

A QUESTIONNAIRE-BASED DATA QUALITY METHODOLOGY

Reza Vaziri ¹ and Mehran Mohsenzadeh ²

¹ Department of Computer Science, Science Research Branch, Azad University of Iran,
Tehran Iran r.vaziri@iauctb.ac.ir

² Department of Computer Science, Science Research Branch, Azad University of Iran,
Tehran Iran mohsenzadeh@srbiau.ac.ir

ABSTRACT:

Data quality (DQ) has been defined as “fitness for use” of the data (also called Information Quality). A single aspect of data quality is defined as a “dimension” such as “consistency”, “accuracy”, “completeness”, or “timeliness”. In order to assess and improve data quality, “methodologies” have been defined. Data quality methodologies are sets of guidelines and techniques that are designed for measurement assessment, and perhaps, improving data quality in a given application or organization. If an appropriate list of dimensions is available for the specific needs of an organization, a questionnaire-based methodology can be designed in order to 1. Measure dimensions and identify “weak” dimensions in the organization 2. Select a proper “strategies” to improve data quality. In this paper we propose a questionnaire-based methodology in order to achieve that.

KEYWORDS

Data Quality, Dimensions, Methodologies

1. INTRODUCTION:

Data quality has been defined as “fitness for use” of data. *Data quality methodologies* have been defined and developed to measure and improve data quality in organizations. A methodology refers to a set of guidelines and techniques that are designed for the assessment and improvement of data quality in a specific application or organization. In [Batini 2009] methodologies have been divided into three main “phases and steps”. The three steps and phases are the following:

1. **State reconstruction:** which collects contextual information on organizational data, processes and services
2. **Assessment / Measurement:** which measures the quality of data along relevant “dimensions”. The term “measurement” refers to measuring the values of data itself, and the term “assessment” refers to comparison against reference values.
3. **Improvement:** which proposes techniques and strategies for reaching higher levels of data quality, perhaps levels specified by the organization’s management.

A dimension is a single aspect of data quality such as “accuracy”. Notice, that the “Improvement / Assessment” phase measures the quality of data along “relevant dimensions”. In [Vaziri 2012] a questionnaire-based *state reconstruction phase* was proposed to identify the relevant dimensions for the organization or application at hand. In this paper we continue the same strategy to measure these relevant dimensions and identify a proper methodology to improve upon them.

The questionnaire-based methodology was selected because of two reasons, 1. It is very simple to implement where many organization cannot afford to hire expert or professional teams for data quality purposes 2. It is very general, namely it can easily be applied to almost all kinds of organizations. In the past some methodologies were designed for very *specific types of organization or information systems*, such as the Canadian Health Institute [Long and Seko 2005] or the Cooperative Information Systems [Scannapieco 2004]. However, we believe a general methodology could be more practical. In [Muller 2012] a data quality improvement strategy is proposed based on the "conflicts" of data about the same object, but from different sources. The idea is that such conflicts highlight critical areas of poor data quality and offer better data. However, we cannot always guarantee that several sources are available for the data at hand, especially in business organizations.

Measuring the *relevant dimensions* by a questionnaire-based methodology could serve several purposes. 1. It gives an overall measurement of the current quality of data in the organization, 2. It identifies “weak” dimensions, namely dimensions that require urgent and thorough attention in the organization, 3. It provides values for assessment purposes. Remember that assessment is when we compare data against “reference values”, in this case perhaps an ideal organization with very high quality data.

Identifying a *proper improvement strategy* in the same fashion can help an organization to reliably pick a proper improvement strategy, leading to many cost-saving benefits for the organization.

This paper is organized as follows: section 1 is the Introduction. In section 2 is the Related Work about data quality measurement and methodologies. Section 3 explains the proposed solution where a new questionnaire-based method is introduced. Section 4 is the evaluation, and Section 5 is the conclusion and future research.

2. RELATED WORK

Data quality dimension has been defined as a set of data quality attributes that represents a single aspect or construct of data quality [Wang 1996]. Many different lists, definitions, and classification of dimensions have been proposed in data quality literature [Wang 1996, Wand 1996, Jarke 1997, Goodhue 1995, Delone 1992, Ballou 1985, Zmud 1978 etc]. For instance, In [Wang 1996] some of the more prominent dimensions are defined as the following:

Table 1: Wang and Strong List of Data Quality Dimensions [Wang 1996]

Dimensions	Definition
Accessibility	The extent to which data is available, or easily or quickly retrievable
Appropriate Amount	The extent to which the volume of data is appropriate for the task at hand
Believability	The extent to which data is regarded as true and credible
Completeness	The extent to which data is not missing and is of sufficient breadth and depth of the task at hand
Concise Representation	The extent to which data is compactly represented
Consistent Representation	The extent to which data is presented in the same format
Ease of Manipulation	The extent to which data is easy to manipulate and apply to different tasks
Free-of-Error	The extent to which data is correct and reliable
Interpretability	The extent to which data is in appropriate languages, symbols, and units, and the definitions are clear
Objectivity	The extent to which data is unbiased, unprejudiced, and impartial
Relevancy	The extent to which data is applicable and helpful for the task at hand
Reputation	The extent to which data is highly regarded in terms of its source or content
Security	The extent to which access to data is restricted appropriately to maintain its security
Timeliness	The extent to which the data is sufficiently up-to-date for the task at hand
Understandability	The extent to which data is easily comprehended
Value added	The extent to which data is beneficial and provides advantages from its use

But how are the above dimensions identified? In [Vaziri 2012] objections are raised to using a pre-defined list of dimensions, because such a pre-defined list may not address the specific needs of the organization at hand. In [Vaziri 2012] a questionnaire-based method is used to identify the “relevant dimensions” of an organization. More specifically, three groups of subjects, namely Information Professionals (IP’s), Information Consumers (IC’s), and Independent Experts (IE’s) are given a questionnaire about the importance of the well-known dimensions in the current organization. The “open-ended” part of the questionnaire also allows the subjects to add, define, and rate new dimensions for the organization. By carefully analysing the results of all the questionnaires a list of more relevant dimensions may be obtained. Now the basic idea is that if a questionnaire-based method works well for identifying the relevant dimensions (i.e. state reconstruction), it could also be used for measurement and improvement of data quality as well.

In [Lee 2002] the AIMQ methodology is proposed, which takes advantage of similar ideas. The AIMQ methodology has three main components. The first component is the PSP/IQ model [Kahn 2002]. This is a 2 X 2 model consisting of two rows and two columns making up four quadrants. The two rows of the model correspond to the fact the information could be treated either as a product or a service. The two columns correspond to the fact that information quality could be measured either by predefined specifications or whether it meets or exceeds user expectations. Each of the four quadrants has its own specific data quality dimensions.

Table2: PSP/IQ Model [Lee2002]

	Conforms to Specifications	Meets or exceeds Consumer Expectation
Product Quality	Sound Information IQ Dimensions <ul style="list-style-type: none"> • Free-of-error • Concise Representation • Completeness • Consistent representation 	Useful information IQ dimensions <ul style="list-style-type: none"> • Appropriate amount • Relevancy • Understandability • Interpretability • Objectivity
Service Quality	Dependable information IQ dimensions <ul style="list-style-type: none"> • Timeliness • Security 	Usable information IQ dimensions <ul style="list-style-type: none"> • Believability • Accessibility • Ease of operation • Reputation

The second component of the AIMQ methodology is a questionnaire. The questionnaire is used to measure information quality along the dimensions of the PSP/IQ model. The collective measurement of the dimensions in a single quadrant makes up the data quality measurement for that quadrant. This instrument can be applied to assess the quality of information in various organizations.

The third component of the AIMQ methodology consists of two different gap analysis techniques. The first gap analysis technique compares an organization’s Information Quality (IQ) to that of a best-practice organization. In other words, it compares the IQ in a typical organization with the IQ to a very high quality organization. In industry this is called benchmarking. The second gap analysis technique compares the measurement of IQ in an organization by different roles, which is with respect to the roles of the IQ collectors, consumers, or professional.

The development of the questionnaire in AIMQ is very involved and detailed. In a three step process the methodology develops the questionnaire in the following fashion:

1) Item development: In the first step “items” are developed. Items are simple general statements addressing a single dimension by different wordings. For instance, for “completeness” the following items could potentially be selected. “This information is presented completely”, “this information includes all the necessary information”, “and this information cannot be considered incomplete.” After the items are written they were checked by the IQ researchers to see if each item addresses a specific dimension, items for a dimension do not overlap with others, and the items are meaningful to the users who eventually complete the questionnaire. This first step of the item development assigns 12-20 items for each dimension. Since the PSP/IQ model has a

total of fifteen dimensions that would produce about 240 items. This number of items is a little bit too high for practical purposes, thus, in the pilot study step the number of items is reduced.

2) Pilot study: The main purpose of this step is to make an assessment of the reliability of the items for each of the dimensions, and to use this reliability check to reduce the number of items per dimension. Reducing the number of items is very important; because the number of items developed in step one is generally too much for practical purposes. For the pilot survey, the items are randomly mixed in a booklet and given to the respondents. The scale of rating an item is 0-10 with "0" meaning "not at all" and "10" meaning "completely".

Here are the "items" that were developed for two of the dimensions "completeness" and "concise representation":

Completeness. (6 items, Cronbach's Alpha 0.87)

This information includes all necessary values.

This information is incomplete. (R)

This information is complete.

This information is sufficiently complete for our needs.

This information covers the needs of our tasks.

This information has sufficient breadth and depth for our task.

Concise Representation. (4 items, Cronbach's Alpha 0.88)

This information is formatted compactly.

This information is presented concisely.

This information is presented in a compact form.

The representation of this information is compact and concise.

The pilot study questionnaire also includes questions about demographic information. In particular the respondents' role in the information production system was identified as being a *collector* of information, a *consumer*, or an information system (IS)*professional*. The respondents' may be from different companies in different sectors such as healthcare, finance, and manufacturing.

To assess the items, Cronbach alphas are calculated and factor analysis performed. The results are used to eliminate items that did not add to the reliability of the scale. In general 4-5 items per dimension is a good number, which adds up to a total of about 65 items.

3) Full study: The final questionnaire includes 65 items plus demographic information from respondents in various organizations. In each organization there must be respondents from different role meaning collectors, consumers and IS professionals. Each respondent just focuses his or her answers on one set of information of importance to the organization, for example, patient information in healthcare organizations. Once the questionnaires are completed, just like the pilot study, Cronbach alphas are calculated and factor analysis is performed.

The AIMQ methodology is very comprehensive; however, several objections may be raised to it:

First, it uses a pre-defined list of dimensions [Wang 1996], which as explained earlier may not address the specific needs of the organization.

Second, the item development (or the questionnaire) seems to be based upon the subjective opinion of the researchers rather than well-defined specifications. In other words, if the idea is to measure a dimension via questionnaire-based study, there must be clear rules to develop the items (questions). Of course, coming up with these rules requires lots of research and case studies.

Third, a critical group of subjects, namely Independent Experts (IE's) are missing. In AIMQ gap analysis, *Consumers* may not have the proper expertise, and *Professionals* may be biased by the sense of belonging to the organization. In the next section we try to present solutions to address the above issues.

Fourth, as useful and practical AIMQ might be it only covers the *measurement/assessment* part of a methodology, and does not suggest any strategies for *improvement*. We believe the same questionnaire-based study could be used to come up with the most effect improvement method for the specific situation at hand,

3. PROPOSED SOLUTION: A QUESTIONNAIRE-BASED MEASUREMENT OF DIMENSIONS

3.1. MEASUREMENT

In this section we will present a questionnaire-based method to measure the data quality along the relevant dimensions and discover a proper method to improve data quality in an organization. The basic assumption is that the relevant dimensions for the specific organization are already identified [Vaziri 2012]. Now we identify three groups of subjects and present them with a comprehensive measuring questionnaire. The three groups of subjects are [Vaziri 2012]:

Information Professionals (IP's): These are the people who collect and maintain the information for an organization. They are also responsible for designing the systems where information resides.

Information Consumers (IC's): These are the people who use the information.

Independent Experts (IE's): These are defined as experts that have appropriate amount of practical or academic experience in the practices of the organization being evaluated. Also they are called independent because they have no vested interest in the organization being evaluated and thus can present an unbiased opinion.

First the questionnaire collects general information about each subject, such as Name, Family Name, Organization, etc. It also asks about the "Role" of each subject in the organization for later "gap analysis". Then the questionnaire covers all the relevant dimensions, and asks the subjects to rate each dimension from a scale of 0 to 10 with 0 being rated as "Not True" and 10 as "Completely True". The exact "definition" of each dimension is included in the questionnaire to avoid incoherent interpretation of the dimensions.

The measurement value of each dimension can be asked by several items (questions). The number of items for each dimension should not be too big. Too many questions can tire the subjects and lower the effectiveness of the questionnaire. The groups of items that target specific dimensions can be mixed at random so that the subjects will not detect that each dimension is

being questioned several times. This way, we could have several independent measurements for each dimension. For each single dimension these are the items that we propose:

Direct Question:

Is the Data in your organization [Dimension name]?

Reverse Question

Is the Data in your organization [Opposite of dimension]?

Synonymy Question:

Is the Data in your organization [Dimension Synonym]?

Definition Question:

Is the Data in your organization [Definition of Dimension]?

For instance, for the dimension “accessibility” our method will yield the following items:

Is the Data in your organization accessible?

Is the Data in your organization inaccessible?

Is the Data in your organization reachable?

Is the data in your organization available, or easily or quickly retrievable?

The first item directly asks the subjects about the specific dimension. This item measures the very first impression of the subjects about a dimension clearly and directly. Its simplicity and directness is its feature advantage. The second item asks about the value of a dimension, but in reverse. This item is intended to obtain the subjects opinion from a negative point of view. Of course, the value of this item must be subtracted from 10. The third item measures the dimension if it were given another name. This items tests to see if the subjects are biased towards a term that they have seen lots of times in the literature before. And the fourth item measures the value of a dimension in terms of its definition. The definition of the dimension could be based on a well-known paper such as [Wang 1996]. This items tests whether the subjects have a good understanding of the meaning of the dimension. For each dimensions the “mean value” of its related items makes up its measured value.

Looking at the AIMQ paper [Lee 2002] Appendix A where the final questionnaire is written, the items for each dimension look very similar to what we have just developed. However, the AIMQ items are developed with much subjective opinion and statistical analysis, whereas in this paper they are developed with simple rules and logic of what we are trying to achieve. Of course, only by accurate case-studies we will be able to see if there are any “items” missing from our strategy.

3.2. IMPROVEMENT

Once the “relevant dimensions” have been measured the next task at hand is to identify a proper method to improve data quality. Since we have followed a questionnaire-based study, we will continue with the same strategy in the third and final phase of our methodology namely the *improvement*. For each dimension’s measured value a system of classification like the following can be devised:

Strong: 7.0 < Dimension Value

Intermediate: 4.0 < Dimension Value < 7.0

Weak: Dimension Value < 4.0

Obviously, for data quality improvement the methodology should begin by the “Weak” dimensions. Depending on the resources and funding available in the organization for data improvement, the organization can decide whether to move on to “Intermediate” and “Strong” dimensions as well. As mentioned earlier a questionnaire must be developed for each of the “weak” dimensions.

As was the case before the questionnaire is started with demographic information and is given to the three groups of subjects IC’s, IP’s, and IE’s. For each of the weak dimensions (e.g. completeness) a separate questionnaire is given to the subjects. In the improvement questionnaire some of the well-known data improvement strategies are named and defined for the subjects. Fortunately, these are listed in [Batini 2009].

Table 3: Standard Strategies for Data Quality Improvement [Batini 2009]

1	<i>acquisition of new data,</i>	which improves data by acquiring higher-quality data to replace the values that raise quality problems;
2	<i>standardization (or normalization),</i>	which replaces or complements nonstandard data values with corresponding values that comply with the standard. For example, nicknames are replaced with corresponding names, for example, Bob with Robert, and abbreviations are replaced with corresponding full names, for example, ChannelStr. with Channel Street.
3	<i>Record linkage,</i>	which identifies that data representations in two (or multiple) tables that might refer to the same real-world object;
4	<i>data and schema integration</i>	which define a unified view of the data provided by heterogeneous data sources. Integration has the main purpose of allowing a user to access the data stored by heterogeneous data sources through a unified view of these data.
5	<i>Source trustworthiness</i>	which selects data sources on the basis of the quality of their data;
6	<i>error localization and correction,</i>	which identify and eliminate data quality errors by detecting the records that do not satisfy a given set of quality rules. These techniques are mainly studied in the statistical domain.
7	<i>Cost optimization,</i>	defines quality improvement actions along a set of dimensions by minimizing costs.
8	<i>Process control</i>	inserts checks and control procedures in the data production process when: (1) new data are created, (2) data sets are updated, or (3) new data sets are accessed by the process. In this way, a reactive strategy is applied to data modification events, thus avoiding data degradation and error propagation.
9	<i>Process redesign</i>	redesigns processes in order to remove the causes of poor quality and introduces new activities that produce data of higher quality. If process redesign is radical, this technique is referred to as <i>business process reengineering</i>

We list the same strategies in the questionnaire along with their respective definitions and ask the subjects to rank each of the strategies from 0 to 10, with 0 being *irrelevant* and 10 as *vital*. For each strategy there is also a check mark “*Already in Use*”. This check mark determines whether the named strategy is already being used in the organization. See figure 2 for the questionnaire. Also notice that last “open-ended” question. This question will help the subjects propose any possible improvement strategies that were not included in the questionnaire. Perhaps, by answering the earlier questions, the subjects will become more familiar with data improvement and come up with strategies of their own.

Figure 1: Questionnaire for DQ Improvement

Name:	Family Name:	
Industry :	Organization:	
Department:	Role: IP, IC, IE:	
Job Title:	Years of Experience:	
<p>Please rate each of the data quality improvement techniques for the organization regarding the “<i>completeness</i>” dimension. Also please state whether each strategy is already in use in the organization or not.</p>		
<p>Acquisition of new data: which improves data by acquiring higher-quality data to replace low quality values</p>		
Irrelevant	Important	Vital
0 1 2 3	4 5 6 7 8	9 10
Already in Use Yes No		
<p>Standardization: which replaces or complements nonstandard data values with standard ones.</p>		
Irrelevant	Important	Vital
0 1 2 3	4 5 6 7 8	9 10
Already in Use Yes No		
<p>Record Linkage: which identifies representations in two (or multiple) tables that refer to the same objects</p>		
Irrelevant	Important	Vital
0 1 2 3	4 5 6 7 8	9 10
Already in Use Yes No		

. . .

“Are there any improvement strategy missing from the above list that could prove useful or important for your organization? Please list them, define them, and rank them from the scale of 0 to 10 just as above.” Also state whether each of these strategies is already in use in organization or not.

Irrelevant									Important			Vital
0	1	2	3	4	5	6	7	8	9	10		
Already in Use		Yes	No									
. . . .												

Since improvement is a more complicated task than measurement, the questionnaire in this phase of the methodology will require a more careful examination than the previous phases. We propose that all the suggested improvement strategies be examined carefully and the top three strategies (i.e. top three most frequent answers) be considered as the proper improvement strategy. Also it will be interesting to see if the three groups of subjects (i.e. IP’s, IC’s, and IP’s) agree or differ significantly on the improvement strategies. If there is general agreement it is safe to assume that a proper improvement strategy has been identified. If there are significant differences more careful study needs to be done in order to detect why different groups propose different improvement strategies.

4. EVALUATION

4.1. THE CONTROL MATRIX

In [Pierce 2004] a control matrix is proposed to relate data quality problems with data quality controls. This could also serve as an evaluation tool for our questionnaire results. In such “Control Matrices” the rows identify the various quality checks or controls that are available or proposed and the columns identify the data quality problems that have been identified in an organization. Each cell in the control matrix specifies how effective a data control or check is for a data problem. Here is a simple example:

Figure 2: The Control Matrix for Improvement Strategies [Pierce 2004]

	DQ Problem 1	DQ Problem 2	DQ Problem 3	DQ Problem 4
Control 1		Medium		
Control 2	Medium			
Control 3		Strong		Week
Control 4		Strong		
Control 5			Week	Week

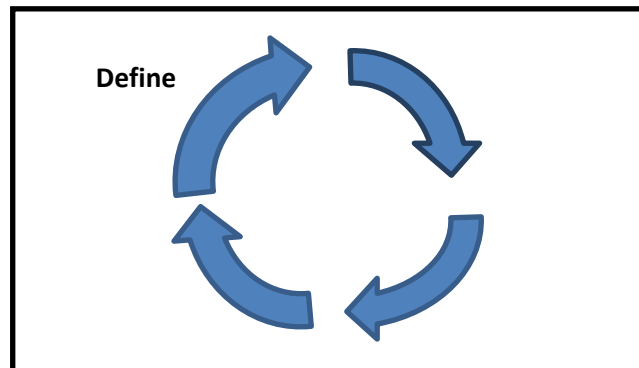
The previously-discussed questionnaires can help us fill out the control matrix. The columns, which are the data quality problems, in the case of our methodology are the *dimensions* that have been identified as *weak* by the subjects. On the other hand, the rows are controls or improvement techniques that have been proposed by the improvement questionnaire. Notice that the results from the same questionnaire can also be used to rate the effectiveness of the strategy so that the associated cell in the matrix could be filled as **Weak**, **Medium**, or **Strong**. Filling out the above matrix can additionally help with the improvement techniques. For instance, in the above matrix DQ Problem 2 has two strong controls associated with it, thus it is safe to assume that the problem is getting good attention. On the other hand, DQ Problem 3 has only one week control associated with it. So in the improvement stage, the IE's, and perhaps IP's, must introduce new controls for DQ Problems that have week or no controls.

4.2. THE TDQM CYCLE

In [Wang 1998] the TDQM methodology is introduced. The methodology views data (or information) as a product entity, because an analogy exists between product manufacturing and information manufacturing. Namely, products are produced from raw material by an Assembly Line, and in a similar way, information is produced from raw data by an Information System. Therefore, the quality issues can be resolved in the same way.

In business and product manufacturing the Deming cycle (or PDCA cycle) has been proposed, which thru a cycle of **Plan**, **Do**, **Check** and **Act** the manufacturing process is constantly being improved. In a similar way, the TDQM cycle includes the following four steps:

Figure 3: The TDQM Cycle [Wang 1998]



1. Define: Define characteristics of an Information Product **2.Measure:**Measure the Information Product by the way of developing metrics **3.Analyse:** Analyse and investigate the root causes of the problems **4. Improvement:** Identify key areas for improvement.

The TDQM cycle can be easily applied to our questionnaire-based methodology. Notice the first questionnaire defines and measures the problematic dimensions. Also the second questionnaire analyses and improves upon the same dimensions. Even though we did not specifically follow the TDQM cycle to devise this methodology, it seems that repeating the two questionnaires repeatedly after each other, it somehow follows the above cycle and will result in the overall improvement of the data quality. In [Vaziri 2012] there is also a third questionnaire which defines a customized list of dimensions. This questionnaire definitely fits into the “Define” phase of the above cycle, but it probably does not need to be repeated as the actual list of dimensions itself does not change very frequently.

4.3. GAP ANALYSIS AND RANKING

Since we are using three groups of subjects to measure and improve data, gap analysis and ranking could be done on both.

1. Gap Analysis: If all groups of subjects agree on a measurement or improvement technique it is safe to say that a reliable result has been reached. However, if there is serious disagreement between the various groups further analysis is required to explain the differences. For instance, if there is a large gap between IP’s and IC’s, the information professionals are not aware of the consumers’ concerns. On the other hand, if there is large gap between the IE’s and other two groups, there seems to be a lack of expertise among professionals and consumers.

2. Ranking: Since all the dimensions and improvement strategies are measured by a scale of 0 to 10 a ranking of dimensions from weakest to strongest could be performed. This is very useful when an organization has limited resources in terms of data quality improvement. In such cases, an organization could concentrate on the weakest dimensions first and, depending on the resources, continue on to the other dimensions. On the other hand, an organization finds the most benefit by following the improvement strategies that have been identified as the most useful.

5. CONCLUSION

In this paper we introduced a new method to measure data quality dimensions in an organization, as well as identifying the proper strategies to improve upon the dimensions. The questionnaire-based method is both simple and general-purpose, so it could be used for most applications. Also the use of three different subjects could lead to interesting “gap-analysis” results. As mentioned in section 1 measurement is measuring the value itself, and assessment is comparison to *reference values*. It will be useful to see how the “assessment” of the questionnaire-based methodology could be performed. Of course, testing the questionnaire-based methodology in real-world organizations will help to identify its shortcomings and improve upon them.

Finally, objections may be raised to the questionnaire-based methodology that it is a purely “subjective” methodology. More specifically, it is a methodology only based on the personal opinions of the subjects being studied. We tried to get around this shortcoming by introducing “three” distinct groups of subjects. This, hopefully, will serve as a cross-check system, so that, if an opinion of one group is mistakenly influenced, the surveys from the other groups will make up for that. However, it will be very useful to introduce an “objective” element to the methodology, or something that does not consider subjects’ opinion (e.g. Formulas for Measurement).

REFERENCES

- [1] BALLOU, D. AND PAZER, H. 1985. Modeling data and process quality in multi-input, multi-output information systems. *Manag. Sci.* 31, 2.
- [2] Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. 2009. Methodologies for data quality assessment and improvement. *ACM Comput. Surv.* 41, 3, Article 16 (July 2009), 52 pages
- [3] W.H. Delone, E.R. McLean, Information systems success: the quest for the dependent variable, *Information systems research* 3 (1), 1992, pp. 60–95.
- [4] D.L. Goodhue, Understanding user evaluations of information systems, *Management Science* 41 (12), 1995, pp. 1827–1844.
- [5] M. Jarke, Y. Vassiliou, Data warehouse quality: a review of the DWQ project, *Proceedings of the Conference on Information Quality*, Cambridge, MA, 1997, pp. 299–313
- [6] LEE, Y.W., STRONG, D. M., KAHN, B. K., AND WANG, R. Y. 2002. AIMQ: A methodology for information quality assessment. *Inform. Manage.* 40, 2, 133–460.
- [7] LONG, J. AND SEKO, C. April 2005. A cyclic-hierarchical method for database data-quality evaluation and improvement. In *Advances in Management Information Systems-Information Quality Monograph (AMISIQ) Monograph*, R. Wang, E. Pierce, S. Madnick, and Fisher C.W.
- [8] Muller, H., Freytag, J.-C., and Leser, U. 2012. Improving data quality by source analysis. *ACM J. Data Inf. Qual.* 2, 4, Article 15 (February 2012), 38 pages.
- [9] SCANNAPIECO, M., A. VIRGILLITO, MARCHETTI, M., MECCELLA, M., AND BALDONI, R. 2004. The DaQuinCIS architecture: a platform for exchanging and improving data quality in Cooperative Information Systems. *Inform. Syst.* 29, 7, 551–582
- [10] Vaziri Yazdi, Reza. Mohsenzadeh, Mehran. 2012. Towards a Practical “State Reconstruction” for Data Quality Methodologies: A Customized List of Dimensions. *Second International Conference on Computer Science, Engineering & Applications (ICCSEA-2012)*
- [11] Y. Wand, R.Y. Wang, Anchoring data quality dimensions in ontological foundations, *Communications of the ACM* 39 (11), 1996, pp. 86–95.

- [12] WANG, R. 1998. A product perspective on total data quality management. Comm. ACM 41, 2.
- [13] WANG, R. AND STRONG, D. 1996. Beyond accuracy: What data quality means to data consumers. J. Manage. Inform. Syst. 12, 4.
- [14] R. Zmud, Concepts, theories and techniques: an empirical investigation of the dimensionality of the concept of information, Decision Sciences 9 (2), 1978, pp. 187–195.

Authors:

Reza Vaziri has obtained his B.S. and M.S. in Computer Science from the University of Pittsburgh, Pennsylvania in 1994 and 1996. Currently he is an Assistant Professor at the Azad University of Iran (Central Tehran Branch), and also a third year PhD student at the Azad University of Iran (Science and Research Branch). He is writing his Ph.D. thesis on data quality under the supervision of the advising professor Dr. Mehran Mohsenzadeh. r.vaziri@iauctb.ac.ir



Mehran Mohsenzadeh has received his Ph.D. in software engineering from Department of Computer engineering, Science and Research Branch, Islamic Azad University of Iran, in 2004. His major interests are Data Integration Data Quality, Data/Web Mining, Software Architecture, Methodologies, Ontological Engineering, and has published more than 50 conference papers and 5 journal papers. He is Assistant Professor of Computer Engineering Department, Science and Research Branch, Islamic Azad University of Iran. mohsenzadeh@srbiau.ac.ir

