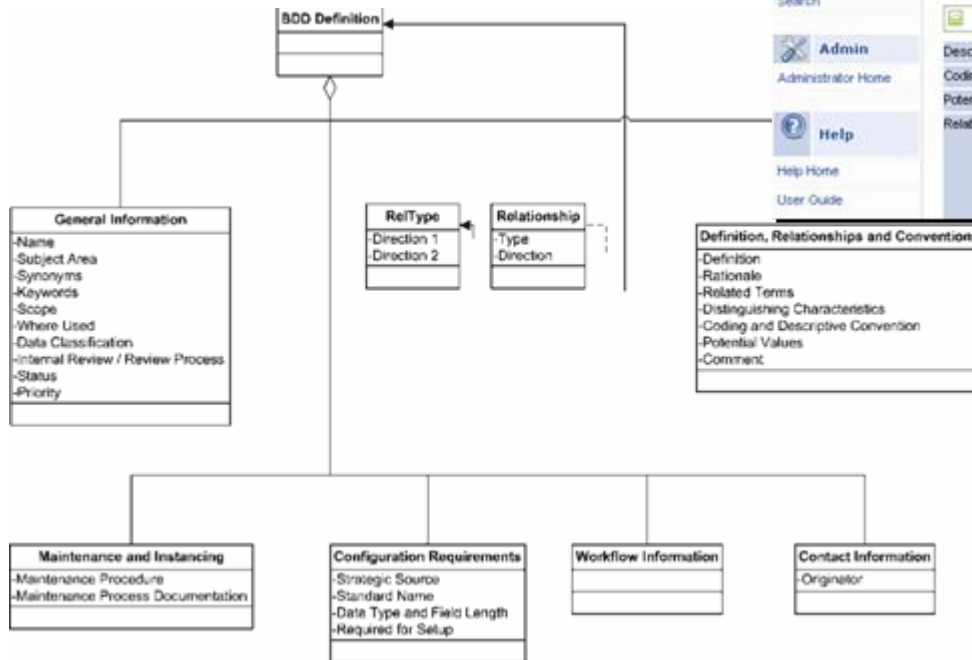
Last change: 19.04.2007 (Created on 19.04.2007 from Samuel Santochi)

Core Data

Name	Uhr	Synonyms	Synonyms MO
Subject Area	Product	Keywords	
Scope	Group	Data Classification	
Where Used		Contact	Santochi, Samuel
Definition	Zelmessegerät, das die Uhrzeit (bürgerliche Zeit) anzeigt.		
Rationale			

Conventions, Relation Ships, Potential Values

Descriptive Convention		
Coding Convention		
Potential Values		
Relation Ships	broader term	narrower term
	replaces	replaced by
	consists of	is part of



CDQ execution deals with the actual execution of data quality policies, the monitoring of data quality and operational problem-solving

Communicate and Control the CDQ Strategy
<ul style="list-style-type: none">• Develop and execute a communication plan• Carry out appropriate organisational change management measures• Cultivate a learning culture• Scan the environment for changes
Execute and Monitor CDQ Processes
<ul style="list-style-type: none">• Monitor data quality levels• Monitor efficiency and quality of data management processes• Introduce an effective training program
Operate and Maintain the CDQ Architecture
<ul style="list-style-type: none">• Operate and maintain systems for the storage and distribution of data• Operate and maintain systems for data quality analysis, for data cleansing and transformation, for meta data management, and for data management processes

Example: Monitoring data quality levels using a data quality scorecard

A1

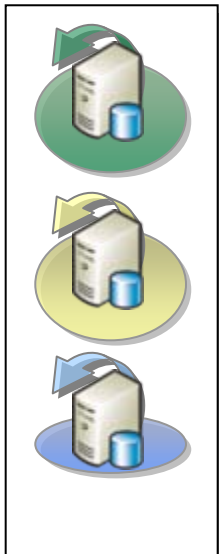
Error Rate Summary

Indicator

Description

Validation Rules are checked on Regional System. Deviations are calculated between Regional System and Golden Box. Error Rates are calculated across all rules and fields.

Links



		Validation Rules			Deviation from E44 (Golden Box)		
		Product Hierarchy	Consistency VP / UVP / Specification	Conversion Factors and Weights	Material Data	Customer Data	Vendor Data
Regional System	RS Europe	3,5%	2,2%	16,3%	1,6%		
	RS Asia Pacific	6,6%	8,0%	30,6%	9,0%		
	RS Americas	4,7%	3,8%	21,1%	0,7%		

Reporting Month: 2007-09

Future research: Extending the framework using the CMMI structure

- A** Data Quality Strategy
 - 1 Strategic data quality goals defined
 - 2 Business benefits identified
 - 3 Implementation plan defined
- B** Data Governance
 - 1 Accountabilities for data quality defined
 - 2 Corporate data quality guidelines and standards implemented
 - 3 Compliance with corporate strategy and legal requirements
- C** Management Information System
 - 1 Data quality measured
 - 2 Data quality monitored and reported
- D** Operational Data Management Workflows
 - 1 Data management processes implemented
 - 2 Cross-organisational data management processes implemented
 - 3 Data management processes monitored
- E** Data Architecture
 - 1 Corporate information object model established
 - 2 Architecture for data disposition designed
 - 3 Metadata management processes implemented
- F** Tool Support
 - 1 Cleansed and normalised data provided
 - 2 High degree of system integration

Key

- A** Process area
- 1 Goal

Summary

- Data quality management (DQM) is a support activity. It creates value by enabling strategic initiatives.
- An approach that combines business-driven and technical perspectives on DQM is needed.
- The CDQM framework encompasses design objects on a strategic, operational and information systems layer.
- It distinguishes two perspectives – preparation and execution.

Contacts



<http://cdq.iwi.unisg.ch>

Kristin Wende

University of St. Gallen

Institute of Information Management

Email: kristin.wende@unisg.ch

Phone: +41 71 224 7054