



How to make a start: “DQ Health Check”

Nigel Totterdell MSc.

Data Diagnostic Services

nigel.totterdell@datadiagnosticservices.com

Executive Summary/Abstract: Although many acknowledge the *Data Quality Risks*, the scale of the challenge and the confusing messages on tackling it can be very off putting, and knowing where to start a major obstacle. This presentation explains how to deal with these dilemmas



Objectives of this presentation

- Discuss what DQ assurance must deliver
- Explain how to implement the TDQM cycle
- Introduce the iterative DQ engagement model
- The role of Data Diagnosis methodology
- Describe Data Diagnostics for incremental DQ testing
- Examine the components of a DQ assurance solution
- Why a Data Health check can trigger DQ assurance
- The DQ Health Check judgement framework
- Explore opportunities for DQ “Vital Signs” reporting
- Explain how a DQ attribute model will evolve

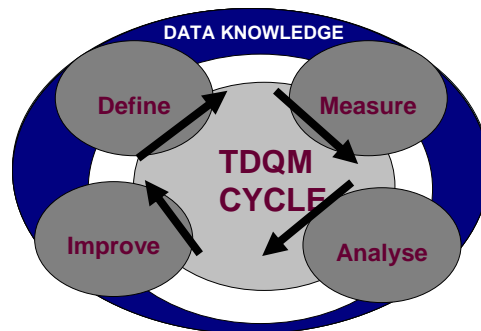


What DQ Assurance must deliver

- To score data quality and monitor its quality as a result, the data must be tested based on itself, with other data stored in the same or in a different system(s) and compared with real-world counterparts.
- DQ measurements need to provide base facts in a scoring scheme composed of several parallel Data Quality Dimensions, or summation hierarchies.
- An Overall Data Quality Score to help Data Consumers in judging data quality, particularly when the data involved come from different sources but relate to the same topic.



How to implement the TDQM cycle



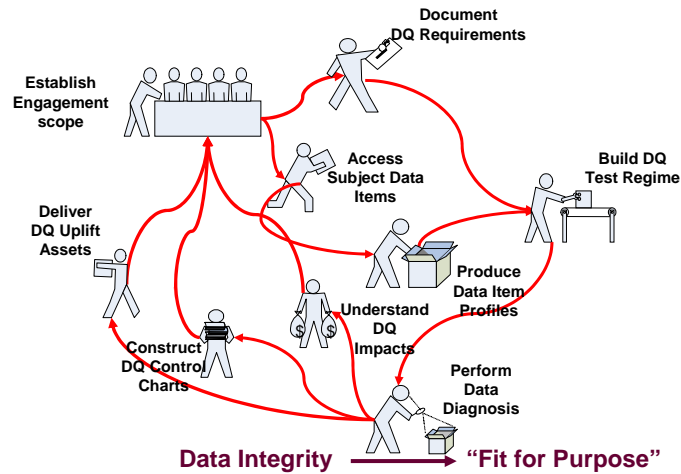
Data Quality Assessment = f (data definitions + data values)



f = Data Knowledge = the understanding gained from **engaging** with the data



Data Diagnosis – an Iterative DQ engagement model



5



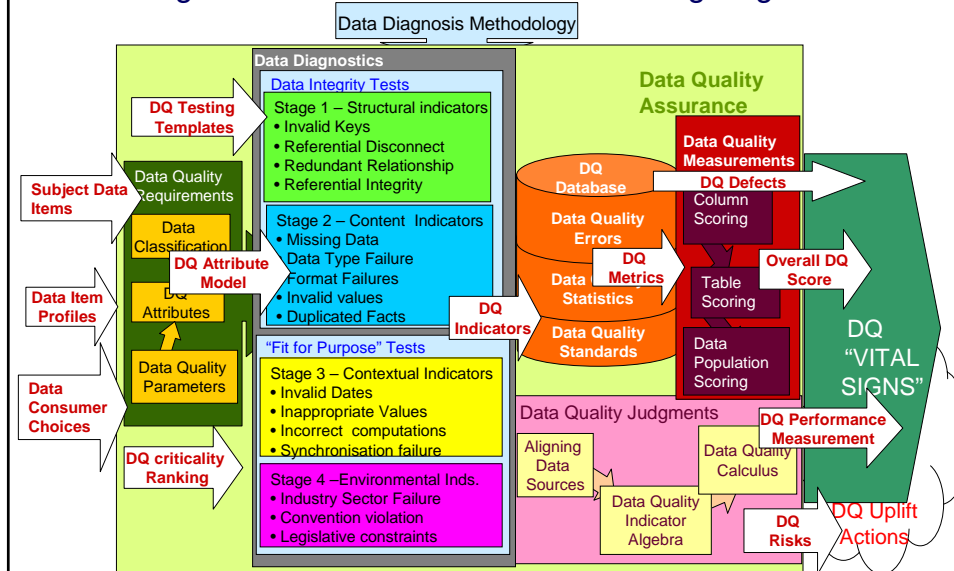
Data Diagnosis Methodology

- Data Diagnosis retrospectively establishes Data Quality Requirements for "live" data base solutions, producing DQ attributes
- Deploying pre-built DQ test templates against the nominated Information Source and selected subject data items to make objective "pass" or "fail" assessments.
- Concentrates on mining for data quality problems related to accuracy, integrity, believability and validity as a pre-cursor and in support of asking questions regarding Information Quality .
- A standard set of Data Quality indicators are produced to establish the Data Quality score of the studied DQ attributes.
- A criticality ranking combined with the data Item profile output and resulting data classification scheme are used to assess the applicable level of Data Quality to expect

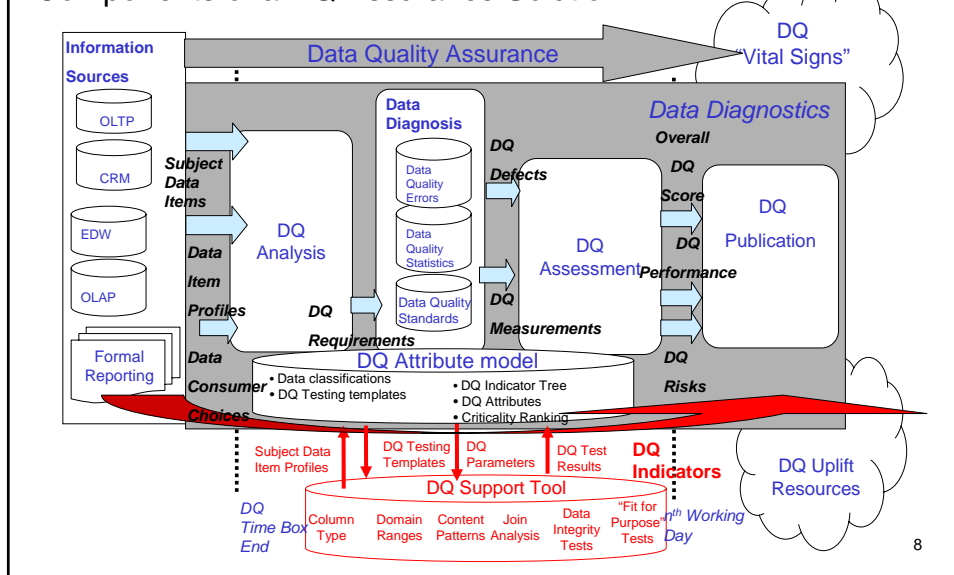
6



Data Diagnostics an "Incremental" DQ Testing Regime



Components of a DQ Assurance Solution





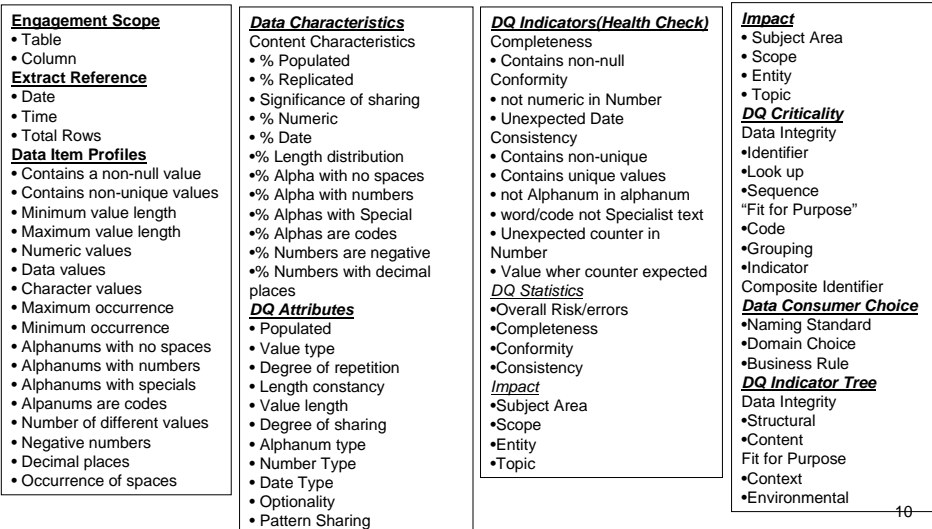
Getting Started ... a DQ "Health Check"

- Data Quality attributes split neatly into those that relate to Data Integrity Indicators and those custom characteristics that establish "fit for purpose" data content criteria.
- **Consistency and completeness are two important dimensions of data quality assessment and require minimal contact with the Data Consumers.**
- Provides an initial DQ Score validates the community held assumption that data stored in databases is mostly of good quality - or provides a "base line" of how bad things really are.
- Whilst the end user may be equally, or more interested, in risks associated with the applicability of the data to their needs, Data Integrity testing is an important pre-requisite to entering into the more sophisticated levels of DQ analysis and assessment.



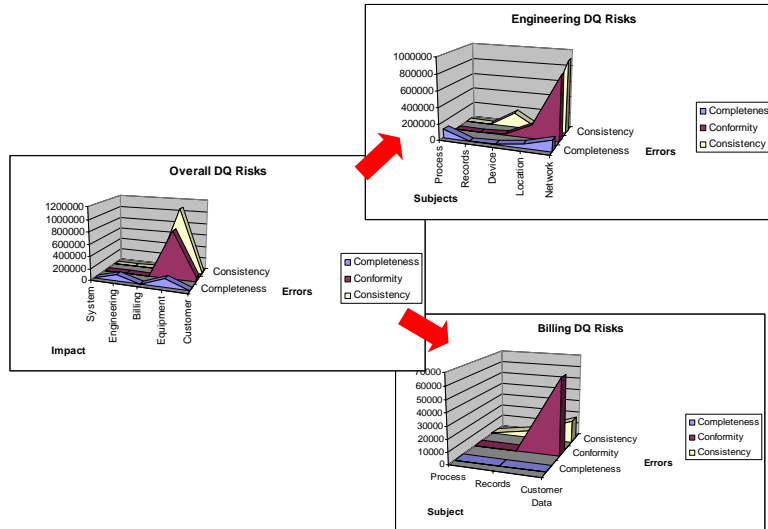
Data Health Check DQ Judgement mechanism

Subject Data Items → DQ Analysis → Data Diagnosis → DQ Assessment





DQ “Vital Signs” ... communicating in Business Terms



DQ Attribute Model from engaging with the data

