Monitoring Information Quality: With Applications for Traffic Management Systems

Markus Helfert[1], Fakir Hossain[1] and Owen Foley[2]

[1]School of Computing, Dublin City University, Glasnevin, Ireland

[2]Galway Mayo Institute of Technology, Dublin Road, Galway, Ireland

Abstract. Information quality (IQ) plays a critical role in all management systems; however for traffic management systems IQ is often of limited consideration. The general approach to the study of IQ has offered numerous management approaches, IQ frameworks and list of IQ criteria. As the volume of data increases, IQ problems become pervasive. An example is decisions within traffic surveillance and management system, with its large amount of real time data and short decision times. Recognizing limitation in the applicability of current IQ frameworks with our work we aim to provide a practical-orient approach, and propose a process centric IQ monitoring framework that can be useful for traffic surveillance and management systems. Key to our work is the perspective of information systems as information manufacturing system (IMS). Objective of our information quality monitoring framework is to develop a comprehensive monitoring system that complements traffic management and surveillance systems.

1 Introduction

Ensuring the quality of information has become an increasingly important factor in decision support and management systems [7,11,33,34]. However for traffic management systems IQ is often of limited consideration. Thus it is not surprising that traffic surveillance and management systems need to address the issues of improving information quality in order to be successful. Recognizing the importance of Information Quality (IQ), practitioners and researchers have considered for many years ways to improve its quality. Scientists have worked on mathematical and statistical models to introduce constrain based mechanism to prevent data quality problems. Management of the process of data generation and the management of Information Manufacturing Systems (IMS) have also attracted many researchers. With the increasing importance of IQ, much research in recent years has been focused on IQ assessment. Researchers have developed many frameworks, criteria lists and approaches for assessing and measuring IQ. The frameworks most widely used have been recently documented and adopted by the International Standards Organizations (ISO) [16].

IQ has been often defined as a measure for 'fitness for use' of information [32]. The discussion follows the general quality literature by viewing quality as the capability to 'meet or exceed users' requirements.' Common examples of IQ

Helfert M., Hossain F. and Foley O. (2011).
Monitoring Information Quality: With Applications for Traffic Management Systems.
In *Proceedings of the 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management*, pages 114-123
DOI: 10.5220/0004473601140123

dimensions are accuracy, completeness, consistency, timeliness, interpretability, and availability. Over the last decade, many studies have confirmed that IQ is a multi-dimensional concept [e.g. 3,15,28,31,32] and its evaluation should consider different aspects. The literature provides numerous definitions and taxonomies of IQ dimensions analyzing the problem in different contexts. Also, literature provides us with numerous case studies, investigating IQ in practice.

However, the practical application of most of the proposed approaches is still very limited and continuous improvement activities for data quality are rarely integrated. Furthermore although some studies examine effects of IQ [e.g. 6], applications of IQ Management to traffic management and surveillance systems are rare. Therefore the IQ problem continues to exist. In addition as the volume of data and the complexity of traffic management system increases [9], IQ problems become pervasive. Most frameworks are aiming to adopt IQ criteria and to develop suitable and domain specific measurements. Currently this process requires intensive domain expertise, as the adoption of the frameworks is limited. Furthermore, Knight and Burn [21] point out that despite the sizeable body of literature available relatively few researchers have tackled quantifying some of the conceptual definitions. We also observed that despite the inherent subjective nature of IQ, most researchers focus on providing a general applicable IQ framework without considerations of its adoption in different environments. Neither process mapping nor data modeling provides sufficient provision to define the required quality that data/information must conform to. Furthermore, on-going monitoring of the conformance of the information production process is not possible without developing a cost and time prohibitive data monitoring system. The problem of data and information quality increases as the volume of data and the time requirements increase.

Recognizing the limitations of current approaches and aiming to provide a practical-orient approach, we propose a process centric data quality approach that can be useful for traffic surveillance and management systems. In contrast to other approaches we do not aim to develop a domain specific IQ approach. In our work we aim to develop a more general approach that can then be used in several contexts, include traffic management and surveillance systems. It also incorporates a technique to specific suitable data product qualities and assesses its conformance. Objective of data quality monitoring framework is to develop a comprehensive monitoring system that for instance can be used to complement the traffic management systems in form of an information manufacturing system. To test the approach on key develop step is to design and implement a process independent monitoring system that will continuously monitor data in traffic management systems to ensure various aspects of data and information quality.

In this paper we present the overall framework, consider the benefit of a process centric framework for on-going data quality monitoring and discuss its application to traffic surveillance and management systems. Our results show that the context dimension is crucial in IQ assessment and that our framework helps to form context-aware IQ assessments that indeed can be applied to other contexts. The paper is structured as follows. In Section 2 we reflect our work with related research and outline limitations of current approaches. In Section 3 we propose a context-aware IQ framework which is then used to outline or data quality monitor framework in Section 4. Section 5 concludes the article and presents indications for further research in form of implementing the framework for traffic management and surveillance systems.

2 Related Work

IQ has been investigated for many years and numerous frameworks and criteria lists have been proposed. Although claims are made to provide generic criteria lists [32], on closer examination most research has been focused on investigating IQ within a specific context [e.g.1,4,10,13,20,24]; however traffic surveillance and management systems are little or not considered [12]. Analyzing some popular IQ frameworks [18,26,32] we can observe a large number of dimensions and criteria associated with IQ. One of the most popular and referenced frameworks was proposed by Wang and Strong [32], and since then has been applied to many contexts and research. A critical element of any IQ assessment is to assign specific values for each IQ criteria through objective, repeatable and reliable measures. Over the years a variety of IQ assessment methodologies have been proposed. [2,8,15,19,22,27,28,30] provide examples of typical methodologies that can be compared by various criteria [14]. On the one hand, IQ is often measured with subjective perceptions from information users. On the other hand, research has developed objective IQ measures on the basis of quality criteria (mostly for intrinsic IQ characteristics such as accuracy, completeness and correctness). But as of today no widely accepted IQ framework with generic, generally applicable measurements is available. This makes the application of IQ concepts to traffic management and surveillance systems challenging. Furthermore, most frameworks do not provide any guidelines to apply the framework to various contexts. Most frameworks only provide very limited assistance for analyzing causes of insufficient IQ and often do not provide any plan for solving identified problems. Furthermore most frameworks are limited in considering specific requirements.

The limited work on IQ in traffic management and surveillance systems together with the challenges to apply foremost IQ approaches to new domains, underpins the requirement for our current work to design a more general applicable information quality management approach.

3 A Context-oriented IQ Framework

In order to develop our general applicable IQ framework, we base our concept on two traditional and well established concepts. In order to structure characteristics of information, we follow the theory of semiotics. In addition, in order to provide different quality views, we follow general quality literature and structure quality along “quality of conformance” and “quality of design”. Semiotic is a relatively widely established discipline, which has recently received increasing attention. Indeed, since the publication of Stamper [29] semiotic has revealed its relevance to information systems (IS) in many research. Stamper extended the traditional three layers of semiotics (syntactics, semantics and pragmatics) with additional aspects (physical, empirical and social aspects) forming the “semiotic ladder” that consists of the views on signs from the perspective of physics, empirics, syntactics, semantic, pragmatics, and the social world [25].

Furthermore, numerous discussions related to quality indicate that defining quality is at least as challenging as the term information itself [12,17]. This approach comprises of two aspects of quality:

(1) Quality represents certain product characteristics, which meet customer needs and thereby provide customer satisfaction.

(2) The absence from deficiencies that result in customer dissatisfaction [17].

In general, the first aspect refers to *quality of design* whereas the second aspect refers to *quality of conformance* [12]. Quality of design addresses the aspect of information requirements and information product design. “How good are the requirements met by the information product design?”

The conformance of the final information product with the product design is addressed by quality of conformance. Quality of conformance takes the divergence of design with the final product into consideration. Because low quality of design and low quality of conformance have different causes and therefore different solutions, it is fundamental to consider both aspects. High quality of design does not mean high quality of conformance and vice versa. Increasing quality of design tends to result in higher costs, whereas increasing in quality of conformance tends to results in lower costs. In addition, higher conformance means fewer complains and therefore increased customer satisfaction. In this article we limit our discussion of this view on IQ and refer to [12], in which an application to Data Warehouse Systems is illustrated.

Having established an IQ framework for relevant IQ dimensions (Table 1), it needs to be applied to a particular context such as traffic management and surveillance systems. In order to evaluate the application context, the application of an IQ framework requires an analysis of the IS environment (e.g. traffic management) prior to the measurement of IQ dimensions.

Table 1. Information quality dimensions based on Semiotic and Quality aspects [12].

Semiotic Level	**Quality Aspects**		**Measurement Approach**
	Quality of Design	**Quality of Conformance**	
Pragmatic	Relevance, completeness	Timeliness, actuality, efficiency	Information process, application
Semantic	Precise data definitions, easy to understand and objective data definitions.	Interpretability, accuracy (free-of error), consistent data values, complete data values, , believability, reliability	Comparison with real world and experience
Syntax	Consistent and adequate syntax	Syntactical correctness, consistent representation, security, accessibility	Syntactical standards and agreements

The dimensions for an environment are many and varied. For example for traffic and surveillance environment timeliness, completeness and accuracy might be of high priority. They are also highly dynamic, time-depended and different user groups will priorities different dimensions. Knight and Burn [21] indicate that the choice and implementation of quality related algorithms for Internet searching is very much dependent on the characteristics of the World Wide Web. In addition, the emergences of new information system architectures and service oriented architecture (SOA) have underpinned the importance of the environment and context to the fore. The ability for organizations to distinguish between the impact of the environment and the traditional view of IQ dimensions is vital. The employment of traditional IQ frameworks does not allow for this. This observation led us to the development of an context-oriented IQ Framework that includes the context dimension and its relation to information quality measurements.

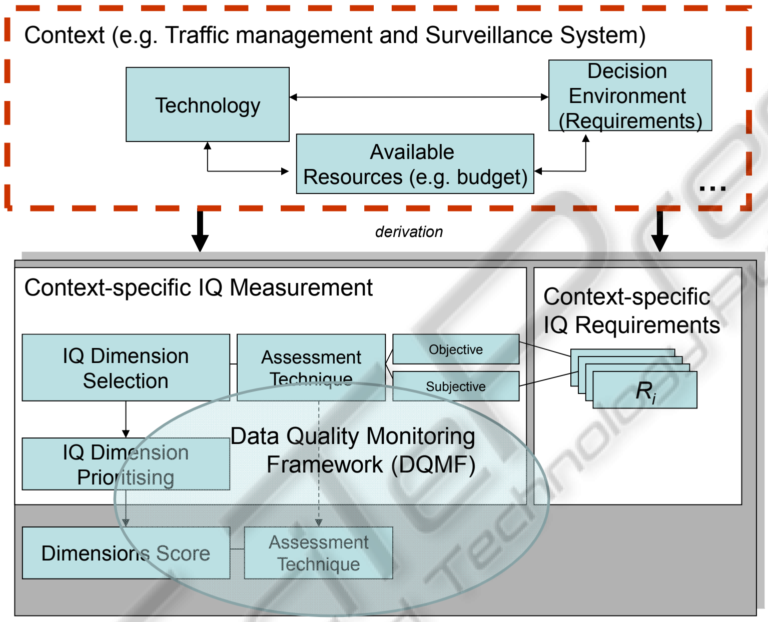

Fig. 1. Information quality framework *(grey area indicates context dimensions derived e.g. from traffic management system).*

Depending on the particular context, we are able to select and prioritize different IQ dimensions. For example, by applying Leung's [23] metric for importance, urgency and cost we can select the dimensions that are most suitable for traffic management and surveillance systems. In general, this will vary from environment to environment and within the same environment different user groups may select different IQ dimensions that most reflect the particular view of quality or the software services they require. These dimensions may change over time for reasons such as user skill level or change in software service.

The application of the framework should be on a regular basis in order to maintain the currency of the dimensions. This dynamic allows for adaptation of IQ dimensions to the constantly changing environment that more and more IS will find themselves. In our future research we aim to apply this framework to traffic management and surveillance systems, in order to identify the most important IQ dimensions. For example following dimensions could be identified (see Table 2).

Table 2. Information quality dimensions for traffice mangement and surveillance system (example).

Semiotic Level	Quality Aspects	
	Quality of Design	Quality of Conformance
Pragmatic	Completeness	Timeliness
Semantic	Easy to understand	Interpretability, believability, reliability
Syntax	Consistent syntax	Syntactical correctness, consistent representation, accessibility

4 Introducing the Data Quality Monitoring Framework (DQMF)

In order to measure the IQ dimensions, the above framework will be extended by a data quality monitoring framework which can be implemented as a comprehensive mentoring system. The key is to develop a process independent monitoring system that will continuously monitor data to ensure various aspects of IQ. In an example scenario, if traffic data could be continuously monitored to ensure notifications are sent timely, a problem could be detected much earlier and rectified with no impact to client/decision maker.

Buildings form the discussion above, we specify a data quality block in form of rules that incorporates data quality constrains and allows us to monitor certain aspects of ensuring quality. In the context of our example, if a text message fails to be sent from the traffic monitoring system, client might be not aware of a new situation. Instead of including the quality block into the system, an IMS independent quality conformance monitor would generate far better results in terms of performance. However, developing a parallel system to monitor data can also be time and cost prohibitive. Our aim in modeling quality block is also to develop data quality rules in such a way so that it can be feed to an independent data quality monitor. The framework consists of three core components. Data Quality Monitor (DQM), Data Product Markup Language (DPML) and Information Quality Markup Language (IQML).

4.1 Data Quality Monitor (DQM)

The data quality monitor is an application that accepts data product quality rules as its input and continuously monitors data product to ensure that it meets the agreed quality as defined. When designing the quality block, usually Business Process Modeling Notation (BPMN) can be supplemented by metadata about each information manufacturing block. Objective of the monitor is not to intervene in the process, but merely to monitor the data products to see if the data meets the quality requirement of the product relevant to the stage of its production. If the product fails to meet the requirement, it will report the inconsistency in accordance with agreed protocol to facilitate immediate intervention for corrective measures.

4.2 Data Product Markup Language (DPML)

A key element of our framework is DPML. In order to be effective quality controller, Information System models must describe sufficiently and accurately static, dynamic and organizational aspect of IMS. In a traditional manufacturing assembly line, as a product reaches various stages of its development, it can be inspected to ensure that it has met the requirement to be achieved at the relevant stage of the production. This is possible because a product in traditional sense will be predefined to achieve certain quality criteria that will be developed as part of designing the product. For our framework to work, we treat data as a product of information manufacturing system. At the design phase, we must then define the quality criteria that a data must meet at various stages of its production. In order to achieve this objective, we developed a Data Product Markup Language as an IP Unified Modeling Language (UML) based data product definition language. By using UML we can build on previous work to create visualized mapping of the data processes [5]. Furthermore, UML/BPMN is widely accepted is that it can be exported to code directly by cutting down on development time. This was further developed by IP MAP which extended a systematic method of representing the process involved in manufacturing of IP. Flow of data at various stages is also visualized by IP MAP. However, it lacks the ability to bridge various process and information product. There is also a need to, as described in the next section, to export the quality rules for automated execution. Hence we also base DPML on BPMN. We extend this model to model an integrated approach to define data quality requirements and business process together.

4.3 Information Quality Markup Language (IQML)

Once we are able to model data product using DPML, as described above, we need to translate it into an executable that can be processed by automated software. Otherwise, for each system a separate monitoring tool have to be developed. This is likely to make it cost and time prohibitive. This is why there needs to be ability to convert this DPML into and XML based rules that can be accepted by the monitoring tool. Information Quality Markup Language (IQML) is an XML based data product definition language. The purpose and nature of IQML is identical to that of DPML. Difference is that while DPML is UML based, IQML is XML based. IQML is either auto generated from DPML or generated independent of it. It is merely a means to facilitate data product definitions to be consumed by the Data Quality Monitor.

5 Conclusions and Further Research

As discussed above, IQ research provided numerous frameworks, criteria and methodologies to guide enterprise in the assessment, analysis, and improvement of

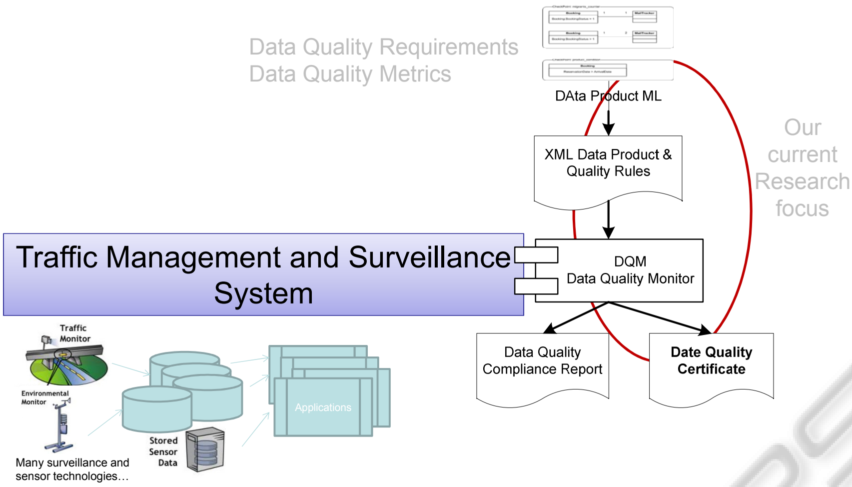

Fig. 2. Data Quality Monitor and Research Focus.

IQ. However, focusing on the critical issues related to the assessment phase, the literature does not provide an exhaustive set of metrics or guidelines that organizations can apply. Indeed examining literature on Traffic Management and Surveillance Systems, IQ in general and IQ assessments in particular are underrepresented. Our current research aims to apply common IQ approaches to various contexts, including traffic management. As illustrated in Figure 2, we aim to build a Data Quality Monitor. Most enterprises are developing their own approaches to address IQ issues although several algorithms have been developed for a subset of dimensions, such as accuracy, completeness, consistency, and timeliness. In fact the practical relevancy and generalization of some frameworks can be argued. Most common approaches used to obtain an IQ assessment is to consider domain specific measures associated with the different quality dimensions. Our research and discussion above shows the importance of context in measuring IQ. Indeed, the variations of IQ frameworks for different application scenarios indicate the significance of context in assessing IQ. Considering this observation we proposed a context-aware IQ framework. A critical element of our framework is the recognition of pragmatics (in the sense of semiotics) within our framework and the differentiation of quality of conformance and quality of design. This was used to propose a process centric Data Quality Monitoring Framework (DQMF), which can be useful for traffic surveillance and management systems incorporating data product quality and conformance. Objective of data quality monitoring framework is to develop a comprehensive mentoring system that can complement traditional traffic surveillance and management systems. The framework consists of three core components. Data Quality Monitor (DQM), Data Product Markup Language (DPML) and Information Quality Markup Language (IQML).

In future work we will implement our monitoring framework and develop a software tool that considers the context of IQ as outlined in our framework. This can be applied in the context of traffic surveillance and management systems. This will be considered for the empirical validation of the proposed context-aware IQ framework.

Furthermore, future work will also focus on the definition of an algorithm to obtain an aggregate quality measure able to assess the organizations' IQ level.

References

1. Alexander, J. E., and Tate, M. A. (1999) Web Wisdom: How to Evaluate and Create Information Quality on the Web, Lawrence Erlbaum, Mahwah, NJ.
2. Amicis, F. D. and Batini, C. (2004), A methodology for data quality assessment on financial data, Studies in Communication Sciences, 4(2), pp. 115-137.
3. Ballou D. P. and Pazer H. L.(1995) Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff, Information Systems Research, 6,1, 51-72.
4. Ballou, D. P. and Tayi, G. K. (1999). Enhancing data quality in data warehouse environments. Communications of the ACM. 42, 1, 73-78.
5. Ballou, D. P., Wang R., Pazer, H., Tayi, G. K. (1998) Modelling information manufacturing systems to determine information product quality, in Management Science, April 1998, Vol. 44, Issue 4, pp. 462-484.
6. Chen, Peter Shen-Te; Karthik K. Srinivasan, Hani S. Mahmassani: Effect of Information Quality on Compliance Behavior of Commuters Under Real-Time Traffic Information, Journal Transportation Research Record : Journal of the Transportation Research Board, 1676 (1999), pp. 53-60.
7. Chengalur-Smith, I. N., Ballou, D. P. and Pazer, H. L. (1999), The impact of data quality information on decision making: an exploratory analysis. IEEE Transactions on Knowledge and Data Engineering, 11(6), pp. 853-864.
8. Dedeke, A. (2000) A conceptual framework for developing quality measures for information systems. Proceedings of 5th International Conference on Information Quality, MIT, USA.
9. Dunkel, Jürgen , Alberto Fernández, Rubén Ortiz, Sascha Ossowski, Event-driven architecture for decision support in traffic management systems, Expert Systems with Applications, 38, 6, June 2011, Pages 6530-6539
10. English, L.: Improving Data Warehouse and Business Information Quality, New York: Wiley & Sons, 1999.
11. Fisher C.W. and Kingma B.R. (2001), Criticality of data quality as exemplified in two disasters, Information & Management, 39, 2, 109-116.
12. Helfert, M. (2001), Managing and measuring data quality in data warehousing, Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics.
13. Helfert, M. and Heinrich B. (2003), Analyzing data quality investments in CRM: a model-based approach, 8th International Conference on Information Quality, MIT, USA.
14. Helfert, M., Foley, O., Ge, M., Cappiello, C. (2009) Limitations of weighted sum measures for Information Quality 15thAmericas Conference on Information Systems, San Francisco, California August 6th-9th 2009
15. Huang K. T., Lee Y. W., Wang R.Y.(1999), Quality Information and Knowledge Management, Prentice Hall.
16. ISO/IEC 25012 (2008) Software engineering -Software product Quality Requirements and Evaluation (SQuaRE) -Data quality model, International Organization for Standardization.
17. Juran, J. M. (1998) How to think about Quality, in Juran, J. M., Godfrey A. B. (ed.): Juran's quality handbook, 5th ed., New York: McGraw-Hill, 1998, pp. 2.1-2.18.
18. Kahn B. K. & Strong D. M. (1998) Product and Service Performance Model for Information Quality: An Update, in Proceedings of the 1998 International Conference on Information Quality, MIT, USA.

19. Kahn, B. & Strong, D. M. (2002) Information Quality Benchmarks: Product and Service. Communications of the ACM, 45, 4,184-192.
20. Katerattanakul, P. & Siau, K. (1999) Measuring information quality of web sites: Development of instrument, in Proceedings of the 20th international conference on Information Systems. Charlotte, North Carolina.
21. Knight, S. & Burn, J. (2005) Developing a Framework for Assessing Information Quality on the World Wide Web. Informing Science, 8, 159-172.
22. Lee Y., Strong D., Kahn B., and Wang R. Y. (2002), AIMQ: A Methodology for Information Quality Assessment, Information & Management, 40, 2, 133-146.
23. Leung, H. K. N. (2001) Quality Metrics for Intranet Applications. Information & Management, 38, 137-152.
24. Li, S. and Lin, B. (2006), Accessing information sharing and information quality in supply chain management, Decision Support Systems, 42(3), pp.1641-1656.
25. Liu, K. (2000) Semiotics in information systems engineering. Cambridge, England: Cambridge University Press.
26. Miller, H. (1996) The multiple dimensions of information quality, in Information Systems Management, Vol. 13, Issue 2, Spring 1996, pp. 79-82.
27. Pipino, L. Lee, Y. W. & Wang, R. Y. (2002) Data Quality Assessment. Communications of the ACM, 45, 4, 211-218.
28. Redman, T.(1996) Data Quality For The Information Age, Publisher: Artech House.
29. Stamper, R. K. (1973) Information in Business and Administrative Systems. New York: John Wiley and Sons (1973)
30. Stvilia, B., Gasser, L., Twidale M. B. and Smith L. C. (2007). A Framework for Information Quality Assessment, Journal of the American Society for Information Science and Technology. 58(12), 1720-1733
31. Wand Y., and Wang R. Y. (1996), Anchoring data quality dimensions in ontological foundations, Communications of the ACM, 39,11, pp.86-95.
32. Wang, R. Y. and Strong, D. M. (1996) Beyond accuracy: What Data Quality Means to Data Consumers. Journal of Management Information System, 12,4, pp. 5-34.
33. Xu, H. and Koronios, A. (2004), Understanding information quality in E-business, Journal of Computer Information Systems, 45(2), pp. 73-82.
34. Xu, H., Horn, J., Brown, N. and Nord G. D. (2002), Data quality issues in implementing an ERP, Industrial Management & Data Systems, 102(1/2), pp. 47-59.