Original Research Paper Management

ISO Certified and Impact Analysis of Data Quality Framework

Srashti Rane MBA-IT, SICSR, affiliated to Symbiosis International University (SIU), Pune, Maharashtra, India

Pravin Metkewar Assoc. Professor, SICSR, affiliated to Symbiosis International, University (SIU), Pune, Maharashtra, India

ABSTRACT

Data Quality is cross-discipline, precondition for analyzing and domain specific problem due to the importance of fitness use in the data quality metric. However, the changing landscape of data quality challenge tells the need for holistic solution. As a step on bridging any gaps in the various research communities, we undertook a comprehensive literature study of data quality is been done. First, this paper summarizes reviews of data quality research. Second, this paper analyzes the data characteristics of the data environment, presents quality challenges faced by data, and expresses a hierarchical data quality framework from the views of various data users. This framework consists of quality dimensions, quality characteristics, and different quality indexes. Finally, on the basis of the framework, the paper constructs a dynamic assessment process for data quality. This process consists of good extensibility and flexibility and can meet the requirements of data quality assessments. The research results enrich the theoretical scope and model of data and lay a solid foundation for the future for establishing an assessment model and studying evaluation algorithms.

KEYWORDS :Data Quality, Literature Survey, Research Framework, Data warehouse, data quality, metadata, stakeholders, quality parameters, framework, ETL, Data Staging, Data Warehouse.

INTRODUCTION

The impact of poor data quality results in decision making, organizational trusts and satisfaction. However, the changing nature and increasing volume of data has exacerbated the level and, thus, increase manifold the stakes involved. Data quality management is complicated by many emerge factor. First, are the clear implications that relate to the sheer volume of data produced by different organizations today. Second, recent years have seen an increase in the diversity of data. Such diversities refers to structured, unstructured, semi-structured data, and multi-media data such as video, maps, images, etc. Data has an increasing number of sources. The use of various technologies, for example, sensor devices, medical instrumentation, RFID readers, etc., further increases the amount and diversity of data being collected. More subtle factors also exist - such as the lack of clear alignments between the intention of data creation and its subsequent usage. A prime and unique example of such lack of alignment is the vast amount of data collected from social networks that can then be used, without assessment of different quality, as a basis for marketing decisions. A related factor exists that relates to difficulty in defining the appropriate data quality metric.

There is an evident need to incorporate data quality considerations into the whole data cycles encompassing managerial/governance as well as technical aspects. Currently it can be observed that contributions from research and industry into data quality has been from three distinct contributing communities:

BA, focus on *organizational* solutions. That is, the development of data quality objectives for the organization, as well as the development of strategies to establishing different roles, processes, policies, and standards required to manage and ensure that the data quality objectives are met.

Solution Architect, work on *architectural* solutions. That is, the technology landscapes required to deploy developed data quality managements processes, standards and policies.

Database Experts and statisticians, contribute to *computational* solutions. That is, effective and efficient IT tools, and computational technique, required to meet data quality objectives. Technique in this regard can include record linkage, lineage and provenance, data uncertainties, semantic integrity constraint, as well as information trust and credibility.

This paper presents the methodologies employed for the study, are followed by a discussion of the key results. We also provide a taxonomy of data quality problem that emerges from the analysis of current research, and identify the top requirements, outputs and main development in data quality research.

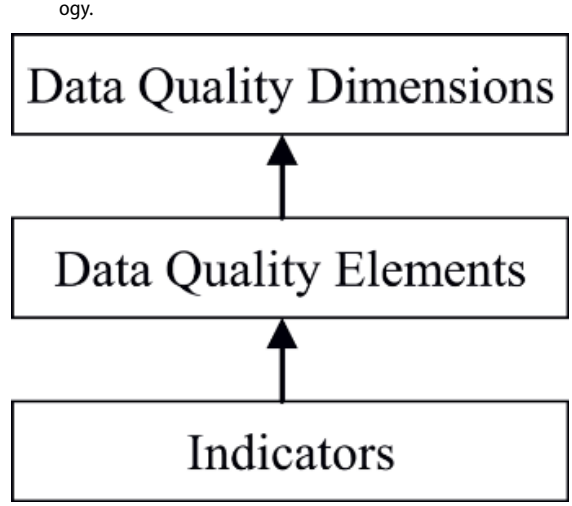
METHEDOLOGY

Our study follows a concept analysis way, in which the challenges are been identified and a model is been repaired to overcome these challenges.

- Database Process Quality
- Database Data Quality
- Database Semantic Quality
- Database Behavior Quality

CHALLENGES

1. The diversity of data source brings up abundant data types and complex data structures and increases the difficulty in data integration.
2. Data Volume is tremendous and it is difficult to judge data quality within a reasonable amount of time.
3. Data change very fast and the "timeliness" of data is very short, which necessitates higher requirements for processing technology.



Dimensions	Elements	Indicators	
1) Availability	1) Convenience	■	Interface for data access is provided or not
		■	Data can be provided easily to general public
	2) Correctness	■	Whether data arrived on time
		■	Data is updated regularly
		■	Time frame from data collection to processing and to release should match according to all the requirements
2) Usability	1) Integrity	■	Data is received from various organizations of a country, or industry
		■	Regular auditing by experts of the correctness of data
		■	Data should exist in acceptable and known values
3) Reliability	1) Precision	■	Accuracy of data
		■	True state of source information should be reflected in data representation
		■	No ambiguity in data representation
	2) Steadiness	■	Concept, value domain, format should match before and after data processing
		■	Data remains consistent and verifiable in given period of time
		■	Consistent and verifiable data from all the data sources
	3) Reliability	■	Format of data is clear and should meet criteria
		■	Consistent data with structural integrity
		■	Consistent data with content integrity
	4) Totality	■	While using data with multi-components, the deficiency will impact use of the data
		■	While using data with multi-components, the deficiency will impact use of the data integrity and accuracy
4)Relevance	1) Aptness	■	Data collected do not completely match requirements but they explain on aspect of it
		■	Datasets retrieved are according and relevant to users need
		■	Information provided matches with users' retrieval requirements
5) Presentation Quality	1) Legible	■	Clear and understandable data
		■	Easy to judge that the data provided meet needs
		■	Everything related to data is easy to understand

CONCLUSION

Data quality is and always an important factor in the success of data warehousing projects. We have identified potential source of error causing quality compromises and presented in the paper. Meta data based quality model is proposed to enforce quality in the data warehouses. While evaluating the qualities value of an object/component, if it fails to meet the specified qualities level, then how to improve upon the quality of that component is also provided. It is not at all easy to quantify some abstract entity. Suitable metric to quantify identified parameters is also provided in this paper.

While setting up a data warehouse projects in an organization, there are various stakeholders and everyone has different perspective for data. And accordingly the importance of the quality concept varies. While the designing and implementing DW, it is really important to understand and incorporate the expectation of all the stakeholders from the DW.

REFERENCES

1. Ballou, D. and H. Pazer,"Designing information systems to optimize the accuracy timeliness tradeoff", *Information Systems Research*, Vol. 6, No. 1, 1995, pp. 51-72.
2. Chignell, M. and P. Kamran, Intelligent Database Tools and Applications, Wiley, Los Angeles, California, 1993.
3. Costin, H., Total Quality Management, Dryden, United States, 1994.
4. Delen, G., and D. Rijsenbrij, "A specification, engineering and measurement of information systems qulisty", Journal of Systems Software, 1992, Vol. 17, No. 3, pp. 205-217.
5. Dvir, R. and Evans, S., "A TQM approach to the improvement of information quality", http://wem.mit.edu/tdqm/papers, accessed 7/97.
6. Dyer, M., The Cleanroom Approach to Quality Software Development, Wiley, 1992.