An Automated Data Validation Approach to Enterprise Asset Management for Power and Utilities Organizations

Kennedy Oyoo
Daniel Berleant
College of Science, Technology, Engineering, and Mathematics
University of Arkansas at Little Rock
Little Rock, AR, USA
kooyoo@ualr.edu, jdberleant@ualr.edu

Abstract—**Power and Utilities (P&U) organizations generate important data from Enterprise Asset Management (EAM) systems, which are used to help manage physical asset life cycle, operations, and related business processes. A range of physical asset types are used in power generation, transmission, and distribution. Asset Data Quality (ADQ) in EAM is one area which is often overlooked during EAM system implementation. The information quality focus has been on the final database, leading to rework and even persistent data quality deficiencies, thereby losing the significant benefits of enforcing data quality by designing data structures correctly at their source before data records are added to the database. Good quality asset data supports organizational objectives, where quality is associated with the "fitness for use" of data as the overarching, multidimensional perspective on the quality of the data. This is about the fitness of data for supporting operations, and distinct from the fitness for use of the equipment itself. A high quality asset data set also directly contributes to the decisions the asset owner must make to increase asset availability, optimize overall cost of asset maintenance, and reduce risks associated with asset operation. This paper proposes the Automatic Data Validation (ADV) Approach for validating fitness for use of asset data using three data quality dimensions: completeness, uniqueness, and consistency. An implementation, ADV Tool, is also presented as a proof of concept to show how it can benefit EAM business processes in P&U organizations by improving asset data quality.**

Keywords—Data Quality; Enterprise Asset Management; Power and Utilities; Data Validation

I. INTRODUCTION

Asset data quality plays an important role in Enterprise Asset Management, or EAM programs in P&U organizations. The data quality plays an important role because decision-making regarding the health of physical assets used in the generation, transmission, and distribution of power depend on it. Thus, the success of any EAM program in a P&U organization is highly dependent on the quality of asset operational data accumulated over multiple years. Indeed, achieving the desired data quality in asset management is a key challenge engineering organizations face today [2]. Nevertheless, the importance of asset data quality in EAM programs is not yet well understood [12].

Understanding the relevant data quality characteristics provides the ability to measure, manage, and report on any data that does not meet the desired levels of these characteristics. ISO 8000 [10] supports this by defining characteristics of information and data that determine its quality. This is valuable because previous studies in asset management data quality suggest a common and critical concern with EAM programs is the lack of a standardized framework to assess data quality. This is due in part to the rapid growth in asset data quantity especially from sensor enabled assets and the increase in data complexity [3]. The literature also reveals that data quality issues experienced P&U organizations are similar to those in other asset-intensive industries [5].

Despite having full access to this data, P&U EAM programs are wrestling with the reality that not all the data accumulated are equally useful in providing key insights required for tracking and managing asset lifecycles and maintenance. This arises in part when data is extracted from disparate systems in different formats and structures before it is loaded into a single comprehensive EAM system, carrying with it any preexisting data quality shortcomings.

The primary disparate systems implemented to support EAM programs mainly exchange financial, human resource, supply chain, asset, and work management data. In order to provide further understanding of the complex asset management lifecycle processes, including data exchange between EAM systems, the collaborative asset lifecycle management model is illustrated in Figure 1.

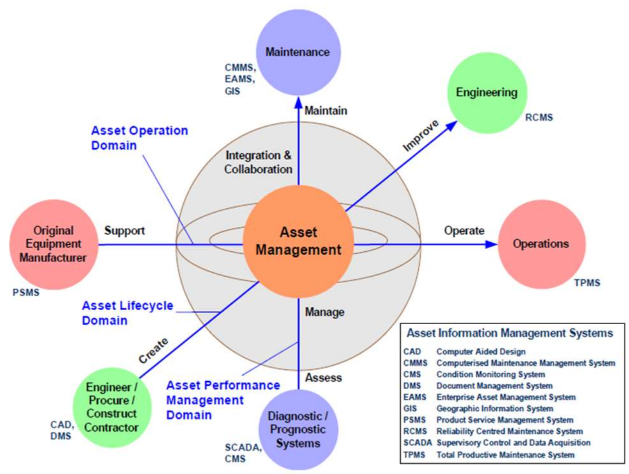

Fig. 1. Collaborative Asset Lifecycle Management, Asset Life-Cycle Managment and Asset Information Management (based on [13], p. 2).

To consolidate the data from these varied and often legacy systems, the standard process used during the implementation of EAM systems has always been extract, transform, load (ETL). However, there are indications [1] that, in a P&U context, the ETL process alone is not enough to guarantee the quality of data required by asset management practitioners. Consequently, the typical reliance on custom solutions to assess the quality of the data only after the ETL process poses a challenge to its quality.

The challenges in finding a good way to assess the quality of asset data in EAM programs are many, but certainly an important one is finding a standardized method that can be used uniformly, repeatedly, and reliably. The proposed Automated Data Validation (ADV) Approach enables asset management practitioners to define the asset data structures and business rules through a set of templates to be used by the implementation teams to better assess the quality of data both before and after the ETL process. The ADV Approach for this data quality assessment should be targeted for use by key asset management stakeholders such as reliability engineers, power plant managers, maintenance technicians, project engineers, asset operations managers and power distribution linemen. These professionals typically have many years of industry experience and deep practical understanding of the underlying asset data. The issue of data quality is intertwined with how users actually use the data in any given system, and EAM systems are no exception [12]. By defining the rules in the templates, users and others will be able, for example, to approve the asset attribute specification data formats and structure given by the templates before the actual ETL process is executed.

The business rules defined in the templates that we propose will be built on the data quality dimensions of completeness, uniqueness, and consistency. This is because P&U asset data quality issues observed by the first author typically fall within these dimensions. For the completeness dimension, data is complete when it fulfills expectations of comprehensiveness. For example, in an asset work order record, the maintenance technician might be required to provide data in mandatory fields such as asset number, description, charge account and location since these are likely to be critical data fields for the business operations. In the same work order record, crew type data might be considered optional. If the first three data fields are provided but crew type is not provided, the asset data would then still be considered complete. An asset record scores high on the uniqueness dimension when it is not stored more than once in a single database. The consistency dimension applies when two or more overlapping asset data representations are compared. Asset data in P&U organizations is typically collected by different processes from multiple sources. As a result, similar asset data might be stored in multiple disparate systems. If an asset identifier matches across these systems, and none of the records contain fields contradicting the others, then the records are considered consistent.

Other studies that fit within the realm of this issue indicate that data quality dimensions have dependency on each other [14]. Improving dependency structure among a set of data quality dimensions is referred to as dependency discovery. Logical interdependence analysis [15], tradeoff analysis [16] and dependency analysis [11] are examples of ways to discover dependency structure among data quality dimensions. These studies reinforce an understanding that the effectiveness of enterprise asset management systems is related to the quality of the data set stored in those systems, in particular its fitness for solving problems and making decisions.

II. Motivation and Problem

As defined by ISO 55000 [4], an asset is an "item, thing or entity that has potential or actual value to an organization." Assets can be classified as tangible (physical) or intangible. In electric power distribution, physical assets include reclosers and transformers. Intangible assets are not physical objects and include such things as knowledge and processes. Enterprise Asset Management (EAM) in the power and utilities industry is a structured program involving use of people, processes, tools, and information to optimize the value to the organization of its physical assets. These are assets used in the generation, transmission, and distribution of power. The primary goal of any EAM program is to reduce the cost of asset ownership while enabling the required level of service to customers. In P&U organizations, an effective EAM program also ensures that assets throughout their life cycle maximize their value to all stakeholders and fulfill core business functionalities such as intended function, supply chain adequacy, distribution management, advanced metering infrastructure and outage management.

Another important challenge to P&U organizations is the fact that EAM programs are in most cases aligned with the information technology (IT) organization [8]. Such alignment tends to put more emphasis on the organization's information technology infrastructure than on the underlying business processes and data. Additionally, that alignment also tends to treat an EAM system as solving a centralized asset repository problem [9] instead of contributing to the strategic goals of the organization. Consequently, data quality tends to be less of a focus than it should be.

This paper seeks to shift the focus back to data quality by proposing an approach that uses a set of measurable attribute values to facilitate evaluating and measuring data to better determine its quality based on the three dimensions of completeness, uniqueness, and consistency. Wang et al. [6], organizes data quality dimensions into four categories, namely intrinsic, contextual, representational and accessibility. P&U organizations have traditionally depended on data warehouses which extract, clean, transform, and integrate data from multiple operational systems, thereby emphasizing the contextual category. These warehouses typically validate data primarily during the initial load and in most cases the data set is not updated much thereafter [7].

To demonstrate how the method we propose, the ADV Approach, will help assess data quality, sample recloser and transformer data records were selected for examination. These assets have specific attributes whose values track their operational status and maintenance conditions. For this research, the data quality of these values is measured by asking relevant EAM personnel and combining their responses.

III. Data Validation Policy Definition

A. Overview

The need for the ADV Approach to asset data validation in EAM was motivated by the first author's frequent observation of data quality problems experienced by P&U organization asset management practitioners. Our solution, and a key differentiator of the approach, is to complement the ETL approach with the flexibility to define custom data quality rules through a set of baseline templates. We address this objective by providing the ability for key asset management stakeholders to define data quality rules that are built into templates based on asset specification attributes associated with each asset class. Subject Matter Experts (SMEs) from the organization who typically have many years of EAM-relevant experience and consequent insights can be identified and leveraged to define and build the baseline templates. Figure 2 below shows a sample template definition in ADV Tool.

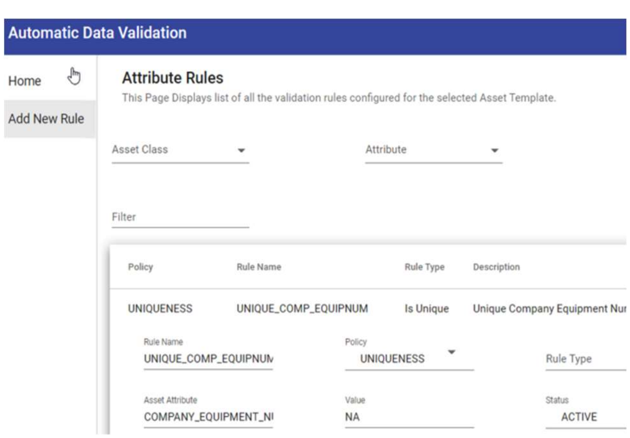

Fig. 2. Example of Template Definition Screen in ADV Tool.

B. ADV Process Flow and System Architecture

Figure 3 shows the high-level architectural plan for the ADV Approach data quality assessment system. Multiple data sources feed into a process that has several steps within each module as described next.

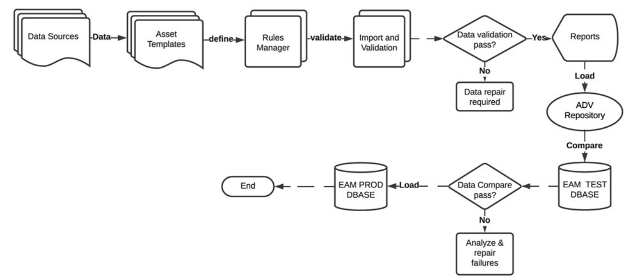

Fig. 3. High Level ADV Process Flow and System Architecture.

Data Sources: These comprise various source systems (see Figure 1) whose output data records will be mapped and consolidated into the records of the EAM system. A mapping document is used to build the new records in ADV Tool.

Asset Templates: These enable using the mapping document to define and edit asset records supplied by the source systems for the different asset classes. The templates are flexible and can also be defined to validate the output of the ETL mapping documents at the target EAM system. Once the templates are defined, they become the baseline for data comparison and validation.

Rules Manager: This module provides the capabilities to define and build data validation rules for the asset templates. These rules are evaluated against the template definitions for each of the three data quality dimensions.

Import and Validation: This implements data comparison and validation of the data obtained from the asset classes' source systems to be imported into the overall EAM system. The process involves extracting asset specification attribute data in "as is" status from source systems, for example in Microsoft Excel spreadsheet files, and importing that data into ADV Tool, which will generate a data quality report.

Reports: Provides the capabilities to display the data generated by the import and validate functions for each of the three data quality dimensions for each template-defined rule.

ADV Repository: This is the database for storing data after it has successfully traversed the import and validation process.

EAM Test Database: This database is not directly part of the solution but needed for purposes of data comparison, testing and quality assessment. The database link between the ADV Tool repository and the EAM system test environment database will involve using an approved third-party tool such as Toad for Oracle or custom scripts to perform data comparisons between these two system components.

EAM Production system: The IBM Maximo EAM system is proposed as suitable for developing and testing of the solution. Maximo EAM is a commercial product external to the proposed solution and is the final environment where validated data is to be loaded and ready for business.

IV. SYSTEM DEPLOYMENT AND IMPLEMENTATION

To put the ADV Approach and ADV Tool into practice, the web-based tool is being deployed in a test environment to provide the functionalities described herein. ADV Tool is built on Angular on the front end and connects to a Spring Boot application deployed on a Tomcat Server. The back-end database is MySQL Server. All the technology stacks are hosted by the Amazon Web Services (AWS) cloud platform. ADV Tool could also be built on other platforms. Once ADV Tool is installed, configured, and initialized, the start screen appears as shown in Figure 4 below. The IBM Maximo EAM system was chosen to host the results of data validation and ETL.

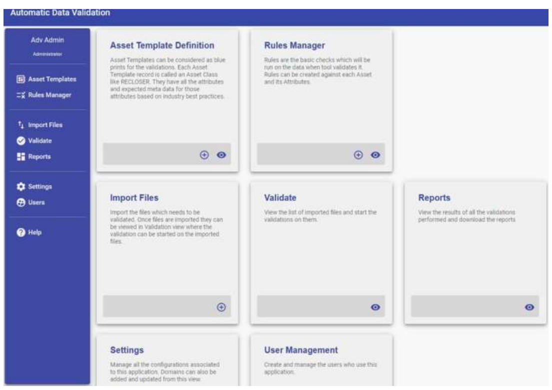

Fig. 4. ADV System Home Page.

The process of performing asset data validation started with the selection of six hundred (600) recloser and transformer assets, three hundred of each. These assets were identified from a pool of 25,000 reclosers, transformers, breakers, switches, substation breaker panels, poles, etc., with incorrect data such as missing values, duplicate values, incorrect descriptions, or missing data in the attribute specification fields. Although we started with only 600 of them, the sizeable larger pool exemplifies the need for ultimately creating a solution based on an automated, scalable method like the ADV Approach to resolving asset data quality problems.

A typical recloser asset has 178 specification attributes that store various values that may be useful for tracking its maintenance history while a 75 KVA, comp 3 transformer used in this research has 66. We selected nine (9) attributes for each of the 600 assets. This translates to 5,400 asset specification attribute values (600 x 9 = 5,400). These attributes are indicated in Table I and Table II and are considered critical to the normal operations of these two asset types.

We next distributed twenty questionnaires (to be followed by a second set of questionnaires later), ten (10) each to Asset Reliability Engineers (AREs) and Asset Maintenance Technicians (AMTs) at randomly selected sites. The questionnaires guided the participants on rating each of the three data quality dimension for each of the nine attributes. These questionnaires enabled us to understand how the two groups of participants viewed the quality of the values for the nine-attribute specifications in the Maximo EAM system. This represented the status of attribute values previously loaded through the ETL process without the ADV Approach.

The participants were asked, for each attribute, to put a check mark next to each data quality dimension for which they believed the value in the Maximo EAM system met appropriate standards. Examples are shown in Figure 5. Each check mark was scored with a value of "1" indicating **high** data quality. If a data quality dimension was not given a check mark, the score was 0 for that respondent's assessment of that dimension of that attribute value, representing **low** data quality. We then collected the responses as shown in Table III.

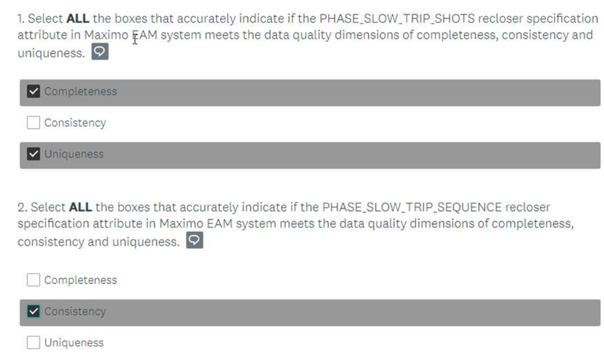

Fig. 5. Sample Questions for Each of the Three Dimensions.

Before the second batch of questionnaires was distributed to the same participants, we examined asset specification attribute values for the same 600 assets stored in the Maximo EAM system test environment and downloaded in Microsoft Excel format. Figure 6 shows a Maximo recloser asset record template. We then compared the Maximo values of the 600 assets with the values defined in the ADV Tool template definitions and corrected the Maximo values as necessary. Based on the outcome of the correction process, we then produced new templates that captured all the attribute values including the corrected ones from Maximo. The next step involved the definition of rules in the ADV Tool Rules Manager for each of the data quality dimensions of the updated templates. We then imported the Excel spreadsheet that was exported from Maximo into ADV Tool using the Import and Validate module. Figure 7 shows the sample validation results in the Reports module for two data quality dimensions, completeness and consistency.

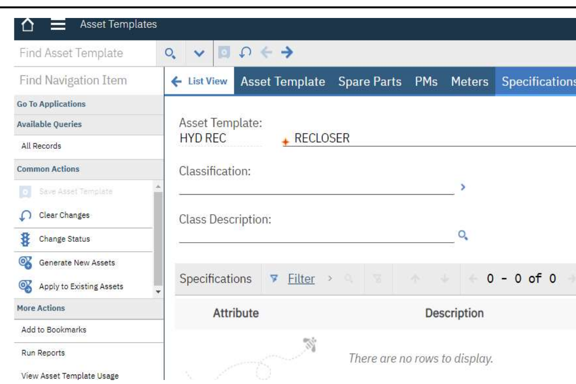

Fig. 6. Asset Specification Attribute Template from EAM System.

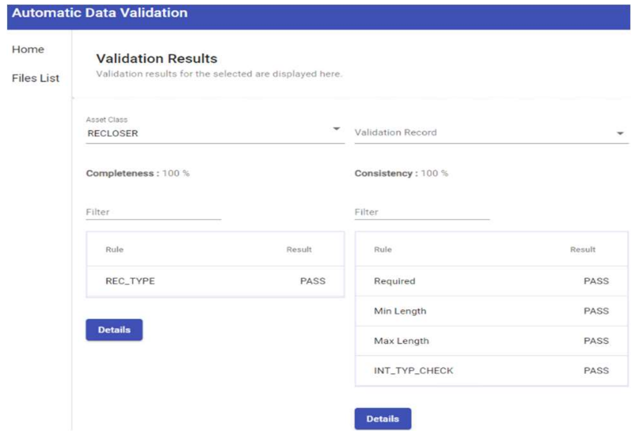

Fig. 7. Sample Validation Report in ADV.

TABLE I. RECLOSER ASSET SPECIFICATION ATTRIBUTES SELECTED.

Asset Class	Attribute Specifications	Data Type	Value
Recloser	PHASE_SLOW_TRIP_SHOTS	ALN	3
Recloser	PHASE_SLOW_TRIP_SEQUENCE	ALN	B
Recloser	PHASE_FAST_TRIP_SHOTS	ALN	1
Recloser	PHASE_FAST_TRIP_SEQUENCE	ALN	A
Recloser	BUSINESS_UNIT	ALN	3-N
Recloser	MDT_AREA	ALN	3N43CHA
Recloser	LONGITUDE	ALN	89.89597
Recloser	LATITUDE	ALN	33.98459
Recloser	LOCATION_TYPE	ALN	BANK

TABLE II. TRANSFORMER ASSET SPECIFICATION ATTRIBUTES SELECTED.

Asset Class	Attribute Specifications	Data Type	Value
Transformer	LOAD_VERIFICATION_TYPE	ALN	NORM
Transformer	INSTALLATION_TYPE	ALN	Pad
Transformer	DESIGN_TYPE	ALN	Padmount
Transformer	INSTALLATION_PURPOSE	ALN	Metered Customer
Transformer	KVA_SIZE	ALN	75
Transformer	INSTALLED_STATUS	ALN	Connected
Transformer	DIS_WR_NO	ALN	BATCH
Transformer	STANDARD_PRIMARY_VOLTAGE	ALN	7620
Transformer	STANDARD_SECONDARY_VOLTAGE	ALN	120/240

TABLE III. FIRST QUESTIONNAIRE RESPONSES RECEIVED BEFORE ADV VALIDATION, 1=HIGH, 0=LOW.

Number of questionnaires distributed = 20, Number of responses received = 20, 10 each for AREs & AMTs							
Asset Type	Number of Assets Observed	Number of Asset attribute specifications Observed	DQ Dimension Observed	Asset Reliability Engineers Response Received		Asset Maintenance Technicians Response Received	
				1=HIGH	0=LOW	1=HIGH	0=LOW
Recloser	100	9	Completeness	75	25	77	23
Recloser	100	9	Consistency	54	46	61	39
Recloser	100	9	Uniqueness	83	17	81	19
Transformer	100	9	Completeness	69	31	67	23
Transformer	100	9	Consistency	55	45	84	16
Transformer	100	9	Uniqueness	78	22	73	27
TOTAL	600						

TABLE IV. SECOND QUESTIONNAIRE RESPONSES RECEIVED AFTER ADV VALIDATION, 1=HIGH, 0=LOW.

Number of questionnaires distributed = 20, Number of responses received = 20, 10 each for AREs & AMTs							
	Number of Assets Observed	Number of Asset attribute specifications Observed	DQ Dimension Observed	Asset Reliability Engineers Response Received		Asset Maintenance Technicians Response Received	
				1=HIGH	0=LOW	1=HIGH	0=LOW
Recloser	100	9	Completeness	93	7	91	9
Recloser	100	9	Consistency	81	19	82	18
Recloser	100	9	Uniqueness	96	4	87	13
Transformer	100	9	Completeness	89	11	94	6
Transformer	100	9	Consistency	88	12	89	11
Transformer	100	9	Uniqueness	95	5	95	5
TOTAL	600						

V. CONCLUSION AND FUTURE WORK

In this paper, we have introduced the ADV Approach and ADV Tool to asset data validation for EAM in P&U organizations. The ADV Approach and ADV Tool do not intend to replace the ETL process but rather complements ETL and adds another layer of asset data validation. ADV Tool introduces asset templates that must be approved by the asset management stakeholders prior to the ETL process. In our study, questionnaires were used to help understand the existing asset data quality problems in the Maximo EAM system. The responses tabulated in Table III show that there were already existing asset data quality issues in the EAM system.

From a second set of questionnaires distributed to the same participants, we can see that the results show a significant improvement from the first round of questionnaires. Based on the pilot work so far, once expanded to the full EAM process of a P&U organization, the ADV Approach and ADV Tool are expected introduce significant asset data quality improvements during EAM system implementation. As an immediate step forward, ADV Tool can benefit asset management practitioners, in our case asset reliability engineers and asset maintenance technicians, to periodically check the quality of asset data loaded in the EAM system through the ETL process.

The templates proposed also introduce data quality health checks for each of the asset specification attributes and can be extended to other assets and data types. Both proposed solutions will benefit EAM system implementation projects in P&U organizations as a concrete step towards remediating their asset data quality issues.

REFERENCES

[1] Oyoo, K., Collaboration-based automatic data validation framework for enterprise asset management, in press.

[2] Koronios, A., S. Lin and J. Gao, A data quality model for asset management in engineering organizations, 10th International Conference on ICIQ (MIT IQ), 2005, mitiq.mit.edu.

[3] Lin, S., J. Gao and A. Koronios, The need for a data quality framework in asset management, in Proceedings of the 1st Australasian Workshop on Information Quality (AUSIQ), Adelaide, June 22-23, 2006.

[4] International Organization for Standardization, ISO 55000:2014 Asset management — overview, principles and terminology, 2014, https://www.iso.org/obp/ui/#iso:std:iso:55000:ed-1:v2:en.

[5] Lin, S., J. Gao, A. Koronios and V. Chanana, Developing a data quality framework for asset management in engineering organisations, Int. J. Inf. Qual., 2007, vol. 1, p. 100–126.

[6] Wang, R. Y. and D. M. Strong. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems 1996, 12, 5–33.

[7] Rao, D., V. N. Gudivada and V. V. Raghavan, Data quality issues in big data, IEEE International Conference on Big Data, Santa Clara, California: IEEE Computer Society, Oct 2015, pp. 2654–2660.

[8] Woodhouse, J., Asset Management, The Woodhouse Partnership Ltd, 2001, accessed online April 10, 2004, http://www.plant-maintenance.com/articles/AMbasicintro.pdf.

[9] Woodhouse, J., Asset management: concepts & practices, The Woodhouse Partnership Ltd., 2003.

[10] https://www.iso.org/obp/ui#iso:std:iso:8000:-63:ed-1:v1:en.

[11] DeAmicis, F., D. Barone and C. Batini, 2006, An analytical framework to analyze dependencies among data quality dimensions, Proceedings of the 11th International Conference on Information Quality, November 10-12, 2006, Cambridge, USA., pp: 369-383.

[12] Orr, K., Data Quality and System Theory, Communications of the ACM, 41(2), 1998, pp. 66-71.

[13] White Paper - ARC, Asset Information Management – A CALM Prerequisite, ARC Advisory Group, 2004.

[14] Barone, D., F. Stella and C. Batini, Dependency discovery in data quality, Proceedings of the 22nd International Conference on Advanced Information Systems Engineering, June 7-9, 2010, Springer, Hammamet, Tunisia, pp. 53-67.

[15] Gackowski, Z., Logical interdependence of data/information quality dimensions — a purpose-focused view on IQ, Proceedings of the 9th International Conference on Information Quality, November 5-7, 2004, Cambridge, USA., pp: 126-140.

[16] Madnick, S. E., R. Y. Wang, Y. W. Lee and H. Zhu, Overview and framework for data and information quality research. J. Data Inform. Qual. 2009, Vol. 1, No. 1. 10.1145/1515693.1516680