One Version of the Truth
The Impact of Poor Data Quality on Organisational Knowledge Management
Gilles Rabaud and Mike Ogle
Company: Simpl.

1. Introduction to Data Quality Hot Spots

Poor data quality impacts not only an organisation's internal effectiveness, ability to manage knowledge and learning, but also its customers. That impact is manifested at five key data quality Hot Spots:

1. Sources – how the data is being collected and the quality commitment from providers
2. Destinations – how the data is being stored, accessed and reported on and the quality commitment from users
3. Ownership – The quality commitment from those who are responsible for a holistic view
4. Usefulness – The ability for the organisation to convert data into actionable information
5. Business Item – Clear organisation-wide understanding of the "data item" and its definition, business rules, parameters etc.

The Data Chain

These data quality Hot Spots can be viewed across the life cycle of a single data unit as depicted, in what we have called the Data Chain below.
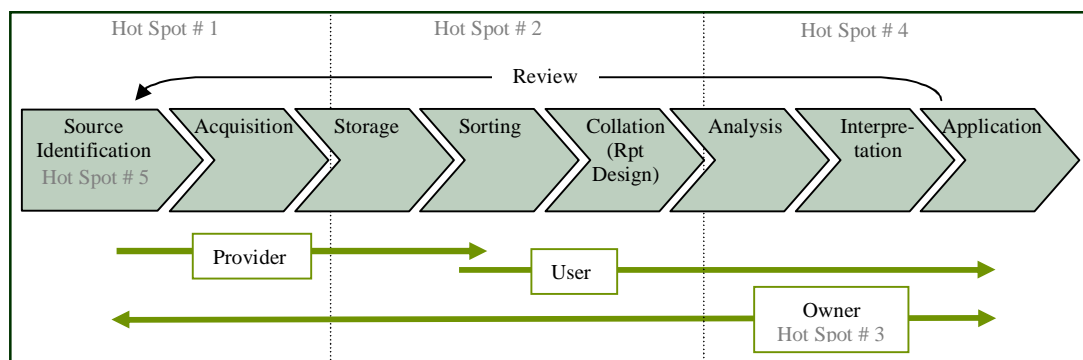
Achieving data quality requires a joint effort across the business. This, coupled with ensuring that quality is intrinsic in the way an organisation does things, will result in 'One Version of the Truth'.

2. Auditing for Quality Data

Initiatives to improve data quality will not stand on their own; they need to be executed in a wider strategic context, becoming inherent in organisation-wide and group business plans.

When auditing data quality, it is important to understand static and dynamic views of data. Cleansing actions include detection, correction, filtering and prevention. These actions must be underpinned by appropriate and collaborative data ownership. Data ownership spans custodianship, business ownership and stewardship. These responsibilities do not reside with one single stakeholder in the business, hence the need for collaboration.

The benefits of high data quality are well documented, with the prime benefits being access to one version of the truth, and increased stakeholder satisfaction.

3. Symptoms Pointing to Poor Data Quality

When business data is subject to differing interpretations it becomes very difficult to gain organisation-wide commitment to any problems that the data might highlight. To regain confidence in the data, agreement

must be gained on the fundamental issue of data unit definition, where data should be sourced from and what the appropriate acquisition method is. This will provide an agreed basis to begin from. Continual communication to providers, users and owners will create understanding of the importance of the data and its possible impact on future decisions.

With the advent of online web applications and customer self service, customer dissatisfaction arising from inaccurate information has become a growing problem. Additionally, due to this greater transparency, other stakeholders such as shareholders and community groups now have the potential to be negatively influenced by inaccurate data.

Poor data quality arising from issues such inconsistent collection, over reliance on "catch all" categories or data deliberately being manipulated increases a need for more reconciliation and double checking of information. This results in process inefficiencies, staff dissatisfaction, increased cost and is a major contributor to a lack of confidence in all data in "the system".

4. Drilling Down on Hot Spots

Examples of specific Hot Spots to consider when evaluating data quality issues are depicted in the diagram below. The Business Item is the specific data e.g. client. This data will have a definition and associated business rules for example, for the client data to be complete it must include salutation, first and last name. The business rule might state that first name is recorded as the real name e.g. Robert instead of Bob.
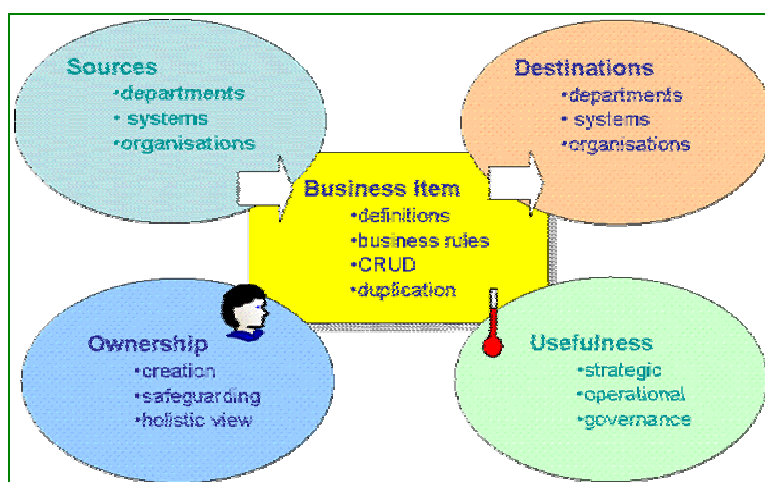
Source, Destination, Ownership and Usefulness are somewhat self explanatory in this diagram. However issues such as changes to source systems have an obvious flow on effect to Destinations (one persons source is another's destination) which raises the issue of Ownership and altermately creates an impact on data Usefulness.

5. Data Quality is a Collaborative 'Way of Life'

Data quality requires attention at all points; front office, back office and customer input. If this can't be achieved a sophisticated information (data) architecture is useless. This is the 'people' and the 'process' component of continuous improvement with benefits from the technology component only being maximised once the former are in place.
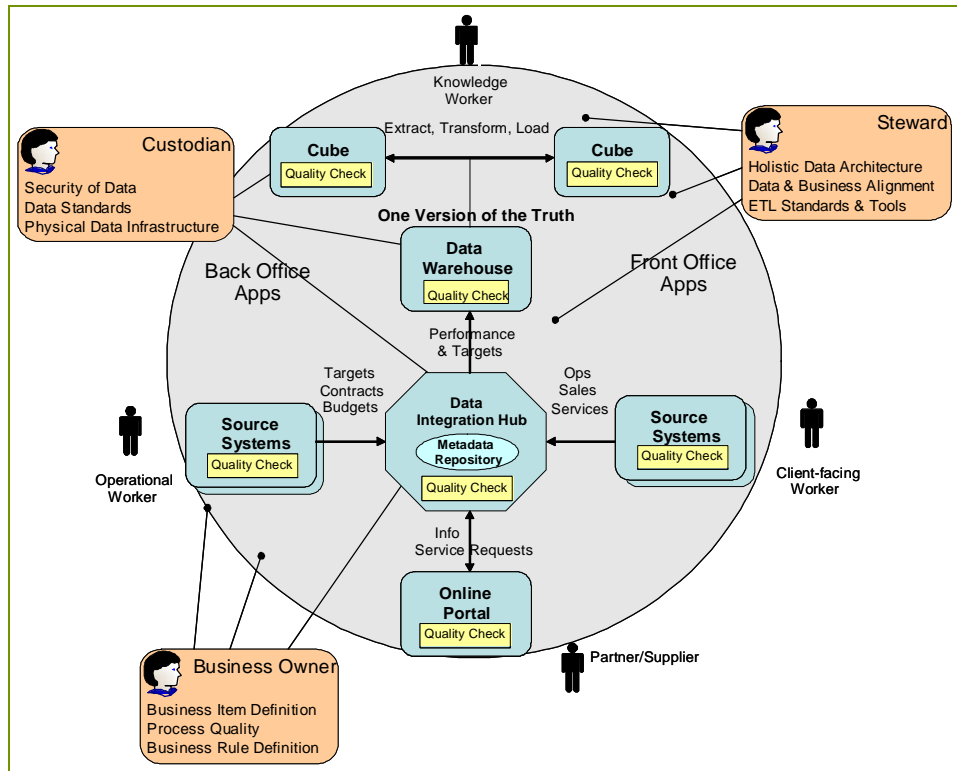
This diagram (above) shows the interrelationships and therefore why data quality is a joint responsibility, not just that of IT. Collaboration should be intrinsic in the way an organisation manages data quality, not an afterthought. If true collaboration is achieved the concept of One Version of the Truth is made possible, delivering the business benefits as shown in the last section of this paper.

6. A Meaningful Context for the Business

Data Quality initiatives need to have a meaningful context for the business as people are often heard to say "we have heaps of data but can't get any information".

Simpl's experience shows that business will not invest in Data Quality for the sake of quality. Because of this, it becomes important to demonstrate the adverse impact of poor Data Quality on the organisation using simple tools such as current / future state reviews, route cause analysis and gap analysis.

7. A Simple Approach to the Data Quality Audit

Simpl's approach to a detailed audit is presented in the diagram below. The audit is complemented with static and dynamic views of the impact on data and 'fix it' actions.
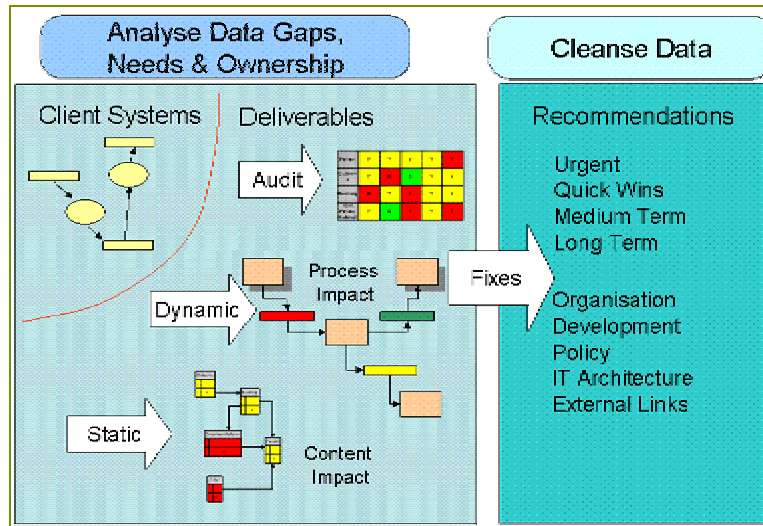
The left of the diagram depicts Data Gaps, Needs and Ownership including:

- Client Systems - Processes, applications, modules in applications
- Data Audit - 'traffic light' analysis of data, based on five Hot Spots
- Process - Dynamic impact of data quality
- Data Entities - Static content impact of data quality

The right of the diagram depicts Cleansing Data and the required investment to:

- Correct – e.g. Mr vs. Mister, Misspellings, SQL for relational navigation
- Filter - Duplicates, missing, nonsensical data elements (e.g. during Extract, Transform, Load)
- Detect and Report - e.g. decide what to do with old data
- Prevent - education of data entry staff, building validations, updating outdated codes

8. Benefits of High Quality Data

Inherently most people in an organisation understand, at some level, the benefits for them of acquiring and maintaining high quality data. These benefits will vary from industry to industry and even between companies in the same industry however there are some common benefits that have been identified.

The graph below shows the six top common issues high quality data is perceived to address for organisations.
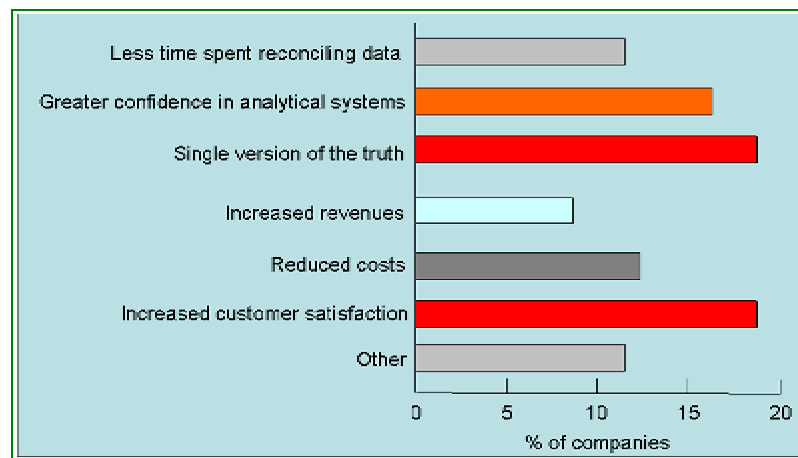
9. The Takeouts from this Paper

- Data quality is experienced in five hot spots
- Data quality must be put in context that's valuable to the business
- Data quality initiatives require collaboration across the business and IT groups
- Data quality is not a project, it must become intrinsic in the way data is acquired, stored and reviewed
- The objective should be "One Version of the Truth" from which many benefits can be derived