Data Quality Enhancing Software for Asset Management – State of the Art Evaluation

SASA BASKARADA, JING GAO, SHIEN LIN, GING SUN YEOH & ANDY KORONIOS
School of Computer and Information Science
University of South Australia
Mawson Lakes SA 5095
AUSTRALIA

This research is conducted through the Centre for Integrated Engineering Assets Management (CIEAM)

Abstract: - This paper places a special focus on data quality (DQ) issues associated with asset management (AM) systems and on how software can assist in dealing with those various DQ issues. It does not intend to merely be a review of DQ software, but rather aims to identify limitations of such software solutions when used in conjunction with AM applications. Therefore, this paper can be valuable to practitioners, researchers and software developers who are specializing in, studying, developing or adopting a computerized software solution for data quality maintenance in AM systems.

Key-Words: - Data Quality, Data Cleansing, Asset Management

1 Introduction

Industry has recently put a strong emphasis in the area of asset management (AM). In order for organizations to generate revenue they need to utilize assets in an effective and efficient way. Often the success of an enterprise depends largely on its ability to utilize assets efficiently. Therefore, effective asset management plays an important role in driving organisations to success in the highly competitive business world. Furthermore, as companies today are running leaner than ever before, physical assets are being pushed to their limits as engineering enterprises attempt to continuously drive more productivity out of their equipment in order to improve their bottom lines. It can therefore be seen that asset management is moving to the forefront of contributing to an organisation's financial objectives. Effective physical asset management optimises utilisation, increases output, maxmises availability, and lengthens asset lifespan, while simultaneously minimising costs.

The area of asset management is quite broad, in addition to examining the whole gummite of lifecycle management of the asset it also takes into account organizational and people issues. A detailed discussion of asset management is beyond the scope of this paper and a special focus will placed on the data quality (DQ) issues associated with AM and on the use of software to assist in dealing with the issues of DQ in asset management. As contemporary data cleansing software are very generic to client databases, this paper will identify the limitations of such software when used in the AM context.

The paper will be of interest to practitioners, researchers and software developers who are specializing in, studying, developing or adopting a computerized software solution for data quality maintenance in AM systems.

2 Asset Management Process

The objective of an asset management system is to minimize the long-term cost of owning, operating, maintaining, and replacing the asset while ensuring reliable and uninterrupted delivery of quality service. At its core, asset management seeks to manage the asset from the time of its commissioning until its retirement and disposal. This is because, in addition to managing the present and active asset, asset management also addresses planning and historical requirements.

The process of asset management is sophisticated because it is an engineering and planning process that requires substantial information to be collected from many different parts of the organization. This information must be maintained for many years in order to identify long-term trends. The asset management engineering and planning process uses this information to plan and schedule asset maintenance, rehabilitation, and replacement activities. The information management system that captures, maintains, and provides the needed asset information is critical in providing effective asset management.

3 Data Quality Issues in Asset Management

Despite its unique nature & role, asset management is not considered as a core business activity by many businesses, which therefore depend on traditional organizational information sources to manage assets. These traditional sources reflect both the tacit and implicit knowledge of engineers, and operators, as well as information contained in information systems, which have been primarily designed to increase productivity rather than to improve the efficiency of the processes involved in production. At the same time, there are a variety of operational and administrative systems in asset management, which not only control and manage the operation of asset equipment but also provide maintenance and administrative support throughout the entire asset lifecycle. In practice, data is collected both electronically and manually, in a variety of formats, processed in isolation, stored in a variety of customized and off the shelf legacy systems, shared among an assortment of operational and administrative systems, and communicated through a range of sources to an array of business partners and sub contractors. Data captured and processed by these systems is not comprehensive and is process dependent, making it difficult to be reused for any other process or process innovation.

The effective process of asset management has to utilize a large number of data for maintenance requirement. There has always been a limited degree to which data has been obtainable, sometimes due to the lack of data acquisition standards, sometimes due to company culture, and often due to the inability of a business to discern operational from strategic data and information. Furthermore, due to the multiplicity of systems, stakeholders, and system requirements, and the level of unpredictability in asset operation within asset management, it is often difficult to tap user requirements. This consequently contributes to the 'dirtiness' of asset data. In managing physical assets through the entire asset life cycle, large amounts of data are needed for long term performance and reliability prediction, as well as informing the decision making process on when to retire an asset. Although very large amounts of data are being generated from condition-monitoring systems, little thought has been given to the quality of such generated data. Thus the quality of data from such systems may suffer from severe quality limitations.

4 Data Quality Impact

DQ is one of the critical problems facing organizations today. As management becomes more dependent on information systems to fulfill their missions, DQ becomes a larger issue in their organizations. Poor data quality is pervasive and costly. There is strong evidence that data stored in organizational databases are neither entirely accurate nor complete [1]. In industry, error rates up to 75% are often reported, while error rates up to 30% are typical [2]. Problems with data quality may lead to real losses, both economic and social. Davenport states "no one can deny that decisions made on useless information have cost companies billions of dollars" [3]. Poor data quality has many impacts on decision-making. People make choices based on limited resources (data), and misinformed people tend to make poor decisions. Therefore, there is a growing need to develop adequate tools to deal with data quality issues in engineering asset management systems.

Since data quality issues are at the heart of effective asset management, it is essential to ensure the quality of data in monitoring systems, control systems, maintenance systems, procurement systems, logistics systems, and range of mission support applications in order to facilitate effective asset management. However, there is strong evidence that data quality issues have become increasingly prevalent in practice [4], [5], and most organizations have experienced some level of data quality problems within their firms [6]. Many of firms have not taken action to deal with these issues. Part of the reason is that data quality is not a stated priority in most organizations [7], even though data quality problems can have significant social and business impacts [8]. Therefore, there is a growing need to develop adequate tools to deal with data quality issues in engineering asset management systems.

5 Data Cleansing Software

A number of third-party software review reports have been found from various sources [9], [10], [11]. Among these, Data Management Review magazine (dmreview.com) considers itself as a major contributor to the DQ Software development. It describes itself as a publication for business intelligence, analytics, integration and data warehousing. DMReview.com provides a wide range of data quality product reviews, which were published in either DM Review magazine or on DMReview.com website.

The reviews provide comprehensive information including:

- Date of the review.
- Details of the company using the product (including a short overview of the company background).
- Platform details (including which operating system is being used, servers, databases, etc.).
- Reviewer details (name and position).
- Product name (product being reviewed).
- Details of the software company making the product being reviewed.
- Problem solved. This includes description of the need for the data quality tool and the problem statement.
- A product functionality section reflects the company's view about the product's functionality. In this section reviewers express their satisfaction or dissatisfaction with the functionality of the product.
- The Strengths section allows the reviewer to enumerate product's strengths.
- The Weakness section allows the reviewer to enumerate product's weaknesses.
- Selection criteria and the reason for choosing this particular product are stated as well.
- The Deliverables section lists the resulting deliverables.
- Vendor support is also evaluated.
- A documentation section allows the reviewer to express their opinion about the documentation, supplied with the product.

It is not intended to reproduce such reviews in this study but rather to evaluate how several commercially available data cleansing software packages would fit in the Asset management context. A comprehensive analysis of the features of such software packages was made and an evaluation on how useful such features were in assisting in the provision of a high level of data quality in asset management we carried out as well.

The following summary provides results of this systematic evaluation of the basic functionalities of a number of data quality enhancing software products. In general, the four main functions incorporated in the majority of DQ solutions are as follows.

5.1 Data Profiling Function

In data profiling, the DQ solution aims to discover the quality and characteristics of various data sources. Through a data profiling function, data anomalies can be uncovered by inspecting the true content, structures and relationships hidden within data sources. More importantly, this function examines the structure of data to find if the information matches the corresponding metadata. Furthermore, it analyses data values to find areas that are incomplete or inaccurate, and verifies relationships across columns and tables [12].

Trillium's Discovery solution takes a snapshot of data from disparate systems and builds a Metabase where users can explore the data offline, finding inconsistencies, anomalies, and errors that will help them tune the business rules [13]. In contrast, other solutions like DataFlux [12] automatically identify data quality issues in a variety of ways, including basic statistics, frequencies, ranges and outliers, multiple spellings of the same content, data patterns and formats, redundant data, primary/foreign key relationships across data sources, etc. Additionally, detailed audit reports are generated by Athanor solution including both low level drill-down reports and high level scorecard reports for identifying and categorizing actual DQ issues based on generic and user-defined business rules [14].

5.2 Data Cleansing & Matching Function

After conducting data profiling, a general DQ solution starts to improve data quality by standardizing, validating and correcting data. The majority of systems parse data into atomic components, standardize and correct customer data (i.e. names, titles, addresses, emails, etc.) to verify and validate data accuracy for better marketing campaigns. Although such concepts can be applied to asset management informational repositories this is has not yet been implemented by vendors of such software packages. FirstLogic and Athanor encompass operational data (i.e. supply chain, financial, logistics, etc.) in their DQ functions to improve business analysis and cost control [15], [14]. Subsequently, the data quality function aids in the matching and consolidation of data into a single record view and provides an enterprise standard to identify multiple occurrences of the same data. After the establishment of matching, information can be consolidated to create households and build/maintain relationships [16].

5.3 Data Enhancement Function

In order to add value to current information, general DQ solutions also include a data enhancement function, sometimes known as data augmentation, to generate or append additional bits of data from other internal or external data sources to information already in use. These additional data sources can extrapolate meaningful details from widespread bits of data within data sources [18]. Data enhancement functions can include postal address augmentation, geocoding, commodity coding and categorization, phone area code validation, consumer demographic information, suppression services, real-time address and phone verification, and watch list compliance such as public criminal watch lists [18]. Again little has adaptation to asset management systems has been considered.

5.4 Data Monitoring Function

The data monitoring function examines data continuously to detect new inconsistencies and to indicate where changes in data sources and business rules may necessitate further tuning [13]. Rules are put into place to assess data regularly and, if data does get out of control, users will be altered immediately and action can be taken to address the problems before the quality of data declines [19].

In doing this, DataFlux solutions maintain accurate and reliable data in a variety of ways, by identifying trends in data quality, providing alerts of violations in established business rules, understanding the costs associated with business rule violations and detecting variances from cyclical runs. Moreover, Athanor solution produces detailed reports to build a scorecard, which loops back into the DQ management process and serves as a framework for continuous improvement. Output from the scorecard can be shared with third party reporting tools providing enterprise-wide intelligence, identifying and tracking critical factors for business decisions.

6 Pilot Study Findings

From the literature review and the pilot study, it is observed that most of the existing DQ software approaches focus on maximizing the contributions of Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), Supply Chain Management (SCM), and sales and marketing promotion. In fact, there is very little emphasis on enhancing data quality in the asset management area, in particular the information quality of physical engineering assets. Thus, this reinforces the need for data quality software to solve data quality challenges for enterprises with large asset portfolios, across all industries.

6.1 Direct Access to Source Data

The first and foremost challenge of data quality in asset management is due to data sources often existing in multiple and disparate locations across a wide geographic area [20]. Collection of this data is the first key challenge in implementing an effective data quality solution. Moreover, the distribution complexity is often coupled with a variety of data integration and exchange technologies that are already in place and this would further increase the complication of precisely accessing these sources [11]. Therefore, in order to eliminate this problem, it is important that the data quality solution for asset management must be able to connect directly (or transparently pass through the integration layer) to this widely-distributed source data. This suggestion is particularly essential for condition monitoring.

6.2 Classification of Source Data

In view of the massive amount of data collected, the ability to recognize and interpret the data of different assets in an assortment of locations is another great challenge to implementing a successful data quality program. In practice, the source data exists in a wide variety of formats depending on various factors such as location, culture, and vendor’s preset standards. In asset management contexts, data is captured both electronically and manually, in a variety of formats, distributed over a complex communication network and shared among key stakeholders [20]. Often, given the complexity of real-time data acquisition and monitoring, the individual elements within a disparate assets record may not be fully separated.

Thus, before the implementation of data quality improvement processes, the source data must be parsed out into the appropriate classified components using mathematical functions or preset business rules. The parsed data can be recognized easily by the data quality systems and, most importantly, the classification of all data sources enables the critical identification of relationships between entities [21].

6.3 Validation of Source Data

In managing engineering assets, there are a large number of vendors, each having individual operational interface and configuring procedures. Often, this may lead to unintended errors in data collection during configuration and/or operation phases. What is more, there are varied standards in data communication; for example, 2Mbps is the European standard format for digital transmission whereas it is 1.5Mbps for North America. Therefore, the classified data must be validated against known, up-to-date reference data for accuracy, integrity, consistency and conformity to ensure the 'cleanliness' of collected data. The sources used for this process may include standardization bodies like IEEE, industry reference sources or the organization's internal sources [22].

7 Conclusion and Future Research

Due to the unique business nature, organizations have different requirements in their asset management processes. For example, a utility company has a different asset management process as compared to a logistics company. Consequently, the data capture, storage, processing and maintenance requirement could be very different. Although there are a number of commercial data cleansing software solutions available, we found only one product (Athanor Suite by Similarity Systems) which has integrated asset management data cleansing modules. These modules are designed to address the general requirements instead of the unique process within the individual organization. Thus, there is an arising need for a scrutinized research into how to improve the current data cleansing software for specific asset management data management requirement.

This study began with an analysis of the need to improve the asset management process mode. A close examination was conducted on each of the four selected data quality enhancing software packages. It evaluated how each package supports the individual data quality requirements in each asset management process. A further review of specific functionalities for asset management will be examined in future research and this will most likely result in functional requirements guidance for software vendors in this important area of information systems. .

8 Acknowledgements

This research is conducted through the Centre for Integrated Engineering Assets Management (CIEAM). The support of CIEAM partners is gratefully acknowledged.

References:

[1] Klein, B. D., 1998, User detection of errors in data: learning through direct and indirect experience, Paper presented at the AIS98.

[2] Redman, T. C., 1996, Data Quality for the Information Age, Artech House, Norwood, MA.

[3] Davenport, T. H., 1997, Information Ecology, Oxford University Press, New York.

[4] Redman, T. C., 1998, "The Impact of Poor Data Quality on the Typical Enterprise", *Communications of the ACM*, Vol. 41, No. 2, pp. 79-82.

[5] Wand, Y. and Wang, R., 1996, "Anchoring Data Quality Dimensions in Ontological Foundations", *Communications of the ACM (CACM)*, Vol. 39, No. 11, pp 86-95.

[6] Huang, K.T., Lee, Y., Wang, R., 1999, Quality Information and Knowledge, Prentice Hall, Upper Saddle River.

[7] Lee, Y. W., and Strong, D. M., 2003, "Knowing-Why About Data Processes and Data Quality", Journal of Management Information Systems, 20(3), pp. 13-39.

[8] Strong, D.M., Lee, Y. W., Wang R.Y., 1997, "Data Quality in context", *Communications of the ACM*, May 1997, pp 103- 110.

[9] DMReview 2005, Product Reviews, viewed on 15 Mar 2005, <http://www.dmreview.com/resources/reviews3.cfm?Topic=230005>.

[10] FirstLogic 2005, Customer Stories, viewed on 12 Jan 2005, <http://www.firstlogic.com/stories/>.

[11] Astech Intermedia 2005, Customer Success, viewed on 20 Mar 2005, <http://www.astech-

intermedia.com/40/2/Case%20Study%20Firstlogic%20Web%20Automation.pdf>.

[12] DataFlux 2005a, Data Profiling, viewed on 24 Jan 2005, <http://www.dataflux.com/Data-Management>.

[13] Trillium Software 2005, Trillium Software Discovery, viewed on 24 Jan 2005, <http://www.trilliumsoftware.com/site/content/products/ts-discovery.asp>.

[14] Similarity Systems 2005, Taking Control of Data Quality with ATHANOR, viewed on 24 Jan 2005, <http://www.similaritysystems.com/download.php>.

[15] FirstLogic 2005a, Data Cleansing, viewed on 24 Jan 2005, <http://www.firstlogic.com/solutions/dataQuality/iqSuite/dataCleansing.asp>.

[16] FirstLogic 2005b, Data Matching and Consolidation, viewed on 24 Jan 2005, < http://www.firstlogic.com/solutions/dataQuality/iqSuite/matching.asp>.

[17] DataFlux 2005b, Data Quality, viewed on 24 Jan 2005, <http://www.dataflux.com/Data-Management/Data-Quality/index.asp>.

[18] DataFlux 2005c, Data Augmentation, viewed on 24 Jan 2005, <http://www.dataflux.com/Data-Management/Data-Augmentation/index.asp>.

[19] DataFlux 2005d, Data Monitoring, viewed on 24 Jan 2005, <http://www.dataflux.com/Data-Management/Data-Monitoring/index.asp>.

[20] Koronios A & Lin S., 2004, Key Issues in Achieving Data Quality in Asset Management, School of Computer and Information Science, UniSA, Adelaide.

[21] Madnick S., Wang R., Chettayar K., Dravis F., Funk J., Katz-Haas R., Lee C., Lee Y. & Xian A., Exemplifing Business Opportunities From Corporate Household Research, MIT, Massachusetts, April 2003.

[22] Group 1 2005, The Data Quality Process, viewed on 24 Jan 2005, <http://www.g1.com/Products/DataQuality/TheDataQualityProcess>.