

'Accounting' for data quality in enterprise systems

O'BRIEN, Anthony

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/13319/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

O'BRIEN, Anthony (2015). 'Accounting' for data quality in enterprise systems. *Procedia Computer Science*, 64, 442-449.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>



Conference on ENTERprise Information Systems / International Conference on Project
MANAGEMENT / Conference on Health and Social Care Information Systems and Technologies,
CENTERIS / ProjMAN / HCist 2015 October 7-9, 2015

‘Accounting’ for data quality in enterprise systems

Tony O’Brien*

Sheffield Business School, Sheffield Hallam University, Howard Street, Sheffield S1 1WB, UK (a.o'brien@shu.ac.uk)

Abstract

Organisations are facing ever more diverse challenges in managing their enterprise systems as emerging technologies bring both added complexities as well as opportunities to the way they conduct their business. Underpinning this ever-increasing volatility is the importance of having quality data to provide information to make those important enterprise-wide decisions. Numerous studies suggest that many organisations are not paying enough attention to their data and that a major cause of this is their failure to measure its quality and value and/or evaluate the costs of having poor data. This study proposes an integrated framework that organisations can adopt as part of their financial and management control processes to provide a mechanism for quantifying data problems, costing potential solutions and monitoring the on-going costs and benefits, to assist them in improving and then sustaining the quality of their data.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge

Keywords: Data Quality; Enterprise Systems; Accounting Information Systems; ERP; SCM; CRM; Big Data

1. Introduction

Over the past three decades in particular, organisations have invested vast resources into the development and maintenance of enterprise-wide information system applications (ERP, SCM, CRM etc.). In more recent times this has taken place within the context of an ever-increasing volatile environment which has witnessed the emergence and advancement of new technologies involving, the expansion of the internet and the growth of E-Business and E-Commerce; the explosion in social media; the greater integration of processes and systems within and between

* Corresponding author. Tel.:

E-mail address: tony.obrien@hotmail.com

businesses; the proliferation of mobile computing including BYOD (bring your own devices); the emergence of Cloud Computing allied to software as a service (SaaS), and the evolution of Big Data, As a consequence organisations are now tied irrevocably to information technologies (IT) and information systems (IS), as never before not just to create competitive advantages but merely to survive.

It is within the context of this ever-changing environment of greater and greater organisational complexity, that the importance of data quality and the quality of information, which all organisations use to make the important decision at all levels, has become even more paramount. Within the last two decades data quality has been identified as a major concern for many enterprises^{1,2,3,4,5,6} none more so than those operating enterprise resource planning and information systems.⁷

In the intervening years however, there does not appear to have been any real improvement. English⁸ outlined a catalogue of corporate disasters emanating from poor quality business information amounting to 'One and a Quarter Trillion Dollars'. A survey sponsored by Pitney Bowes, reported that a third of the respondents rated their data quality as poor at best and only 4% reported it as excellent⁹ whilst a further survey found that less than one third of organisations regularly monitor data quality.¹⁰ More recently a Gartner report predicts that by 2017, 33% of Fortune 100 organisations will experience an information crisis, due to their inability to effectively value, govern and trust their enterprise information.¹¹ In addition an Experian report published in January 2015 identified that global companies feel that 26% of their data is inaccurate (32% in the US), up 25% from the previous year and that almost 80% of organisations do not have a sophisticated approach to data quality.¹²

From this there is evidence to suggest that many organisations are unaware of the extent of the data quality problems which exist within their enterprises or either choose to ignore, or do not prioritise such issues. A major underlying element of this apparent 'indifference' is that many organisations fail to value either the quality of the data they hold, or the cost of having poor and inaccurate data. If an organisation is not able to evaluate the quality of its data how can it determine its value in relation to the corporate decision making process? A number of studies have attempted to develop forms of cost classification models^{3,4,13,14,15,16,17} and whilst these have developed focussed taxonomies on the related major elements, they may be perceived to be somewhat generic.

2. Research Approach

This on-going investigation is attempting to build upon the work of these studies and to develop a specific cost/benefit framework to enable *individual* organisations to: a) analyse the costs of low quality data (consequential costs); b) determine the costs of improving/assuring data quality (investment costs) and c) evaluate 'other' benefits of having quality data. The intended outcomes are to provide mechanisms to: d) identify and analyse the data quality issues; e) build a strong business case to promote improvements, where applicable; f) implement improvement processes; g) establish the on-going monitoring of the quality of the data. It is intended that the outcomes of this study will provide organisations with the opportunity to build this framework within their procedures and systems, both operational and financial, so that the processes will become a permanent integrated management and financial control mechanism to add real value, rather than an occasional one-off ad hoc 'data clean up' exercise. In this way the organisation is able to take real 'ownership' of its data.

In this paper one has scoped the problem and based the discussion on reviewing relevant literature, feedback from a related case study, together with one's own experiences from having worked with major organisations related to the quality of organisational data, from which the proposed framework summarised above has been developed. The intention is to conduct a research investigation with a number of Small and Medium-Sized Enterprises (SMEs) to test and refine the proposed framework.

3. Related Work

This section discusses major elements of the literature relating to the classification of the impact of poor data quality together with the costs of attempting to improve and assure the on-going quality of the data. English³ divides data quality costs into three broad categories, costs caused by low data quality, assessment and inspection costs incurred to verify if processes are performing correctly and costs resulting from activities to improve the quality of

data. Loshin⁴ focusses upon grouping costs according to their organisational impact relating to, operational, typically short-term process issues; tactical, system problems essentially of a medium term nature; strategic, addressing decision which address the long term future. Loshin⁴ further identifies several sub-categories relating to detection, correction and prevention costs.

Following on from the themes of English³ and Loshin⁴, Eppler and Helfert¹³ proposed a model which dissects data quality costs into two major classifications relating to those costs incurred as a result of low quality data and the consequential costs of improving or assuring ongoing data quality. Each classification then consists of subordinate categories relating to the direct and indirect costs of poor data and the prevention, detention and repair costs associated with data quality improvement processes as shown in Table 1. Each subordinate category is then further subdivided into six quality costs elements and seven cost improvement elements.

Table 1: “A data quality cost taxonomy”. Source ¹³ Adapted by ¹⁶

Data quality costs	Costs caused by low data quality	Direct costs	Verification costs
			Re-entry costs
			Compensation costs
		Indirect costs	Costs based on lower reputation
			Costs based on wrong decisions or actions
			Sunk investment costs
	Costs of improving or assuring data quality	Prevention costs	Training costs
			Monitoring costs
			Standard development and deployment costs
		Detection costs	Analysis costs
			Reporting costs
		Repair costs	Repair planning costs
			Repair implementation costs

The Eppler and Helfert¹³ study differentiated between direct and indirect costs. In attempting to estimate the costs of poor data quality the question of devising a credible evaluation of the costs is very important otherwise arguments can develop around the basis of the evaluation rather than the ‘message’ that is being generated. Certain of the cost elements in Table 1, in particular direct costs and improvement costs, may be visible to the organisation and therefore capable of attracting some form of valuation. However indirect costs may not be so visible and the actual consequences may be intangible. Therefore a further differentiation is necessary between direct costs (those that can be valued) and hidden costs which require some form of ‘estimate’ of their effect, but may not be capable of a monetary evaluation. Following on Ge and Helfert¹⁴ identified three major components relating to this area: (1) information quality assessment, (2) information quality management, and (3) contextual information quality. In a subsequent study Ge and Helfert¹⁵ showed that the improvement of data quality in the intrinsic category (e.g. accuracy) and the contextual category (e.g. completeness) can enhance decision quality.

Haug, Zachariassen, and van Liempd¹⁶ in Table 2 provide examples of various types of costs, direct (tangible) and hidden (intangible) from both an organisational and strategic perspective.

Table 2: Types of Data Quality Costs

Hidden costs	E.g. long lead times, data being registered multiple times, employee dissatisfaction, etc.	E.g. focus on wrong customer segments, poor overall production planning, poor price policies, etc.
Direct costs	E.g. manufacturing errors, wrong deliveries, payment errors, etc.	E.g. few sales, low efficiency, problems in keeping delivery times, etc.
	Effects of poor quality data on operational tasks	Effects of poor quality data on strategic decisions

Data can be seen as an important organisational asset as well as a resource. Its quality is directly related to business value and organisational performance. In addition to measuring the effect on business processes, organisational performance has always been of consideration to IS/IT researchers and practitioners, resulting in a plethora of performance related contributions.¹⁷

4. Related Case Study

This article has provided illustrations from the literature to highlight examples of the costs of poor data quality and potential benefits of related improvement programmes. A further example of the effects of such an initiative may be seen from a recent study conducted with a large quasi-public sector organisation which has again highlighted the impacts of poor data quality.¹⁷ The organisation faced numerous problems relating to data quality whilst providing its services. The study conducted in the form of focus groups, highlighted a number of key themes relating to data quality.

The main themes identified are as follows. Firstly, in the discussion among the cross section of the workforce, it was noted that data and information governance were of low priority. Employees' awareness of data governance issues and the associated responsibilities were low; the communication channels used to highlight and promote data quality issues were either non-existent or clogged. Secondly, there was an absence of any formal mechanism or a procedure to report data problems. However one of the positive aspects of the discussion was that the senior management were aware of the data quality issues and the pressures of compliance and were highly supportive in improving the current practices and procedures, but the existing organisational culture and the remains of its public sector heritage made their task harder and less effective.

Each of the six focus groups, comprising practitioners from a similar function or department, was asked to undertake separate individual projects investigating areas of actual/potential information risk which were within the groups' sphere of influence. This work was then developed into six mini case studies from which the following risks and issues were identified and summarised below:

1. The introduction of a new Master Data Management (MDM) system
2. Migration of a major company-wide communications system to a new provider
3. Data quality issues within a major company-wide application
4. Issues with a core data system

- 5. Master data issues
- 6. Mismatch between transactional and master data

Each of the identified risks was summarised (numbered 1-6) and sub-divided further into their more detailed elements and identified (lettered a-d) as appropriate.

Whilst it is not possible to measure the above risks and issues in strict monetary terms, an evaluation matrix has been developed based upon a) the level of risk (high, medium, low) and b) the related organisational decision making level (strategic, tactical, operational). Each of the sub-risks (analysed by major risk 1-6 and detailed risk a-d) was then evaluated as to its potential risk level (high, medium, low) and to which organisational level it related (strategic, tactical, and operational). This evaluation is detailed in Table 3 below

Table 3: Decision/Risk Level Matrix (each Case Study and sub-element)

Decisions/ Risk Level	High	Medium	Low
Strategic	1b New MDM system 5a Regular service interruption causing public embarrassment	2a Data quality problems 2b Migration issues 3b Lack of data ownership 4a Data accuracy in doubt 4c Information quality issues 4d Lack of data ownership 5b Master data issues 5c Multiple systems and data formats 5d process changes not managed	
Tactical	1b Data download failure	1b Customer dissatisfaction 2a Increased costs 2a Lost revenues 3c Users losing confidence 4b Procedures not always followed 6a Data inconsistencies- manual intervention 6b Asset records inaccurate	1a Existing data problems
Operational	1b Non-availability of system	1a Existing data problems 2a Manual intervention 2a Integration issues 3a Data inconsistencies 3a Manual corrections	1a Existing data problems

A further outcome from this case study was the appreciation that the Decision/Risk Level Matrix model above has the potential to be applied in other instances where some form of evaluation of a risk, issue or benefit is required which cannot be measured strictly in monetary terms.

5. Development of a Data Quality Cost/Benefit Framework

The initial research objectives are to develop a framework to enable organisations to: a) analyse the costs of low quality data (consequential costs); b) determine the costs of improving/assuring data quality (investment costs) and c) evaluate ‘other’ benefits of having quality data and thereby provide them with mechanisms to: d) identify and analyse the data quality issues; e) build a strong business case to promote improvements, where applicable; f) implement improvement processes; g) establish the on-going monitoring of the quality of the data

Following on from the review of the relevant literature, the outcomes from the above case study, together with one’s own experiences from having worked with major organisations related to the quality of organisational data, an

initial conceptual framework has been developed which attempts to embrace the requirements and outcomes of the research objectives as shown in the Initial Data Quality Evaluation Framework Figure 1 below.

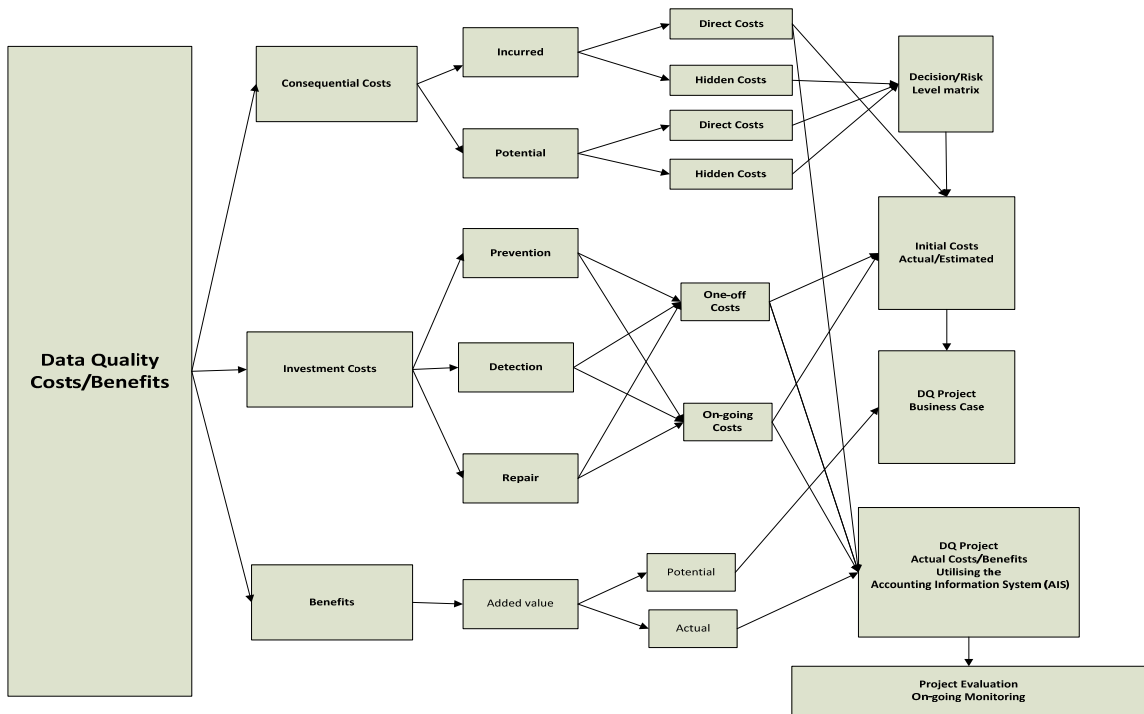


Figure 1: Initial Data Quality Evaluation Framework

5.1. Research Project Outline

In relation to the Initial Data Quality Evaluation Framework: Figure 1 above.

The costs and benefits are broken down into three sections:

- Consequential costs- the costs resulting from having low quality data, analysed between:
 - Costs already incurred
 - Potential costs, which could accrue in the future
 - The above costs are then analysed further into:
 - Direct costs which can be quantified (incurred) or estimated (Potential)
 - Hidden costs which are intangible and difficult to estimate in monetary terms

- Investment costs- the actual costs relating to the process of improving the data, analysed by
 - Detection costs
 - Repair costs
 - Prevention costs
 - The above costs are then analysed further into:
 - One-off costs as part of the initial programme
 - On-going costs which will be incurred into the future in order to sustain the programme

- Benefits, those outcomes of the improvement programme which add real value to the business as against reducing costs, analysed between those that have:
 - Potential to occur
 - Those that actually occur in the future which require to be captured

The essence of the project is to provide organisations with the ability to capture, analyse and evaluate all of the above consequences and outcomes as effectively as possible, initially to evaluate the size of the problem; to prepare a business case for an improvement programme if this is applicable; monitor the actual initial improvement process if this is implemented and then to monitor subsequent events into the future utilising the organisation's existing Accounting Information System (AIS) to determine whether progress if any is being made on an on-going basis. It is suggested that a number of analytical tools be employed to analyse each of the components described above, which can be linked together to provide an integrated evaluation process. Whilst it is appreciated that this is a 'working concept' at this time the process provides an initial robust framework on which to base the initial research.

5.2. The Evaluation and Monitoring Process

The process of the cost/benefit evaluation, the monitoring and the ultimate overall project evaluation corresponds to the *right hand portion* of The Initial Data Quality Evaluation Framework: Figure 1 above. The Decision/Risk Level Matrix detailed in Section 4, Table 4 above provides a format to analyse, evaluate and prioritise the 'hidden' intangible consequential costs and risks (*box 1*). The actual direct consequential costs together with the estimated one-off and on-going improvement (investment) costs can be analysed more easily within some form of database/spreadsheet (*box 2*). The outcomes of these two analyses, together with any estimated potential additional value added benefits can be integrated to form the basis of evaluating and subsequently building a valid business case to initiate improvements (*box 3*). Whilst the 'hidden' costs may not be determined strictly in monetary terms, the matrix can provide a means of evaluating the potential risks, their impact and chances of occurrence which can influence the overall business case decision.

The monitoring of costs and benefits is essential if an organisation is to manage and control any form of project or programme. Failure to do so is a common source of project failure¹⁸. It is suggested that an organisation can utilise the analysis and reporting features of its Accounting Information System (AIS) to identify those direct consequential and investment costs and tangible benefits over periods of time. Within a typical AIS the 'general ledger' 'collects' and analyses all types of transactions (costs, revenues, income, assets and liabilities) by way of the 'chart of accounts' and is also able to relate the transactions to a specific business, factory, department, function, location, employee etc. by a designated 'cost or profit centre' (*box 4*). Modern systems have additional features by which transactions can be analysed usually in the form of 'dimensions'. It is suggested that a specific 'dimension' be set up and allocated to each transaction relating to the data quality project whether consequential and investment costs or added value benefits. In this way all actual transactions relating to the data quality programme can be identified by the designated dimension code and subsequently analysed by type of transaction (cost/benefit) and by location (factory, department etc.) via the AIS reporting structure.

The outcomes of the business case will provide projections, forecasts, targets, milestones over time, against which the organisation can measure the project's actual on-going performance from the AIS general ledger reporting as the essential part of the Project Evaluation (*box 5*). The intention is that the above framework will be built into an organisation's procedures and systems, both operational and financial so that the processes will become permanent business activities rather than occasional one off ad-hoc exercises.

5.3. Further Research

It is contended that this project has real potential to make considerable progress towards achieving the initial research objectives as detailed at the beginning of this section. Further research is needed and to this aim the intention is to conduct a research investigation with a number of SMEs to test and refine the proposed framework. At this stage SMEs are considered to be the most appropriate type of organisation to approach as they appear to be more accessible and provide a wider scope for cooperation than larger organisations.¹⁹

6. Concluding Remarks

This study does not purport to provide solutions as to how organisations may improve the quality of their data or to implement changes to sustain such improvements. Rather it argues that practical improvement programmes and real process change, cannot take place successfully without some form of evaluation and monitoring to establish, 'where one is starting from', 'where one wants to go' and 'where one is now' within the overall process. This study therefore attempts to assist organisations in making that journey, thereby taking real 'ownership' of its data, by providing a mechanism for quantifying data problems, costing potential solutions and monitoring costs and benefits via an integrated management and financial control mechanism to add real value to its operations.

References

1. Redman, T. C. (1995) Improve Data Quality for Competitive Advantage. *Sloan Management Review*, Winter: 99-107.
2. Redman, T. C. (1998) The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, 41(2): 79-82.
3. English, L. P. (1999) *Improving Data Warehouse and Business Information Quality- Methods for Reducing Costs and Increasing Profits*. New York: Wiley Computer Publishing: 518.
4. Loshin, D. (2001) The Cost of Poor Data Quality. *DM Review Magazine*, June.
5. Eckerson, W. (2002) *Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data: The Data Warehouse Institute*: 1-33.
6. Redman, T. C. (2004) Data: An Unfolding Quality Disaster. *DM Review Magazine*, August: 1-6
7. Deloitte Consulting LLP. (1999) *ERP's Second Wave- Maximizing the Value of Enterprise Applications and Processes*: 1-45.
8. English, L. P. (2009) *Information Quality Applied*. Indianapolis: Wiley Publications Inc: 802.
9. Information Difference (2009) *The State of Data Quality Today*: 33: The Information Difference Company.
<http://www.pbinsight.com/files/resource-library/resource-files/The-State-of-Data-Quality-Today.pdf>
10. Hayter, A. (2010) *The State of Data Governance*. Paper presented at the Data Governance Conference Europe 2010, London.
11. Gartner (2014) Gartner press release Egham UK <http://www.gartner.com/newsroom/id/2672515>
12. Experian (2015) *The data quality benchmark report*- <http://go.experian.com/global-research-report-2015>
13. Eppler M, Helfert M (2004) A classification and analysis of data quality costs. 9th MIT International Conference on Information Quality, November 5-6, 2004, Boston, USA
14. Ge, M. and Helfert, M. (2007) *A Review of Information Quality Research - Develop a Research Agenda*. International Conference on Information Quality, November 9-11, 2007, Cambridge, Massachusetts, USA
15. Ge, M. and Helfert, M. (2008), Effects of information quality on inventory management. *International Journal of Information Quality*, 2(2), pp. 176-191.
16. Haug, A., Zachariassen, F., & van Liempd, D. (2011). The cost of poor data quality. *Journal of Industrial Engineering and Management*, 4(2), 168-193.
17. O'Brien, T, Sukumar, A and Helfert, (2013) *The Value of Good Data- A Quality Perspective*. The International Conference of Enterprise Information Systems, Angers, France
18. Meredith, R. and Mantel, S. (2002) *Project Management- A managerial approach*. Wiley Text Books
19. ERC (2015) *Collaboration between SMEs and universities- Report to HEFCE by The Enterprise Research Centre (ERC)*, May