

Data Quality – Trusted Data Across the Enterprise

By Martin Spratt

Contents

- Executive summary IX
- About the Author XI
- Acknowledgements XIII

- Part 1: Approaching Data Quality for your Enterprise**

- Chapter 1: Trends driving data quality stress points 3**
 - Data quality – the perfect storm 3

- Chapter 2: Revealing the data quality issue 9**
 - How is poor data quality first recognised or identified? 9
 - Why is data quality important? 10
 - Moral obligation for accuracy 10
 - Core business instrumentation 10
 - Accuracy in recording and reporting data 10
 - Reliability and fairness 11
 - Accountability 11
 - Probity 11
 - Performance 11
 - Social demographics – the social cost of poor data quality 11

- Chapter 3: Data governance and data quality 13**
 - The politics of data 13
 - How can data be governed? 13
 - Defining data ownership and responsibility of the data quality issue 14

- Chapter 4: Data quality remediation concepts and challenges 19**
 - Solving the right problem 19
 - Data integrity (is it the right data?) versus data quality (is the data right?) 19
 - Conflict – apathy versus pragmatism 19
 - Identifying and scoping the problem 20
 - Measuring the problem – how bad is it? 20
 - Assigning responsibilities 20
 - Data ownership (business) versus data stewardship (ICT) 21
 - Winning funding and sponsorship for data quality 21

- Chapter 5: Data-quality-methodology best practices 23**
 - Data trustee – who governs the data? 23
 - Problem recognition – is there a problem with the data? 23
 - Root cause identification – where is the problem originating? 23

Data quality measurement – how bad is the data quality problem?	23
Data quality remediation	24
Data quality – continual improvement	24
Remediation approaches	25
Recommendations summary	26
Chapter 6: Quality data – practical approaches	29
The five-stage approach to quality data (DataFlux)	29
Conclusion	32
Chapter 7: Ten factors for enterprise-wide data quality	33
Lessons learnt by the experts: does your data support successful implementations?	33
10 critical factors for successful enterprise-wide data quality	33
Factor one: establish measurable business goals	34
Factor two: align business and IT expectations	34
Factor three: confirm senior management buy-in	35
Factor four: ensure that business goals drive functionality	36
Factor five: understand the costs of building a solution in-house	37
Factor six: commit trained personnel	37
Factor seven: understand the real costs and causes of poor data quality	38
Factor eight: employ a proven methodology	39
Factor nine: use a phased roll-out schedule	40
Factor 10: tracking ROI	40
Conclusion	41
Chapter 8: Project manager’s guide to data quality	43
Phase one: project preparation	43
Phase two: making the blueprint	46
Phase three: implementation	49
Phase four: rollout preparation	51
Phase five: going live	53
SWAT team	53
Phase six: maintenance	54
Chapter 9: Data quality in BI and performance management	59
Data quality is central to BI initiatives	59
Data quality in data warehouse and BI	59
Data quality: the first metric for BI and business performance management (BPM) success	60
The key dimensions of data quality	60
A business-focused approach to PM and data quality	61
Who owns data quality – business or IT?	62
Better quality data for better performance	62

Chapter 10: Top tips for customer address data quality	65
Tip one: start at the end	65
Tip two: consider the data elements	65
Tip three: measure data quality	65
Tip four: how to get from here to there	65
Tip five: how to secure buy-in	66
Tip six: how to win support for the investment	66
Tip seven: put effective processes in place	66
Tip eight: use technology	66
Tip nine: measure improvement	67
Tip 10: institutionalise (and start again)	67
Chapter 11: Data quality and EDM	69
Enterprise application integration (EAI)	70
Extract, transform and load (ETL) tools	71
Master data management (MDM)	72
Enterprise data management (EDM) versus MDM	72
Service-oriented architecture (SOA)	73
EDM in summary	74
Chapter 12: Data quality and MDM	75
Evaluating MDM solutions	75
The MDM advantage	76
Other data quality benefits from a MDM solution	77
Data quality baseline to ongoing MDM	78
Profiling reduces MDM migration risks	78
Part 2: Case Studies – Industry-Specific Challenges	
Case study 1: Banking and finance	81
Banking and data quality	81
Basel II and data quality implications	81
Australian Prudential Regulatory Authority (APRA) calls for data quality improvements	82
Non-banking finance sector and data quality	82
APRA letter to the banking industry regarding data quality	83
Deutsche Bourse: sharable and trusted data	90
UMB Bank: the remediation of failing CRM projects with better data	91
ING Americas: high-quality data reduces costs	94
Standard & Poor's: financial system reference data	95
Marks & Spencer (M&S) Money: Basel II data quality initiative	98
Banco Popular: high-quality customer information system (CIS)	102
HSBC Bank Canada: advanced data cleansing delivers product insight	103

Case study 2: Healthcare	107
Making patient data flexible, reusable and productive	107
Clalit Health Services: integrated patient care data.....	109
The University of Texas MD Anderson Cancer Center.....	109
Sutter Health: Enterprise Master Patient Index (EMPI).....	114
Leukaemia Foundation: cleaning up patient address data	116
NSW Nurses’ Association: improving membership data quality.....	117
New South Wales Cancer Council: accurate patient identification and follow-up care.....	119
Case study 3: Retail and channel sales	121
An in-depth, holistic view of retail data – wherever it resides	121
ACE Hardware: high-quality customer insight.....	123
Carphone Warehouse: real-time view of accurate sales	123
Choice Hotels International: customer data quality using MDM.....	124
Rochford Wines: data quality saves over \$10,000 in one mailing.....	126
Cendant Hotel Group: clean loyalty data in 90 days	127
Microsoft: improving channel management with accurate data.....	129
EMI Music Publishing: data quality improves copyright compliance.....	131
Case study 4: Pharmaceutical industry – culture change in life science regulatory compliance	133
Data quality – part of a larger compliance culture change.....	133
Case study 5: Utilities and energy	137
Optimising asset management with quality metrics	137
Ameren Corporation: high-quality single customer view.....	140
Southern Company – rejuvenating legacy data	142
British Gas, AA and Centrica – merging customer data accurately.....	143
Case study 6: Government, defence and education	147
Accurate visibility across department silos	147
Joondalup City Council – cleaning up customer data quality.....	148
Inland Revenue, UK: centrally-managed accurate customer data.....	150
Defence Acquisition University (DAU): quality student faculty and finance data.....	152
Ministry of Defence: cutting costs by £20m in dirty data clean-up.....	152
Department of Industry, Tourism and Resources (DITR) – cleans up web data	154
South African Revenue Service: cleaning data accurately identifies citizens	156
Insurance: Prudential UK – improving call centre effectiveness with CRM data quality.....	156
Case study 7: Telecommunications	161
Convergence, consolidation and competition.....	161
Telemar: improving customer loyalty with accurate customer data.....	163
XO Communications: monitoring traffic flow.....	163
Eircom Europe: high-quality customer directory data	164

The Carphone Warehouse: data quality rescues CRM initiative.....	165
BT Group: integrating customer view across the enterprise.....	167
Dutch Yellow Pages: customer satisfaction up with accurate data.....	169
Case study 8: Law enforcement	173
Humberside Police: clean and accurate crime-fighting data	173
Case study 9: Transport and logistics	177
Stale, disjointed data reduces profit, performance and compliance.....	177
US Airways: data quality efforts improve safety and maintenance.....	179
Burlington Northern Santa Fe (BNSF) Railway: quality data drives profit model	179
FedEx: accurate, web-based tracking data kept clean	180
Washington State Department of Transportation (WSDOT)	182
Case study 10: Manufacturing.....	183
Porsche: rapid customer data quality for CRM and marketing.....	183
3M: cleaning trading partner database.....	184
Appendix: Who’s who.....	187
Index	193

Executive summary

“Fast is fine, but accuracy is everything.”

Wyatt Earp

DATA QUALITY is fast becoming the Achilles’ heel for contemporary computer systems. Time and economic pressures are forcing organisations into faster transaction speed and richer computerised relationship interactions, and by default, data collection volumes are escalating at breakneck speed as vast arrays of complex data assets fail to be well managed.

Added to the speed and volume of computerised data collection and data management systems, is the modern mantra of ‘agility’, an industry buzzword for rapid change. Global commerce is universally pursuing a vision to create computer systems that quickly and efficiently accommodate change, even at the ‘speed of thought’ as proposed long ago by visionaries like Bill Gates.

Agility is fast becoming a reality supported by commodity, virtualisation computing capabilities in hardware and software, and emerging architectural approaches like Services Oriented Architecture (SOA). The result is a journey towards a fast, fluid computerised environment that ironically is failing to accurately capture and recall information. The focus on agility in contrast to immature data quality disciplines is resulting in systems where the data cannot be trusted.

Compliance at all levels of government and industry is driving renewed scrutiny and vigour into data management systems and data quality in particular. Increased disciplines concerning accountability and

the quality of raw data are what compliance efforts are based upon.

The need to address data quality is not only mandatory in commerce, it is becoming painfully acute in the wake of the recent 2007 – 2008 US sub-prime mortgage crisis, which demonstrates the toxic combination of poor governance and poor fundamental data quality.

This report explores the costs and penalties associated with poor data quality, and reviews remediation methodologies, best practices and leading technologies to help restore confidence in the most basic building block of computer systems – the data.

Data quality is ubiquitous. It has emerged clearly as an issue wherever data is present; therefore data quality participates as a consideration in every computer application, as well as every major information system’s theme; such as business intelligence (BI), enterprise resource planning (ERP), customer relationship management (CRM), master data management (MDM), service oriented architecture (SOA) and security.

Time and space won’t allow the deep dive into the relationship and application of data quality within all these areas, so the focus in this report is on several marquee technology themes at present as they intersect with data quality issues. These are:

- Service oriented architecture;
- Master data management and its cousin – customer data integration (CDI);
- Business intelligence and performance management; and
- Compliance efforts at all levels requiring accurate, trusted data.

Within data quality, as a specialty, we also see specific sub-disciplines at work. These sub-disciplines follow generally accepted approaches to the management and implementation of data quality disciplines. These fall into the following generic categories:

- Data governance and data ownership – who owns the data, and who is best able to know if the data is wrong, and knows what rules/logic to apply to repair the data;
- Assessment and profiling – examining the *status quo* to identify core data quality issues;
- Matching and cleansing – the process of cleaning the data;
- Enrichment – optionally adding additional data (external or otherwise);
- Monitoring and improvement – the ongoing process of monitoring and improving the overall data quality of systems.

Across our research, a few common themes rang true from customers and vendors alike as various practitioners developed successful approaches to data quality. Some of the common themes to emerge were:

- Data quality knowledge – the business personnel, rather than technical IT personnel, are in the best position to rate and understand the semantic quality of data;
- Master data – data should ideally reside in one main, core, or central location and moved as infrequently as possible;
- Ban ungoverned copying – data should ideally be referenced from one main, core, or central location, rather than being randomly copied throughout the organisation;

- Assess the cost of poor quality – do an assessment of the costs and impacts of using poor quality data;
- Getting business buy-in – business needs to drive, own and manage data quality initiatives if they are to stick;
- Use technology – automate as much of the data quality workload as possible;
- Institutionalise data quality – data will degenerate over time so data quality must be an embedded discipline. Embed data quality as a culture, measurement tool and as an improvement tool on an ongoing basis, rather than a one-off effort.

Leveraging commentary from global luminaries and case studies on the subject of data quality, this report aims to assist readers on their journey to measure and improve their own data quality initiatives, and restore confidence and trust in their data. It does this through:

- Raising critical awareness of the cost of poor data quality;
- Identifying key methods and disciplines to drive and measure data quality improvements;
- Highlighting technologies and vendors in the market with expert focus on data quality; and
- Reviewing case studies that showcase the benefits of improved data quality.

About the Author

MARTIN SPRATT is a veteran data specialist with 27 years international experience in data intensive projects and technologies. Working in a variety of jurisdictions, he has applied his mastery of deep data management disciplines to a broad range of business problems in the airline, banking, insurance, telecommunications and heavy manufacturing sectors, with household corporate names like Oracle, IBM, Platinum Technology, Candle Corporation, Bell South, Bell Canada, State Street Bank, John Deere, Caterpillar, Rockwell, Qantas, Westpac, Norwich Union Insurance, Royal Bank of Scotland, Telstra, Transurban and Mitsubishi Motors, to name just a few.

Martin's career highlights include: conducting engineering due diligence for IBM acquisitions in its Laboratory Research Community, such as Unicorn (Metadata), SRD (Entity Analytics), Venetica (Unstructured Data Federation), DWL (WebSphere Customer Centre), Ascential (ETL, Data Quality) and CrossAccess (MVS Mainframe Data Access); pioneering product design and deployment work on IBM's Information Integration technology; undertaking global competitive intelligence work across IBM's Information Integration portfolio including competitors like Informatica, Composite Software, Data Mirror, Siperian and many others; carrying out design teamwork on IBM's Database Migration Toolkit (MTK) and working on joint engineering projects with global IBM partners such as Unicorn, CrossAccess, Microstrategy, Business Objects, Initiate Systems and many more.

Based in Melbourne, Australia, Martin advises companies in several key data intensive areas including fraud detection, anti-money laundering (AML), counter-terrorism financing (CTF), data quality, data governance, and high speed real-time business performance measurement systems with a view to improving corporate compliance initiatives such as APRA Data Quality, AUSTRAC Reporting and BASEL II efforts, as well as SOA Data Services delivery as part of the Anatas SOA Competency practice.

Martin also chairs the Australian chapter of the EDMCouncil.org, which is an executive peer network of the world's largest data users in the finance sector, coaching organisations to manage data as a valuable corporate asset.

Martin can be contacted at martinspratt@gmail.com

Acknowledgements

MAJOR CONTRIBUTORS and supporters of this effort were many. First, thanks to Ark Group for investing into the research effort on the subject of data quality in the local marketplace, and for sponsoring this project to drive greater awareness, education and excellence to help corporations and governments improve data quality.

Special thanks is owed to local management teams and individuals from several data integration and data quality technology organisations who contributed comprehensive data quality materials and project case studies of great interest.

Special thanks go to:

- Identity Systems – Michael Dunkerley, vice president of global marketing;
- IBM – Adrian Gaule, Information Integration Solutions; Harald Smith, product manager information management; Bob Zureck, director of advanced technologies and Mei Selvage SOA data architect;
- Informatica – Laurie Newman, country manager; Malcolm Pooley, Australian NZ data integration manager; Dominic Micic, regional data integration and data quality trainer and Neil Gow, Asia Pacific data quality and integration manager;
- Initiate Systems – Alex Paris, Australian country manager and Piers Wilson CDI and MDM data quality specialist;
- QAS – Frank McKenna, Senior Product Manager APAC for QAS;
- SAS – Jillian MacMurchy, data integration solution manager;
- DataFlux – Tony Fisher, president and general manager;
- Standard & Poors, Australia – Rory Manchee, managing director;
- Trillium Software – Caroline Lim-Brown director, Asia Pacific and Leonard A. Dubois VP, marketing and support;
- Veda Solutions Group – Ian Davies, Australian product manager.

Martin Spratt, May 2008.

