Data Quality Management: Framework and Approach for Data Governance

5th German Information Quality Management Conference

Dr. Boris Otto
Agenda

- Introduction to University of St. Gallen and the Research Context
- Business Drivers for Data Quality Management
- Data Governance: Design Elements and Implementation
- Summary
Introduction to University of St. Gallen (HSG)

- Founded in 1898

- 5,000 students on bachelor, master, and doctorate level
  - Two thirds of the students in business administration
  - Largest business administration faculty in Switzerland

- 80 professors, 35 permanent lecturers, 260 associate lecturers

- 30+ institutes

- 1,000 staff

- 50 per cent industry funding
Project partners of the Competence Center Corporate Data Quality (CC CDQ)

Bayer CropScience AG
Monheim, Germany

DB Netz AG
Frankfurt am Main, Germany

ETA SA Manufacture Horlogère Suisse
Grenchen, Switzerland

ZF Friedrichshafen AG
Friedrichshafen, Germany

Robert Bosch GmbH
Stuttgart-Feuerbach, Germany

Daimler AG
Stuttgart, Germany

Deutsche Telekom AG
Bonn, Germany

IBM Deutschland GmbH
Stuttgart, Germany

Data Governance
Bad Soden, November 22nd, 2007
© IWI-HSG / CC CDQ
CC CDQ research scope in the context of Business Engineering
Agenda

- Introduction to University of St. Gallen and the Research Context
- Business Drivers for Data Quality Management
- Data Governance: Design Elements and Implementation
- Summary
Business drivers for data quality management can be found literally across the entire enterprise

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
</table>
| 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 |
Consequences of poor data quality

- Inaccurate reporting: 81%
- Arguments over which data is appropriate or trusted: 78%
- Bad decisions based on incorrect definitions: 54%
- Data governance and stewardship limitations: 53%
- Limited visibility for data lineage and linkage: 52%
- No understanding of master data homonyms, synonyms: 46%
- Poor customer service: 35%
- Inefficient marketing: 32%
- Inefficient purchasing/sourcing: 18%
- Delay in new product introductions: 17%
- Other: 8%

Source: Russom, P.: Master Data Management - Consensus-Driven Data Definitions for Cross-Application Consistency. The Data Warehousing Institute, 2006
Agenda

- Introduction to University of St. Gallen and the Research Context
- Business Drivers for Data Quality Management
- Data Governance: Design Elements and Implementation
- Summary
Data quality management in a corporate-wide context

Companies have difficulties with institutionalizing data quality management

• Definition of responsibilities
• Integration in organizational structure
• Enforcement of mandates

Different internal and external stakeholders with divergent interests

• CxOs, IT, Sales, Procurement, Controlling, Business Units, Customers etc.
• Company-wide requirements, laws and regulations vs. regional and local differences
• Separation of data maintenance and data usage
Data quality management without data governance

Data governance specifies the framework for decision rights and accountabilities as part of data quality management

**Data Governance**

- Definition of roles and assignment of responsibilities for decision-areas to these roles
- Definition of organization-wide guidelines and standards and assurance of compliance to corporate strategy and laws governing data.

**Data Governance Model**

- Combination of roles, tasks and responsibilities for data quality management.
Data governance defines clear roles and responsibilities for data quality management

The current state-of-the-art suggests 4 roles and 1 committee for data quality management as a common denominator*

- **Executive Sponsor**
  - Provides sponsorship, strategic direction, funding, advocacy and oversight for data quality management.

- **Data Quality Board**
  - Defines the data governance framework for the whole enterprise and controls its implementation.

- **Chief Steward**
  - Puts the council’s decisions into practice, enforces the adoption of standards, helps establishing data quality metrics and targets.

- **Business Data Steward**
  - Details the corporate-wide data quality standards and policies for his area of responsibility from a business perspective.

- **Technical Data Steward**
  - Provides standardised data element definitions and formats and profiles source system details and data flows between systems.

* The actual number of roles may vary from company to company.
Data governance sets standards and guidelines for the major decision areas and tasks in data quality management

**Data quality strategy**
Alignment with business strategy, business benefits of DQM, measurement and control of DQM

**DQM controlling**
Measures and metrics, scorecards, incentive systems

**Organization and processes**
Assignment of roles, definition of data management processes

**Data architecture**
Meta data definition, logical data modeling, global vs. local distribution of information objects

**System architecture**
Functional architecture for data management, standards for data repository, data cleansing, data distribution
Excursus: A benefit tree for data quality management

Support Activities

- Firm Infrastructure
- Human Resource Management
- Technological Development
- Procurement

Primary Activities

- Inbound Logistics
- Operations
- Outbound Logistics
- Marketing & Sales
- Service

Data Governance
Bad Soden, November 22th, 2007
© IWI-HSG / CC CDQ
### Example: Key performance indicators for data quality in the retail industry

<table>
<thead>
<tr>
<th>Kennzahl</th>
<th>Bezugsgroße</th>
<th>Berechnungsvorschrift</th>
<th>Ebene</th>
<th>Periodizität</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formatwechsel</td>
<td>Wert</td>
<td>(Absolutwert der Formatwechsel) / Verbrauch * 100</td>
<td>Filiale, Abteilung</td>
<td>Monatlich</td>
</tr>
<tr>
<td>Pseudo-Bepo</td>
<td>Wert</td>
<td>(Wert Pseudo-Bepo) / Wert Bepo Gesamt * 100</td>
<td>Filiale, Abteilung</td>
<td>Monatlich</td>
</tr>
<tr>
<td>Minusbestand</td>
<td>Anzahl</td>
<td>(Anzahl Bepo ohne Bestand) / (Anzahl Bepo Gesamt) * 100</td>
<td>Filiale, Abteilung</td>
<td>Monatlich</td>
</tr>
<tr>
<td>Inventurbestand ohne Bestellpositionen</td>
<td>Wert</td>
<td>(Inventurbestand ohne Bepo) / Inventurbestand * 100</td>
<td>Filiale, Abteilung</td>
<td>Jährlich</td>
</tr>
<tr>
<td>EK-Differenzen</td>
<td>Anzahl</td>
<td>(Anzahl fehlerhafte Repo) / (Anzahl Repo Gesamt) * 100</td>
<td>Abteilung</td>
<td>Monatlich</td>
</tr>
<tr>
<td>Rechnungen ohne Auftrag</td>
<td>Anzahl</td>
<td>(Anzahl Rechnungen ohne Auftrag) / (Anzahl Rechnungen Gesamt) * 100</td>
<td>Filiale, Abteilung</td>
<td>Monatlich</td>
</tr>
<tr>
<td>Fehlerlisten</td>
<td>Anzahl</td>
<td>(Absolutwert der Menge mit Fehlern) / (Absolutwert der Gesamtmenge) * 100</td>
<td>Filiale, Abteilung</td>
<td>Monatlich</td>
</tr>
<tr>
<td>Stapf-Korrekturen</td>
<td>Wert</td>
<td>Wert der nachträglichen Ergebniskorrekturen</td>
<td>Filiale, Abteilung</td>
<td>Monatlich</td>
</tr>
</tbody>
</table>

## Excursus: Demarcating business data dictionaries from related concepts

<table>
<thead>
<tr>
<th>Evaluation Criteria</th>
<th>Glossary</th>
<th>Data Model</th>
<th>Data Dictionary</th>
<th>BDD</th>
<th>Classification Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantically enriched definitions</td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
</tr>
<tr>
<td>Mapping of relationships</td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
</tr>
<tr>
<td>Intelligibility for potential users</td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
</tr>
<tr>
<td>Manageability, Maintainability</td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
</tr>
<tr>
<td>User friendly integration into other applications</td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
<td><img src="Image" alt="Rating" /></td>
</tr>
</tbody>
</table>

### Key:
- **Very high**
- **High**
- **Medium**
- **Low**
- **Very low**

*Data Governance*  
Bad Soden, November 22nd, 2007  
© IWI-HSG / CC CDQ
### Core Data

<table>
<thead>
<tr>
<th>Name</th>
<th>Finished Product Dimension</th>
<th>Alias</th>
<th>Product Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Area</td>
<td>Item - Outbound</td>
<td>Keywords</td>
<td>Height Width Depth</td>
</tr>
<tr>
<td>Lead Project</td>
<td>GRD</td>
<td>Data Classification</td>
<td>Unclassified</td>
</tr>
<tr>
<td>Version</td>
<td>GRD V1.0</td>
<td>Scope</td>
<td>Global</td>
</tr>
<tr>
<td>Where Used</td>
<td>Apollo ATLAS Outbound CPF</td>
<td>Internal Review</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Mercury</td>
<td>Originator</td>
<td>David A. Ellis</td>
</tr>
</tbody>
</table>

| Definition | Finished Product Dimension is the dimensions of a finished product in terms of its depth, width and height as defined by the Mars measuring standards for Finished Products and consistent with the GS1 Standards. |
| Definition Links | General EAN.UCC Specifications, See Section 6.0.1.4 ---- http://194.203.97.138/EANUCC/|

| Rationale | The size of a finished product is required throughout the business and also by our customers for shelf planning etc. It is very important that we have a consistent approach to defining the measurements which is in line with the global standards such as EAN.UCC (GS1) etc. The attached give the standards that we should adopt throughout the business. In the meantime we are at Version 1.5, revised November 2006. |

For the assignment of responsibilities, concepts like the RACI notation may be used:

**R**esponsible: Responsible roles perform tasks or specify the way of performing tasks.

**A**ccountable: Accountable roles authorize decisions or the results of tasks.

**C**onsulted: Consulted roles have specific knowledge necessary for decision-making or performing tasks.

**I**nformed: Informed roles will be informed about decisions made or results of tasks.
Data governance materializes in a responsibility matrix

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Roles</th>
<th>Executive Sponsor</th>
<th>Data Governance Council</th>
<th>Chief Data Steward</th>
<th>Technical Data Steward</th>
<th>Business Data Steward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data quality strategy</td>
<td></td>
<td>A</td>
<td>R</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measures and controls</td>
<td></td>
<td>A</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data management processes</td>
<td></td>
<td>A</td>
<td>R</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Metadata management and business data dictionary</td>
<td></td>
<td>A</td>
<td>R</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data architecture</td>
<td></td>
<td>A</td>
<td>R</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System architecture</td>
<td></td>
<td>I</td>
<td>A</td>
<td>R</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Data quality tools</td>
<td></td>
<td>A</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key: R - Responsible; A - Approver; C - Consulted; I - Informed.
There is no “one fits all” solution - in analogy to IT management, contingencies determine the individual data governance model.
The establishment of a data governance model is a change process.

- Analyze contingencies
- 1st draft of the model
- Kick-off workshop
- 1-on-1 workshops w/ business units
- Agreed data governance model
- Roll-out

Involved roles:
- Core team
- Sponsor
- Data governance council
- Process owners

Data Governance
Bad Soden, November 22nd, 2007
© IWI-HSG / CC CDQ
How to ensure failure in a data governance program

- Silver bullet syndrome
- Ivory tower syndrome
- Poor organizational change management
- Lack of executive sponsorship

Don’t forget: Simplicity is the clearest policy

Agenda

- Introduction to University of St. Gallen and the Research Context
- Business Drivers for Data Quality Management
- Data Governance: Design Elements and Implementation
- Summary
Summary

- Multiple business drivers for data quality management exist - throughout the entire organization of a company.

- However, organizations have difficulties assigning responsibilities for data quality management.

- Data governance is an instrument for the strategic alignment of data quality management and the definition of standards.

- It is no “one fits all” concept, but is rather influenced by a company’s contingencies.
Contact

Dr. Boris Otto
University of St. Gallen
Institute of Information Management
boris.otto@unisg.ch
++41 71 224 32 20