

8th International Conference on Information Quality, 2003

Data Diagnosis: Making DQ Assessment Work

Nigel Totterdell MSc.
MahindraBT
nigel.totterdell@mahindraBT.com

Executive Summary/Abstract
Data Diagnosis - a practical "RAD" approach to data quality from several years working with tricky telecommunications data. The lessons learned are applicable across any data in any sector. This presentation shares how these tactics can address gaps in data knowledge, ensure early wins, at the same time underpinning a "TQM" approach to data quality across an organization.

1

8th International Conference on Information Quality, 2003

Objectives of this presentation

- Set the scene – the "Telco" origins
- Introduce data knowledge, it's significance and how TDQM helps
- Position data discovery and profiling
- Introduce the data diagnosis framework
- Discuss "Low Hanging Fruit" payback opportunity(s)
- Align data diagnosis to competitive advantage
- Show data diagnosis underpinning Information Quality system(s)

2

8th International Conference on Information Quality, 2003

Outline and Structure

- Brief overview of typical global telecommunications delivery processes and data resource requirements
- How Data Quality "Best practice" can be brought to bear
- The data quality assessment process followed
- The typical detail result and business benefit
- Reference(s)

3

8th International Conference on Information Quality, 2003

Brief overview of typical global telecommunications delivery processes and data resource requirements

- The typical Telecommunication's provisioning information chain
- The intricacies of telecommunication's data
- "Real World" – the data quality risks experienced

4

8th International Conference on Information Quality, 2003

Tier 1 Telco's service operation

5

8th International Conference on Information Quality, 2003

Network centric data is tricky!!

Domain Identification	Engineering detail	Technical Attribution	Reporting Attribution
<ul style="list-style-type: none"> • Location • Device Type • Device Identifier • Component Type • Component Identifier • Date • Subject 	<ul style="list-style-type: none"> • Capacity (STM4) • Component Identifier (1+P1+1) • Configuration (JK-LM) • Assignment (connect-u) • Connection (a to z) 	<ul style="list-style-type: none"> • Status • Protection • VC12/DOS Equivalences • a or z end characteristics • Facility • Service 	<ul style="list-style-type: none"> • Geographic • LOB • Period • Forecast/actual • Monetary equivalent • Performance measure • Organisation • Customer • Ownership • Confidence limit

6

8th International Conference on Information Quality, 2003

Network data domain the reality

- Missing or imprecise data integrity rules are common due to difficult data structures, lack of domain expertise and poor documentation
- Data error detection and correction is expensive and tricky, weighed down by inadequate metadata
- Everyone recognizes meta data shortcomings, just don't know what to do about it, it's not uncommon for meta data to be missing altogether
- Only a fraction of data created automatically is **a) used** and **b) important**
- Many major data quality problems are down to poor training and lack of validation, they are relatively easy to fix.

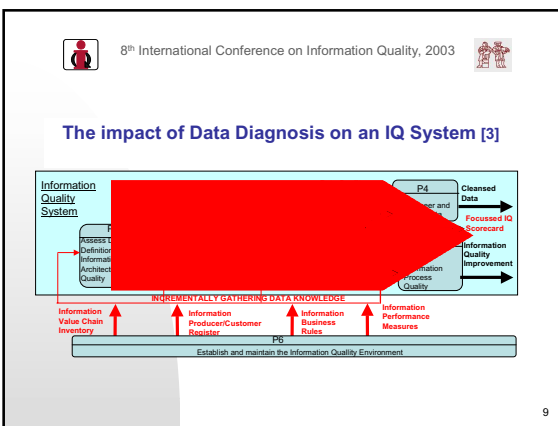
7

8th International Conference on Information Quality, 2003

How Data Quality "Best practice" can be brought to bear

- Data Diagnosis within TDQM based Information System
- What does Data Quality mean
- How to make Data Quality Assessment work

8



8th International Conference on Information Quality, 2003

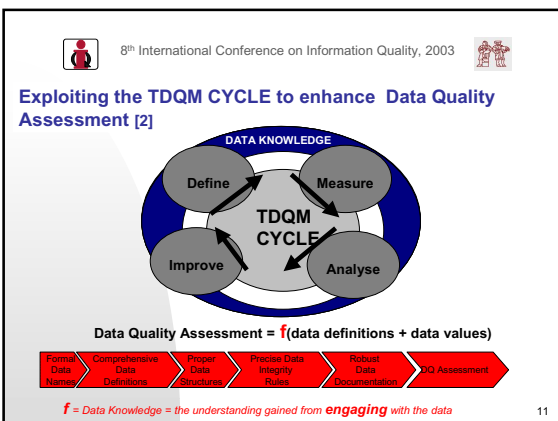
The challenge "Fit for purpose" data health check [1]

Information Quality - Defined
Data are of high quality if they are fit for their intended uses in Operations, decision-making, and planning (after Juran).

Importance and worth

- Free from defects**
 - accuracy
 - integrity
 - correctness
 - completeness
 - consistency
 - ...etc
- Possess desired features**
 - comprehensive
 - relevant
 - current
 - appropriate
 - ...etc

10





8th International Conference on Information Quality, 2003

The data quality assessment process followed

- A Data Discovery phase to build Data resource knowledge
- Data Diagnosis as a systematic data content test and scoring mechanism
- The resulting Data Quality scorecard

12

154

8th International Conference on Information Quality, 2003

References

[1] Redman Thomas C. Data Quality: The field guide, Digital Press, ISBN 1-55558-251-6
[2] Wang Richard "Raising the bar for Data Quality in the new Millenium"
(<http://www.niss.org/affiliates/dqworkshop/presentations/wang/presentation.pdf>)
[3] English Larry P. Improving Data Warehouse and Business Information Quality,
John Wiley & Sons, ISBN 0-471-25383-9

19