Journal of Information Technology Management

ISSN #1042-1319

A Publication of the Association of Management

THE ROLE OF PROCESS METADATA AND DATA QUALITY PERCEPTIONS IN DECISION MAKING: AN EMPIRICAL FRAMEWORK AND INVESTIGATION

G. SHANKARANARAYANAN
BOSTON UNIVERSITY SCHOOL OF MANAGEMENT
gshankar@bu.edu

ADIR EVEN
BOSTON UNIVERSITY SCHOOL OF MANAGEMENT
adir@bu.edu

STEPHANIE WATTS
BOSTON UNIVERSITY SCHOOL OF MANAGEMENT
swatts@bu.edu

ABSTRACT

The quality of the data used in decision-making tasks has important implications for the outcome of these tasks. Data quality researchers have defined various dimensions for measuring data quality, such as accuracy, currency, and completeness. Such measurements are intrinsic to the data itself and do not take into account contextual factors related to the decision-maker or the decision-task. However, recent research suggests that data quality, when assessed by the decision-makers who use it, is not necessarily perceived as intrinsic, but as subjective and context-dependent. This research investigates the provision of process metadata - an abstracted description of how datasets are acquired, processed, stored, and delivered – as a mechanism that affects the end-user's assessment of data quality. In this study we develop a model for understanding the associations of both perceptions of intrinsic data quality and process metadata with the outcome of a data-driven decision task. An exploratory test of the model suggests that both data quality perceptions and the associated process metadata have beneficial effects on outcomes, when mediated by decision-making process efficiency. The model developed in this study and the preliminary empirical results highlight the value of embedding quality and process metadata in computer-supported decision environments to facilitate assessment of data quality.

Keywords: *Decision-making, Information Systems, Data Quality, Decision Support Systems, Database, Metadata*

INTRODUCTION

Data quality is becoming a critical issue in information systems due to the rapid growth of data volumes and their complexity. Poor quality customer data is estimated to cost U.S. businesses $611 billion a year in printing, postage and staff overhead alone [8]. Add to this the potential for capital losses and heightened risk exposure due to poor data quality and it is clear why executives view data quality management as critical to their organizations [42]. Researchers have responded to this issue by developing techniques for improving data quality, such as data cleansing [15], data tracking and statistical process control [30], data source calculus and algebra [28], data stewardship [10], and dimensional gap analysis [17]. These techniques are clearly useful, but they tend to treat the data in isolation from the business context in which it is used. Often they do not take into account important contextual factors such as the task for which the data is used, available time, and individual characteristics of the decision-maker. Such contextual factors have been shown to strongly influence perceptions of data quality [36], [16] – perceptions that affect the decision-making process. Researchers are beginning to take contextual factors [11] and individual differences [43] into account when examining data quality. This paper furthers this research stream, presenting a theory and preliminary evidence associating process metadata and perceptions of data quality with objective decision-making outcomes.

Decisions are made in the context of a particular task, hence understanding contextual evaluation of data quality is particularly important when the data is being used for managerial decision-making. Managerial decision-making tasks are sophisticated and often relatively unstructured [24]. Such tasks are activated by business needs and consist of multiple stages - specifying requirements, gathering information, evaluating alternatives, and formulating decision outcomes [24], [26]. The efficiency and success of managerial decision-making are influenced by a number of organizational and individual factors, one of which is the quality of informational inputs [9], [25], [12]. And since the quality of the data used in decision-making is an important informational input, understanding data quality is critical for analytical, data-driven decision-making processes. With the exception of studies by Chengalur-Smith et al. [7] and Fisher et al. [11], research on the impact of data quality on managerial decision-making lags behind the importance of this phenomenon for organizations.

Metadata is abstracted data about data. Information systems capture and manage different types of metadata such as data dictionary metadata, administrative metadata, and metadata about the system infrastructure [33]. *Quality metadata* is an important component of the metadata layer - a set of quality indicators attached to the data set such as accuracy, currency, and completeness. Quality metadata, also referred to as data tags [38], and data quality information [11], is considered to be intrinsic to the data and context-independent [38], [11]. Prior research has established a link between quality metadata and decision outcomes - providing users with *quality metadata* during the decision process can improve decision outcomes [7], [11].

Another type of metadata that is relevant to data quality assessment is *process metadata* [21], [31]. *Process metadata* describes the processes that created and delivered the data. It captures information such as data sources, processing methods, storage units, and end-usage targets. This study focuses on the role of process metadata in supporting managerial decision-making in context. Process metadata can be macro-level metadata – at the level of the whole dataset – and/or micro-level metadata – at the level of a specific data element within a large dataset. Implemented appropriately, *process metadata* can help users assess data quality in the context of a particular decision task. Providing process metadata to business users can improve their ability to assess data quality and thus enhance decision-making [32].

Data quality practitioners acknowledge the potential benefits of metadata for managerial tasks including decision-support [10], [21]. However, metadata requirements are difficult to capture, and the costs of associated software development and training are high. Further, there are no accepted models for assessing the value and benefits of metadata and hence organizations are unable to quantify its value and find it difficult to justify metadata investments [33]. Our objective in this research is to understand the impact of providing process metadata on decision outcomes as a first step towards justifying the provision of process metadata in decision environments.

We focus on process metadata because intrinsic indicators of data quality do not have the same impact on all decision tasks across contexts. Further, decision makers may find intrinsic assessments insufficient and seek additional, external sources of information concerning the quality of data [32]. Context-independent, quality metadata can enhance decision-making [7], [11]. At the same time, process metadata can affect perceptions of quality metadata, influencing decision-making efficiency and ultimately decision outcomes. This research investigates process metadata as an additional source of value that organizations can provide to their decision-makers to improve decision outcomes.

In this study we present a theoretical framework and use it to examine how data quality assessment and the provision of process metadata affect analytical, data-driven decision-making. Unlike previous studies that have focused on quality metadata alone, a key contribution of this study is that it explores the effects of process-metadata. Following our theoretical assumptions, process metadata can contribute to data-quality assessment in context above-and-beyond the contribution of quality metadata alone. We begin by presenting theoretical background that reviews the literature on decision-making processes, data quality assessment, and metadata. We then propose an exploratory model for examining the impact of process metadata on decision-making. We use this model to develop propositions concerning the role of metadata in data quality assessment and its effects on the decision-making process and outcome. In section 3, we describe an exploratory, non-experimental study conducted as a preliminary empirical examination of this model. In section 4 we discuss the implications of this study, offer our conclusions, and suggest directions for extending this work.

THEORETICAL BACKGROUND AND RESEARCH MODEL

We first review the data driven decision-making literature in order to define the scope of the phenomenon and describe the research model. We then discuss the two relevant types of metadata: quality metadata that represents intrinsic data quality, and process metadata that supports assessment of contextual data quality. The research model links users' perceptions of quality metadata and the usefulness of process metadata to the efficiency of the decision-making process and consequent decision outcomes.

Data Driven Decision-Making Processes

A rich body of literature on managerial decision-making reflects decades of research, both at the individual and organizational levels[1]. Simon established the notion of decision-making in organizations as a complex and often unpredictable process [34]. Decision making, according to Simon, is a complex transformation of inputs – data gathered from internal organizational sources and from the external environment – into outputs. Since the purpose of this study is to understand the effects of process metadata and data quality on decision-making, we focus on processes where input data is provided in the form of a dataset, and the quality of this input data affects the decision-making process and its outcome.

A high-level and widely accepted classification of decision-making processes identifies four basic categories – analytical, judgmental, bargaining and inspirational [37]. Of these four, we focus on analytical decision-making processes because of the importance of data quality to them. Such processes, according to Thompson, are assumed to be of high certainty due to their known cause/effect relationships and desired outcomes. Nutt suggests that often such analytical decision-making processes are data-driven and use data obtained from archives, pilots, or simulations to draw inferences [24], [26]. Data may be presented in various forms, such as database records, documents, or visual illustrations such as graphs. In many cases the decision maker applies quantitative methods and computational approaches during the analysis process. Analytical, data-driven decision-making falls within the broader category of *rational* decision-making. In rational decision-making, decision-makers enter decision situations with known objectives, gather appropriate information, develop a set of alternative actions, and select the optimal one [9]. The selection here of the analytical, data-driven decision making archetype reflects our need to control for potentially confounding characteristics (such as intuitive-choices of the decision-maker, choices driven by politics, or the use of information not obtainable from the data provided) that might obscure the effects of data quality assessment and the availability of associated metadata.

Efficiency and effectiveness are two important measures for assessing organizational decision-making [34]. Effectiveness measures the success of outcomes in terms of how well they resolve the problem in context, while efficiency assesses the process that led to these outcomes. For example, efficiency measures may take into account the ability to evaluate alternatives, recognize constraints and fulfill secondary goals with minimal investment of resources. Researchers have examined factors that impact both the effectiveness and the efficiency of decision-making processes, along with possible synergies between the two. Studies have investigated how the decision task is defined and framed [25], managerial creativity [12], the availability and the performance of decision support technology [22], the characteristics of the task [6], and the skills and motivation of the decision-makers [29. Nutt's findings indicate a positive association between the efficiency of a decision-making process and its outcome [27].

Mackay and Elam specifically examine the interaction between end-users and the decision-making aid provided [20]. Their findings identify the significance of the level of expertise on both the outcome and also on the process that led to the decision. Users who were both task and software experts were able to reach a good solution in a more systematic and direct manner, compared with novice users who tended to spend time on trial-and-error attempts. Speier and Morris found that the design of the decision-making aid has significant effects on the decision outcome, moderated by the complexity of the decision-making task [35]. Text-based aid was found to be more efficient for simple tasks, while complex tasks benefited from the availability of a graphic/visual user interface. Perceptions of quality metadata (discussed in the following section) have also been shown to have a significant effect on the efficiency and the effectiveness of decision-making [7], [11].

The accumulated theoretical and empirical research on decision-making reflects the complexity of the phenomenon. Many factors appear to have significant effects, as do interactions between these factors. The goal of this study is to explore the effect of two specific constructs – data quality perceptions and process

[1] See [9] and [12] for a comprehensive review of the managerial decision-making research

metadata usefulness – on the outcome of a decision task. Efficient decision-making processes enable decision-makers to consider more alternatives, detect constraints correctly, and apply appropriate evaluation mechanisms, which can all lead to better outcomes. Thus decision-making process efficiency plays a central role in determining outcomes, and we take this into account by giving decision process efficiency a central place in our model.

Data Quality and Metadata

Data quality scholars and practitioners define data quality as a multi-dimensional construct [30], [14], [10]. Wang and Strong suggest that business users recognize the multi-dimensionality of data quality and evaluate data elements accordingly [39]. They identify a parsimonious set of data quality dimensions that are perceived by business users as important – *accuracy, relevancy, representation* and *accessibility*. *Accuracy* reflects the extent that the data is error free, correct, and reliable. *Relevancy* indicates how applicable and usable the data is for the task at hand. *Representation* describes the extent that the data is presented in a manner that is easy to understand, and *accessibility* indicates whether the data is secure and cost-effective to access and use. Reliability, believability, currency, and completeness are some other dimensions of data quality that have been discussed in the literature [2], [39], 14], [32].

An important issue with respect to how people evaluate data quality is the distinction between *intrinsic* and *contextual* assessment of data quality. Intrinsic data quality is based on the data elements themselves, independent of the context in which they are used. Examples include the accuracy, representation, and accessibility of the data. The majority of data quality research has focused on intrinsic data quality. However, individuals' perceptions of data quality are influenced by contextual factors such as the decision task in which the data is used, the timing of use, and characteristics of the individual user [36], [16]. Researchers that take into account the context of the decision-making task acknowledge the role played by contextual assessment of data quality. Indeed, one of the most widely accepted definitions of data quality is the fitness of the data for *use* [30].

The need to consider contextual assessment adds another level of complexity to data quality management. With intrinsic assessment, measurement can focus on the dataset alone, making the improvement goals easier to specify (e.g., 99.999% of the data records ought to be accurate). However, intrinsic measurements present only a partial picture of the usefulness of the data to a particular user. Consider a sales report (i.e., showing item codes, quantities, cost and selling prices) where some of the "selling price" values are missing. For decisions regarding the "shelf-placement" of products, a manager would need to know which products have the potential to generate higher profits, and the report with missing "selling price" data would be "incomplete" for this decision task. Yet for making inventory decisions (i.e., reordering, stocking etc.) this report would be "complete" since the "quantity" data is available for all products. This example illustrates the contextual, task-dependent nature of data quality and suggests that a better understanding of contextual data quality could improve data quality management methods.

Campbell's view of task complexity is useful for understanding contextual data quality [6]. Campbell identifies three research perspectives on complexity as it pertains to information processing: the *psychological* perspective looks at subjective aspects of the task such as enrichment, challenge, and stimulation. The *objective* perspective looks at characteristics intrinsic to the task itself, such as information load and constraints to be satisfied. Finally, the *person-task interaction* perspective investigates the role played by factors relating to both the person and to the particular task, such as the difficulty of the task for the person undertaking it, the person's experience and familiarity with it, and their motivation to perform it. Analogously, data quality assessment in the context of managerial decision-making has objective, psychological, and task-person interaction components. *Intrinsic* data quality invokes objective, invariant aspects of the data, whereas viewing data quality in the context of a particular task brings to bear a task-person interaction perspective to it. For example, when two different individuals assess the accuracy of the same data element differently, this reflects differences in their respective psychologies. When an individual views a data element as more relevant for the current task than for a different task, this exemplifies the task-person interaction perspective. Our interest in this study is to understand how process metadata affects the perceptions of quality metadata and vice versa and how, together, these are associated with the decision process and decision outcome. In this study we do not examine or manipulate objective or psychological components per-se, but focus rather on the person-task interaction perspective to explore how this interaction affects the efficiency and outcomes of data-driven decision-making.

For large complex datasets, the human ability to detect data quality problems is limited. In such cases, data quantity and complexity exceed the bounded rationality of users, who can consequently benefit from the provision of data quality metadata [18]. Helpful abstractions can take the form of intrinsic data quality measurements of the data, based on, among other things, the number of records, error rate, count of missing values, or the time/date of last update. As discussed earlier, such intrinsic measurements are limited in their usefulness since decision-makers must gauge data quality within the context of the particular task and these measurements do not aid in this. So far, information systems have failed to provide adequate support for assessing contextual data quality of large and complex datasets.

Process metadata is explored here as a means for helping decision-makers gauge data quality in context. It accomplishes this by providing information about the processes that were used to generate and deliver the data to the decision-maker. Process modeling and documentation are not new concepts in information systems. Information systems professionals commonly use such techniques for the design and on-going maintenance of data processing and delivery systems [30], [3]. Such process abstraction is now recognized as a special form of metadata [21], [31], [33]. One graphical representation that communicates process metadata, and the one used in the empirical investigation described here, is called an Information Product Map (IPMAP) [31]. The IPMAP is a modeling technique based on managing information as a product [40]. It allows the decision-maker to visualize the flow of data elements and the sequence in which they were processed. Visualization is done by a set of constructs that represent different stages of the data manufacturing process – data source, processing, storage, quality inspection, organizational or IS boundary, and data consumer. Each construct is supplemented with metadata about the stage it represents, such as a unique identifier, the composition of the data at that stage, ownership, processing requirements, and physical location where the step is performed. In this research we have used the IPMAP to communicate micro-level[2] process metadata. While clearly process metadata may be communicated in many different ways, in this study we have chosen to use the IPMAP to communicate process metadata.

Figure 1 shows an example of an IPMAP. The information product (IP) described is an exposure report prepared for an advertising director. Regional data is used to compute the exposure effectiveness (EE) for that region (e.g., "Compute East EE"). Regional evaluations are then checked for quality and combined to provide data that reports the overall exposure effectiveness, which is stored in an "Exposure" database. The exposure report is generated by another process that reads data from this database and distributed to the consumer of the report.

We illustrate, using a brief example, how the process metadata in an IPMAP can change users' perceptions of the quality of the data. Consider the case where the quality metadata associated with the data collected from the Northeast indicates that the data is not of exceptionally high quality but is not very low quality either (i.e. has a quality level of 70% aggregated over all relevant data quality dimensions). By examining the IPMAP the decision-maker can see that the data was collected, for example, by a market survey agency that the decision-maker trusts and whose data he has used in the past with success. Further, the IPMAP reveals that the data was cleansed using a specific algorithm that has a proven track-record. These two pieces of process metadata obtained from the IPMAP may convince the decision-maker to weight the data from the Northeast higher than he might have without the process metadata.

Processing data prior to delivery to a user can include many stages. In such cases the resulting visual representation is complex – potentially too complex to be useful to non-technical users. One of the goals of this exploratory study is to investigate the effects of process metadata on non-technical users in a visual form - the IPMAP representation in this case. We are interested in understanding the extent that decision makers find visualized process metadata to be useful. And to the extent that they do, how this affects the efficiency and outcome of the decision making process. In particular, we want to explore the impact of process metadata on the context-dependent evaluation of data quality. While intrinsic assessment is derived from the granular details of the dataset contents, process metadata adds an extrinsic layer to the assessment process, locating the data in the context of the processes that produced it. For this reason, we expect it to affect users' perceptions and assessment of data quality differently than data quality metadata. This is not to say that quality metadata is unimportant – it is a significant informant of quality. However, combining it with process metadata can provide the business user with a more comprehensive picture of quality and increase the user's confidence on the data and its quality. This leads the user to believe that he will perform well in the decision task. Social Cognitive Theory suggests that when a user is confident of doing well, on an average, the user does perform well [4]. Thus, process metadata improves the decision process efficiency of the user and consequently the decision outcome. Our model below suggests that both process metadata and contextual quality assessments may have significant positive effects

[2] In this paper we define micro-level is metadata related to individual data elements in a dataset as opposed to macro-level metadata which is metadata about the dataset as a whole.

on decision-making outcomes, mediated through their effect on the efficiency of the decision-making process.

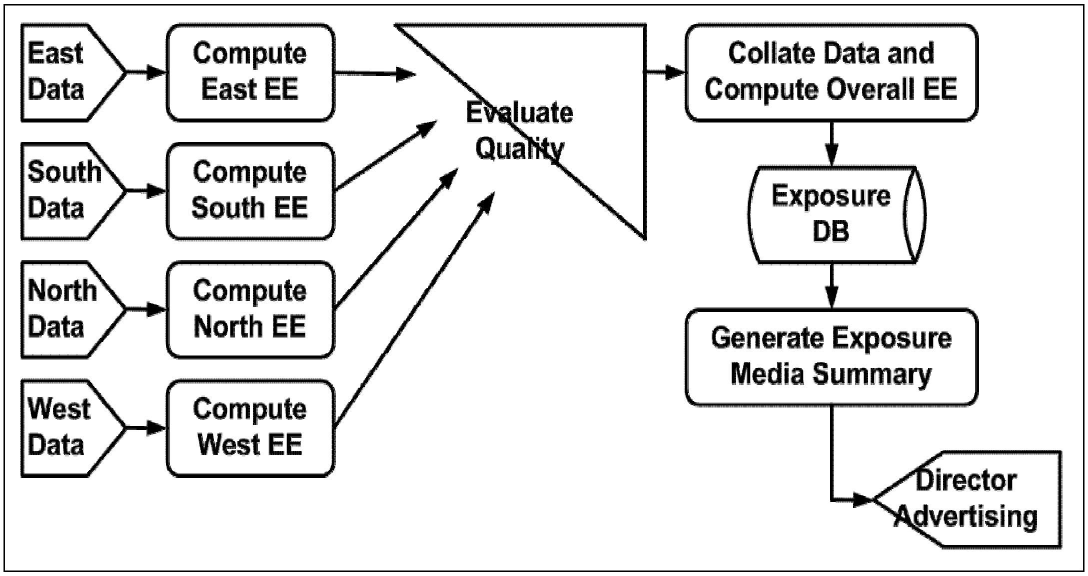

Figure 1 - Sample Information Product Map (IPMAP)

The Research Model

The model presented in Figure 2 incorporates the efficiency of the decision-making process, data quality assessment and the perceived usefulness of process metadata into a framework for understanding how metadata can influence decision outcomes.

The model posits that perceptions of the usefulness of process metadata, and assessments of data quality, both have a significant effect on the decision outcome. Prior research found that intrinsic assessment, in the form of quality metadata, directly influences decision outcomes, as reflected in the model [7], [11]. However, these studies do not consider the role of process metadata in the data-driven decision-making. Thus a major contribution of the model is the addition of perceptions of process metadata.

The other major contribution of the model is the proposed mediation of outcomes by the efficiency of the decision-making process. The logic underlying this mediation effect is as follows. Organizations can improve the decision-making processes by complementing data quality assessment with the provision of process metadata. Both data quality metadata and process metadata support the capability of the decision-maker to perform their decision-making processes efficiently, and it is this consequent efficiency that drives positive outcomes. Efficient decision-making processes enable decision-makers to consider more complex and larger data sets, support the exploration of more alternatives, detect constraints properly and apply correct evaluation mechanisms. Ceteris paribus, support for these capabilities increase the likelihood that an optimal outcome will be identified. In this way, the more efficient the decision-making process, the better the expected outcome of that process.

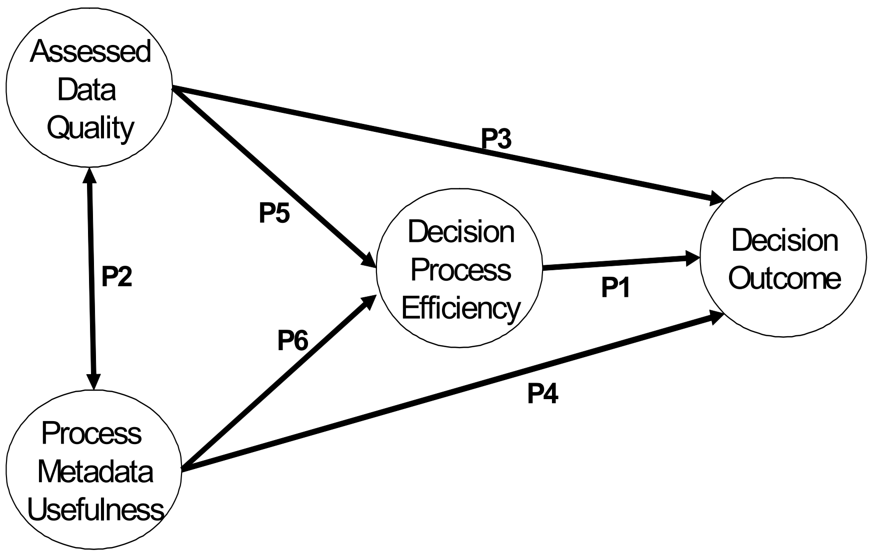

Figure 2 - The effect of quality and process metadata in a decision-making process

Proposition 1(*P1 in figure 2) – Decision-making process efficiency is positively associated with decision outcomes.*

The next proposition investigates the relationship between assessed data quality and the perceived usefulness of process metadata. Data quality is assessed on the basis of actual data values provided to the user, addressing whether or not the data meets acceptable quality standards. The process metadata describes factors that are external to the dataset – such as the data source and the manipulations that were applied while delivering it. Process metadata is at a higher level of abstraction than quality metadata, since it provides details about the processes used to create the data. Whereas quality metadata directly describes the data delivered, process metadata describes the processes used to deliver it. We suggest that this higher level of abstraction will be most useful when the quality of data is ambiguous, motivating the user to seek additional information to resolve this ambiguity. Consequently we expect that data quality assessment will affect the perceived usefulness of process metadata. For example, in cases where the data quality assessment indicates that the data is of very poor quality, the user will have an unambiguous notion of the quality of the data and have little need for additional quality information in the form of process metadata. On the other hand, where the results of data quality assessment are ambiguous or positive, users will be most likely to find the process metadata useful, either for confirming the positive quality assessment or for better understanding it. Thus we expect to find that the usefulness of process metadata is associated with perceptions of data quality.

Proposition 2 *(P2 in figure 2) – Assessed data quality is positively associated with the perceived usefulness of process metadata*

We next examine the effects of data quality assessment and process metadata on the decision-making outcome. Higher quality data reduces the likelihood of errors, hence is likely to improve the decision-making outcome. Process metadata serves to confirm or disconfirm quality metadata and affects the assessment of data quality. The more useful the process metadata, the more information available to inform the decision, with consequent increased potential for achieving successful decision outcomes**.** Hence:

Proposition 3 *(P3 in figure 2) – Assessed data quality is positively associated with decision outcomes.*

Proposition 4 *(P4 in figure 2) – Perceived usefulness of process metadata is positively associated with decision outcomes.*

Finally, we suggest that the effects of both assessed data quality and the usefulness of the process metadata on the decision outcome will be mediated by the perceived efficiency of the decision making process. As discussed above, we expect the efficiency of the entire decision-making process to mediate the effect of quality assessment.

> ***Proposition 5*** *(P5 in figure 2) – The efficiency of the decision-making process mediates the association of assessed data quality on decision outcomes.*

> ***Proposition 6*** *(P6 in figure 2) – The efficiency of the decision-making process mediates the association of the usefulness of process metadata on decision outcomes.*

Our premise in this research is that there is no clear precedence relationship between perceptions of process and quality metadata, and the efficiency of the decision process. In absence of prior theory suggesting a temporal ordering of these factors, we suggest that these occur concurrently and recursively.

Our claim in P5 is that the assessed data quality affects decision outcomes and this effect is mediated by the decision process. Likewise, P6 suggests that process metadata affects decision outcomes and that this effect is mediated by the decision process. Note that hypotheses P5 and P6 do not presume causal relationships, but mediation effects. In our model, we are not testing for causality. Further, we are exploring the possibility that the synergistic effect of the two types of metadata is greater than the individual effects of either type of metadata. We hypothesize that this synergistic effect manifests in a superior decision process and consequently better decision outcomes.

EXPLORATORY STUDY

The study described in this section is a preliminary attempt to empirically assess the concept of process metadata and its usefulness to the decision maker. Since the effects of process metadata have not been previously tested empirically, this study is presented as exploratory and non-experimental. It seeks to assess the constructs in the theoretical model, illustrate a method to conceptualize those constructs, and begins to validate the model's propositions. A full-scale test of the model would require an experimental design based on a robust conceptualization of the model – a conceptualization that this exploratory study begins to develop.

Research tool, Task, Participants, and Procedures

The effect of process metadata on decision outcomes has not been explored before and introduces significant empirical challenges, particularly with the operationalization of the model constructs. A subjective operationalization (e.g., in a form of a questionnaire) is vulnerable to method bias, while objective operationalization is difficult to implement, particularly for the "usefulness" and "efficiency" constructs. Therefore the research tool (Figure 3), utilizing macros within MS Excel, attempts to combine both objective and subjective aspects of the model. The tool implements a marketing decision task of allocating an advertising budget across multiple types of media – Billboards, Magazines, Radio, TV – and geographical locations – East, North, South, West – given a fixed budget. The allocation aims to maximize the expected number of people exposed to the campaign, based upon past exposure efficiency (i.e. the number of people exposed per advertising dollar spent). To support this decision task, spreadsheet data was provided indicating the estimated number of people who would be exposed to each type of media within each geographical region. The data also provided estimations of exposure efficiency history, calculated as the average number of people exposed to the product per dollar spent on advertising. However, given the possibility of poor quality data, the subjects were advised to consider the data quality when allocating the budget. The performance of the participants in the decision task was measured by calculating a geometric-average of two of their scores: the first score consisted of the estimated number of people exposed (E) to the advertisement campaign given the budget allocation – the final solution to the assigned task. The second score reflected the quality of this solution (Q), based on the extent that high quality data elements were used to calculate it.

The input data, including past information on exposure efficiency, was provided at both an aggregated level (Figure 3) and at a detailed level of granularity (Figure 4) for all participants. In addition to the actual data, users were provided with metadata to enable data quality assessment. One category of metadata provided was a set of information product maps (IPMAP), linked through the main screen. One set of participants received an IPMAP at a higher level of representation detail (low process metadata) and another set were given access to a more detailed IPMAP that was hyper-linked to the first IPMAP. However, this manipulation of the level of process metadata was removed from subsequent data analyses due to the lack of variance it generated. This is not surprising considering it is a new operationalization of an exploratory construct. A pilot study was conducted using ten doctoral students who did not participate in the final study. Unfortunately, this limitation was not apparent in the pilot study, possibly due to the small sample size.

A second metadata element was a quality assessment of the input data, pre-evaluated for accuracy, completeness, currency, consistency and relevance. This quality metadata was presented to the users in a visual "traffic light" format – pieces of information with high quality were highlighted in green, medium quality in orange and low quality in red. Similar to the input data, quality assessment was made available at two levels of granularity – aggregated, on the main screen (Figure 3), and detailed (Figure 4), available through hyperlinks. All participants received both the aggregated and detailed quality metadata.

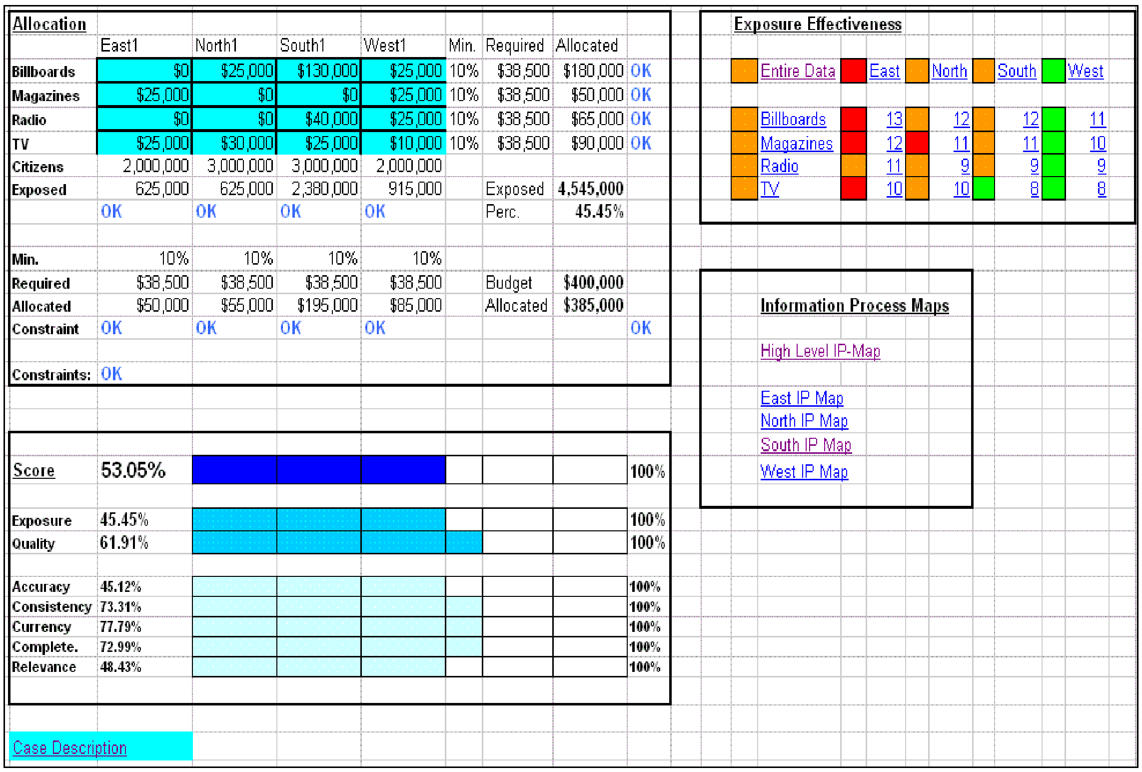

Figure 3: Research tool – Main Screen

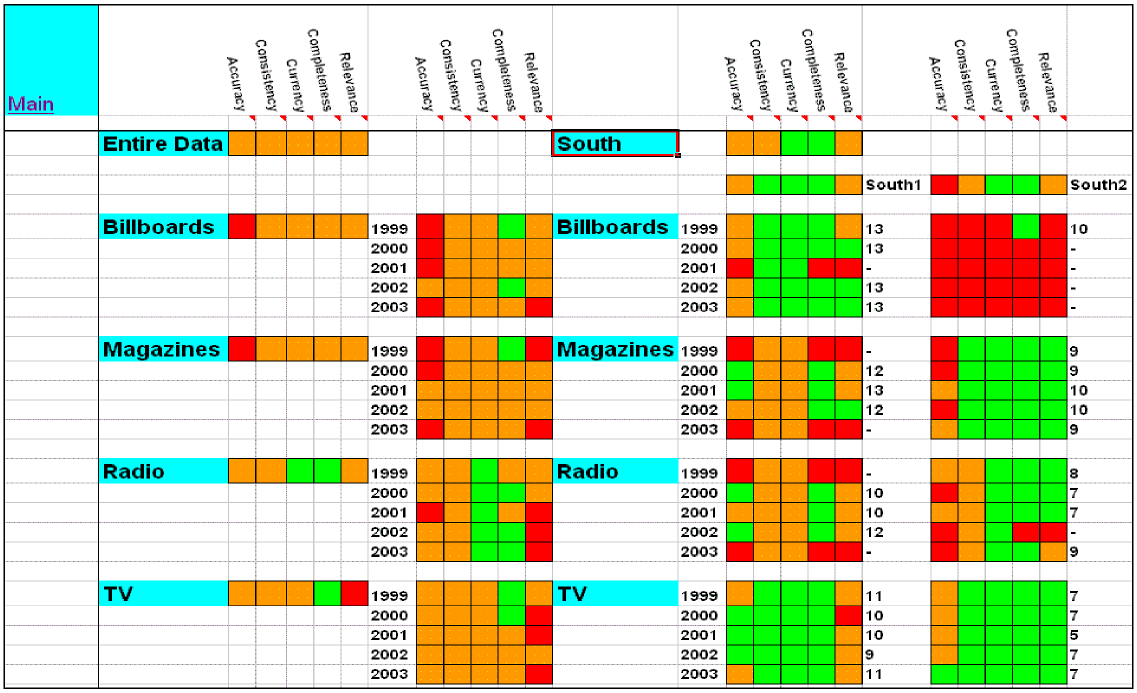

Figure 4: Detailed Data and Quality Metadata

The Excel-based decision support tool allowed the participants to explore different budget allocations using a "what if" style to get immediate feedback on the overall expected level of exposure (E) and an indication of the decision quality based upon the underlying data (Q). The final outcome score (S) was calculated as a geometrical-average ($S=\sqrt{E*Q}$). While the user navigated through the different screens, a hidden back-end process tracked the navigation and the time-spent on each screen. The progress with budget allocation as well as the use of different metadata elements was recorded to a database with an accurate time stamp, that later enabled an insightful analysis of the users' work process.

Fifty-one masters-level information systems graduate students participated in a computer-based decision-making task. All of these students have organizational work experience. Participants were all assigned the same computer-supported, data driven, decision-making task described earlier. Participants were given an initial overview of the task by the experiment coordinator. Included in this overview were the details of how the final score was to be computed. Participants were then given 20 minutes to complete the task and were directed to allocate the budget so as to maximize the final score. At all times during the task the spreadsheet interface provided participants' with their composite performance score of current allocation and associated data quality. Participants were not aware of the progress or score of the other participants. In order to motivate them to work to achieve an optimal decision outcome, the task was conducted as a competition in which cash prizes were offered to the four participants that obtained the highest final scores. Participants completed a survey instrument after finishing the task. The survey consisted of measures of the model constructs, discussed below. Participants were assigned a secure code to ensure their anonymity, and their demographics were reported as part of the final survey.

Measurement Model and Demographics

To assess the extent that perceptions of metadata are associated with the decision process and outcome, the users were provided with a survey. The survey included previously validated items that had been used to measure constructs similar to those investigated in this study. It also included items that were created for this study where no previously validated ones were available. These items were developed by one author, revised by the others and then pre-tested on a sample of ten graduate students who did not participate in the final test. The dependent measure consisted of the composite score discussed above, calculated by weighting the raw decision outcome score with the intrinsic data quality used to calculate it.

For quantification purposes these composite scores were linearly rescaled by setting the maximum score obtained by the top scoring participant to one and linearly adjusting all other scores to this scale.

To test perceptions of intrinsic data quality, users were asked to assess data quality attributes as they reflected the dataset. Bailey and Pearson offer a large set of quality attributes for assessing user perception of information systems and their outputs, which we adopted as a measurement tool for this study [1]. However, their instrument does not differentiate between intrinsic and contextual assessment. Wang and Strong [39], and Gendron et al. [13] offer a classification of data quality attributes that differentiates intrinsic from contextual. The four intrinsic characteristics that are defined by these studies are believability, accuracy, objectivity, and reputation. Thus we included items from Bailey and Pearson's instrument in the survey that map to these intrinsic data quality characteristics – reliable, accurate, reasonable, helpful and useful. Each dimension was introduced to the users as a separate 7-point Likert scale question. The results indicate relatively high consistency among the questions (Cronbach alphas of 0.709). Hence the results were averaged into one a single measure.

The perception of process metadata usefulness, as provided in the form of IPMAP, was measured using two 7-point Likert scale items:

- *The IP Map was helpful for understanding data quality*
- *I had no problem understanding the information provided in the IP Map*

In order to measure perceived process efficiency, the following five 7-point Likert-scale survey questions were developed and validated as discussed above:

- *To what extent were you able to work quickly as you did this exercise?*
- *The time allocated to the task was sufficient*
- *This exercise was easy for me*
- *The process I used to solve this exercise was a very efficient one*
- *I found this exercise to be very difficult*

The independent constructs show acceptable internal reliability and discriminant validity. All Cronbach's alphas for the tested constructs are within acceptable bounds (> 0.7) for an exploratory research, while all Cronbach alphas between constructs were less than 0.5.

Navigation analysis was used to operationalize the process efficiency construct objectively by measuring the actual metadata use, and so provided the outcome score measurement (linearly rescaled to a 0-1 range, for convenience). In this way a proxy for decision-making efficiency was obtained by tracking the score progression (Figure 5). While some users showed systematic process by consistently improving their optimal score, others showed inconsistency and no significant progression. Inconsistency (F) was measured as the average absolute residuals around a regression line that represents score progress over time:

$$F_k = \frac{1}{N_k}\sum_{i-1}^{N_k}\left|Y_{k,i} - \alpha_k * T_{k,i} - \beta_k\right|,\ \text{where:}$$

F_k – Average absolute residual for user [k]
N_k – The number of data points for user [k]
α_k, β_k – The slope and the constant coefficients of the linear regression for user (k)
$T_{k,i}$ – Time point [i] of user [k]
$Y_{k,i}$ – Score obtained at time [i] of user [k]

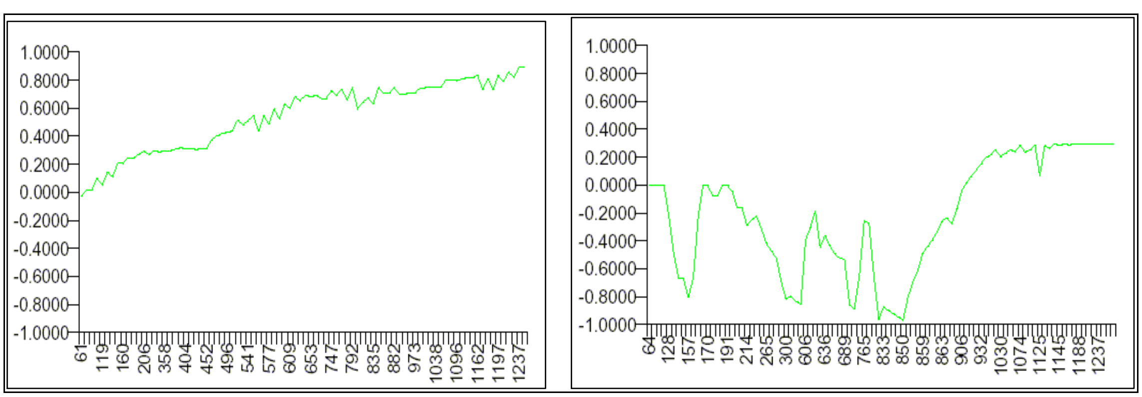

Figure 5: Tracking Efficient versus Inefficient Progress (score) over time

Participants consisted of 35 males and 16 females, reflecting a gender distribution typical of graduates programs in management. Participants' mean age was 27.6, with standard deviation of 2.6. The mean of their years of work experience was 4.8, with a standard deviation of 2.2. There were no significant differences among constructs due to gender, age or years of work experience.

To gauge mediation effects (suggested by hypotheses P5 and P6), the model in figure 2 was tested using path analysis by examining each half of the model separately. Standardized path coefficients for the model shown in figures 6a and 6b were first derived using a series of OLS regressions ($O = \beta_{11} * Q + \beta_{12} * E + \in$; $E = \beta_{21} * Q + \in$ shown in figure 6a and $O = \beta_{31} * P + \beta_{32} * E + \in$; $E = \beta_{41} * P + \in$ shown in figure 6b). The partial direct effect of assessed data quality ($\beta_{11} = 0.201$ in 6a) and process metadata usefulness ($\beta_{31} = 0.201$ in 6b) on decision outcome are small, but significantly non-zero in the presence of decision process efficiency. The path coefficients of assessed data quality to decision outcome when assessed without decision process efficiency ($\beta = 0.356$) and process metadata usefulness to decision outcome when assessed without decision process efficiency ($\beta = 0.371$) are significantly larger than the path coefficients between the same two and decision outcome (β_{11}, $\beta_{31} = 0.201$ respectively) in the presence of decision process efficiency. Further, the path coefficients of the paths between decision process efficiency and decision outcome ($\beta_{12} = 0.545$ in 6a and $\beta_{32} = 0.529$ in 6b) is significantly greater than zero in both cases. This preliminary examination suggests that decision process efficiency may partially mediate the effects of both assessed data quality and process metadata usefulness on decision outcome. This analysis also suggests that the direct effects of assessed data quality and process metadata usefulness on decision outcome do exist and that the mediation by decision process efficiency is only partial.

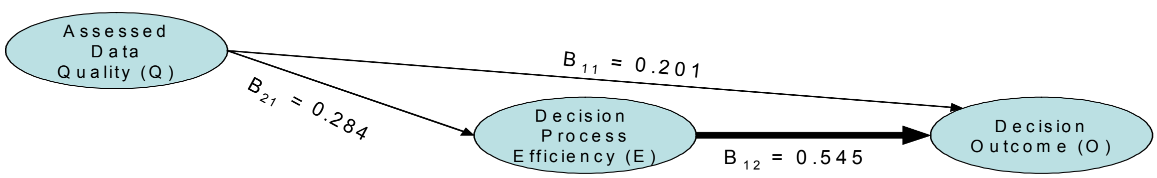

Figure 6a: Path model with standardized path coefficients for Q and E on Outcome

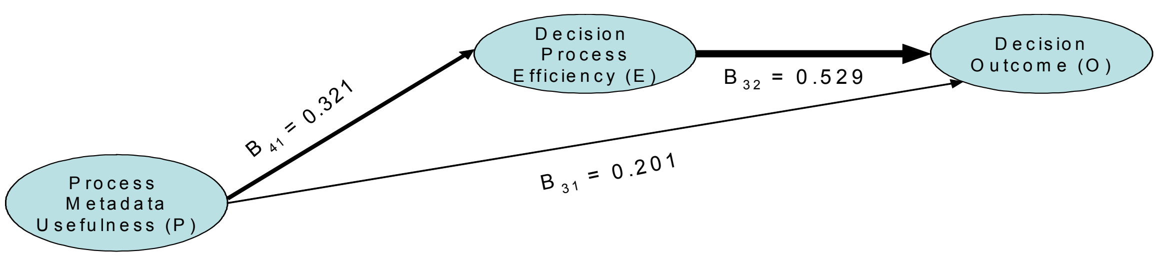

Figure 6b: Path model with standardized path coefficients for P and E on Outcome

Table 1 – Means, Standard Deviations and Correlations of the tested variables

Variables	Mean	STD	Items	Cronbach Alpha	1	2	3	4
1. Intrinsic Data Quality	4.53	0.88	5	0.709		0.35*	0.28*	0.36*
2. Process Metadata Usefulness	3.42	1.62	2	0.737	0.35*		0.32*	0.37**
3. Decision-Making Process Efficiency	4.00	1.22	5	0.760	0.28*	0.32*		0.59**
4. Decision-Making Outcome	0.55	0.04	1	N/A	0.36*	0.37**	0.59**	

n = 51
* p < 0.05 (2-tailed)
** p < 0.01 (2-tailed)

Study Results

Due to the relatively small sample size, OLS regression analyses were used to assess the propositions. Results of these analyses are shown in table 1. For all propositions discussed below, regression slopes were positive. We began by testing the effect of decision-making process efficiency on the decision outcome (P1) by regressing the final performance score onto process efficiency. Results were highly significant with an adjusted r^2 of 0.34, F=26.68 (d.f. = 1, 49; $p < 0.001$). We then investigated the association between assessment of data quality and the usefulness of the process metadata. To test proposition P2, we regressed usefulness of process metadata onto assessed data quality. Results were highly significant with an adjusted r^2 of 0.085, F=5.62 (d. f. = 1, 48; $p < 0.05$). To test P3, the final outcome score was regressed onto assessed data quality. Results were again significant with an adjusted r^2 of 0.108, F=6.946 (d. f. = 1, 48; $p < 0.05$). Similarly, to test P4, the final outcome score was regressed onto perceptions of process metadata usefulness. Results were again significant, with an adjusted r^2 of 0.12, F=7.823 (d. f. = 1, 48; $p < 0.01$).

To test for and confirm mediation effects (P5 and P6) we took a typical four-step analytical approach [5]. This necessitates showing that (a) the independent variable has a direct effect on the dependent variable, that (b) the mediator has a direct effect on the dependent variable, that (c) the independent variable has a direct effect on the mediator and, finally, showing that, (d) in the presence of the mediator, the effect of the independent variable on the dependent variable becomes insignificant. For proposition P5, the independent variable is the perception of quality metadata or assessed data quality, the dependent variable is the performance score, and the mediator is process efficiency. Step (a) has been confirmed for P3 and step (b) has been confirmed for P1. For step (c), perception of process efficiency was regressed onto assessed data quality. Results were significant, with an adjusted r^2 of 0.061, F=4.196 (d. f. = 1, 48; $p < 0.05$). For step (d), the performance score was regressed onto perceptions of both assessed data quality and decision-making process efficiency as independent variables. Using a stepwise regression, process efficiency was retained in the model while assessed data quality was excluded. The final model, excluding assessed data quality, results in an adjusted r^2 of 0.35, F=27.30 (d. f. = 1, 48; $p < 0.001$). Assessed data quality was excluded on the basis of t = 1.705 and $p = 0.095$. All four steps were thus tested, revealing a significant positive effect mediation effect of assessed data quality on the decision-making outcome, confirming P5. The same process was used to test proposition P6. In this case the independent variable was the perception of process metadata usefulness, the dependent variable was again the performance score, and the mediator was again process efficiency. Step (a) was confirmed for P4 and step (b) was been confirmed for P1. For step (c), perception of process efficiency was regressed onto the perception of process metadata usefulness. Results were significant, with an adjusted r^2 of 0.085, F=5.623 (d. f. = 1, 48; $p < 0.05$). For step (d), the performance score was regressed onto perceptions of both process metadata usefulness and process efficiency as independent variables. Using a stepwise regression, process efficiency was retained in the model while process metadata usefulness was excluded. The final model, excluding process metadata usefulness, results in an adjusted r^2 of .33, F=26.676 (d. f. = 1, 48; $p < 0.001$). Process metadata usefulness was excluded on the basis of t = 1.689, $p = 0.098$. Thus all four steps were tested and decision process efficiency was found to significantly mediate the effect of process metadata usefulness on the decision outcome, confirming proposition P6. These results confirm the preliminary observations from the path analysis that the mediation is partial and that the direct effects of both assessed data quality and process metadata on decision outcome are still significant. Table 2 summarizes the results of all the propositions tested in this study.

Table 2 – Summary of Propositions Tested

P.	Independent Variables	Dependent Variable	Adj. R^2	F	P
P1	Decision Making Process Efficiency	Decision Making Outcome	0.34	26.7	P < 0.001
P2	Assessed Data Quality	Process Metadata Usefulness	0.09	5.62	P < 0.05
P3	Assessed Data Quality	Decision Making Outcome	0.11	6.95	P < 0.05
P4	Process Metadata Usefulness	Decision Making Outcome	0.12	7.8	P < 0.01
P5	Decision Process Efficiency mediates Assessed Data Quality	Decision Making Outcome	0.35	27.3	P < 0.001
P6	Decision Process Efficiency mediates Process Metadata Usefulness	Decision Making Outcome	0.33	26.7	P < 0.001

These exploratory test results provide preliminary validation of the proposed theoretical model and encourage further exploration. Even with the small sample size and relatively homogenous participants, variability in the tested constructs was detected and many of the propositions were supported with high statistical significance. The propositions suggested by the model were all supported - some to a greater extent (P1, P5 and P6) and others to a lesser extent (P2, P3, and P4). Overall, the results suggest that perception of good data quality and process metadata do contribute to improved decision-making process efficiency and consequent outcomes. It should be pointed out, however, that the adjusted R-Squares obtained are relatively small, suggesting that while quality assessment and process metadata had positive effects on decision-making outcomes in this test, these were certainly not the only influential factors. This result is not surprising – literature suggests that decision making processes are complex and are likely to be influenced by a large number of organizational, cognitive and technological factors.

In this study, we tested for both direct effects of process metadata on decision outcomes and indirect effects in the form of mediation by process efficiency. The results of the study indicate that these indirect mediation effects are far more significant than the direct effects. Similarly, we also tested for the direct and indirect effects of assessed data quality on decision outcome. These results were similar, with the indirect effects being more significant than direct effects.

Although we collected objective measurements of time spent on the task as a proxy for decision-process efficiency, the analyses did not indicate significant effects of time spent. Also, though we manipulated the level of detail of process metadata (aggregated process metadata versus granular process metadata) the results did not indicate any significant effect associated with this manipulation. When the "extra" process metadata was available, a human decision-maker, as a "cognitive miser", was hesitant to make the additional cognitive effort required to integrate the metadata into the decision making processes and possibly ignored it. Superior presentation and visualization can alleviate the information overload and must be examined.

While this preliminary validation appears positive, we acknowledge some limitations of the experiment. The use of a laboratory setting permitted us to control for noise but prevented us from replicating a real-life decision-making scenario. Participants were all graduate students who were similar in age and were pursuing the same graduate degree in management. The decision-making process investigated here focuses on a specific domain – marketing; a specific type of input data – numeric; and a specific methodology for obtaining a solution – optimization within a given set of constraints. The results of this study can only be generalized to similar decision scenarios that can be mapped to this context. The test utilized a relatively small sample size hence the ability to apply advanced statistical analysis, such as structural equation modeling, is limited. The process metadata usefulness construct is conceptualized with two items – future tests should utilize a more robust measure or an objective measurement of actual use. Similarly, future research should consider objective assessment of decision-making process efficiency (e.g., time to complete, or consistent progress toward solution) rather than subjective perception. For these reasons then, we present our theoretical model as the contribution of this research, and the validation of it as preliminary only.

IMPLICATIONS AND CONCLUSION

This study was conducted to investigate data quality assessment in decision-making and the role of metadata in this process. High quality input data is critical for managerial decision-making. Users need to be able to gauge the quality of their input data for each decision context they face. To that end, we have proposed a model that incorporates quality assessment and process metadata. We have also presented a preliminary empirical validation suggesting that these factors can affect objective decision-making outcomes.

In the test performed, decision process efficiency was shown to have a significant association with performance outcomes, confirming the results of prior research linking the efficiency of the decision making process to its outcome [27]. The mediation effects detected in this study have not been previously identified and contribute to our understanding of the data quality assessment process. They support the notion that the decision-making process is a primary determinant of outcome, and that factors that affect this process will ultimately affect its outcome. Since the effects of metadata are indirect, these mediation findings align with the view that the merits of data quality assessment and associated metadata are not easy to detect or to quantify. They may also explain why previous researchers [7], [11] in this area have had difficulty identifying direct effects of metadata on decision-making outcomes.

Results support the hypothesized positive association between assessed data quality and the perceived usefulness of the process metadata. One explanation of this phenomenon is that process metadata is complementary to quality metadata, providing additional information about the sense of quality achieved through the assessment of data quality. When users assess data as being of poor quality on the basis of intrinsic indicators, the addition of supporting extrinsic information is redundant – the data has already been shown to be unacceptable for use in that task. On the other hand, when users assess the data quality to be high or even moderately high, process metadata can help to confirm (or not) this level of quality by enriching the context with metadata about its production process. Further, the process metadata may provide additional insights that can explain assessed data quality values. These insights in turn can help decision-makers better gauge the quality of the data they are using in the context of the particular decision-task. We did not explore the mechanisms underlying how decision-makers integrate their understanding of data quality with the decision context to gauge data quality *in situ.* A better understanding of this synergy, as reflected in the association between the two types of metadata, presents an interesting avenue for further research.

Despite the empirical limitations of the study, its results have important implications for future research on data quality management and for the design of decision-support environments. Large data volumes, widely distributed data sources and multiple stakeholders (i.e., data providers and data consumers) characterize current organizational settings. Mobile and wireless technologies have increased these volumes, further distributing data sources while permitting access to data anywhere, anytime. Such environments empower and necessitate decision-makers to react more quickly to events that demand decision-making, including mission-critical ones. Decision-support in such environments demands efficient data quality management. Our findings support the common-sense notion that when participants perceive good input data quality, their sense of decision-making efficiency as well as their decision outcomes are improved. This, pending further corroboration, has clear implications for user-interface design and the importance of communicating data quality to end-users. Having a sense of the quality of their data seems to enable users' to utilize the data more efficiently and effectively.

An association between assessed data quality and the usefulness of process metadata also has implications for the design of decision support systems. Providing the decision maker with intrinsic data quality measurements and/or providing tools to gauge the same is important but insufficient. Since the relevant data may come from multiple sources that span organizational and business boundaries, it is helpful to be able to gauge these sources and the processes used by them to create and transfer the data. Providing an additional layer of information about the data manufacturing process complements intrinsic quality assessment and enhances the efficiency of the decision process.

A related issue is the presentation of intrinsic quality metadata and how decision-makers absorb this information. It was interesting to note the high correlation among indicators of quality along the different quality attributes included in the survey. In this study, participants were not able to distinguish significantly between these attributes. This contradicts the widely accepted precept of the multi-dimensionality of perceived data quality. In prior research, data quality has been represented as a set of attributes [3], [30]. Such representation is useful because it enables information system practitioners to identify specific areas for quality improvement [14]. Other findings support the notion that multi-dimensional presentation is beneficial to business decision-makers, by showing that they can distinguish between different aspects of quality [39]. Our results indicate that, in the context of this experiment, users could not clearly distinguish between the various data quality dimensions. Since these results contradict previous findings regarding multi-dimensional quality perceptions, these results suggest that the topic needs further investigation.

Findings from this study support our assertion that the metadata layer can serve as a tool for communicating data quality to business users. The provision of both types of metadata proved to have a significant effect on both perceptions of decision process efficiency and final outcomes, for this task. However, there are many different presentation formats for communicating both quality and process metadata. This research is an initial foray into understanding how indicators of quality can aid organizational decision-makers. We need to understand which presentation formats are most helpful for communicating which types of metadata. For example, we need to investigate whether indicating intrinsic quality using each relevant dimension (e.g., accuracy, completeness, timeliness) is preferable to indicating intrinsic quality using a single aggregated value – an overall quality indicator. We also need to explore which aspects of the process metadata are most useful: Is the entire IPMAP useful or are there particular parts of it that are more useful than others? These are only two of the many issues that this stream of research needs to address.

This study focuses on the effects of quality assessment and process metadata, but does not rule out other factors that could also have affected the perceived efficiency of this decision process, for example users' experience with similar tasks, level of comfort with the user-interface, familiarity with spreadsheet software, or individual motivation. Thus while results indicate that perceptions of data quality and usefulness of process metadata significantly affect decision-making process efficiency, we cannot eliminate the possibility that other factors may be operating here. We argue that perceptions of data quality and the usefulness of process metadata are transformed into an efficient decision-making process that in turn results in high performance outcomes, but this study was not designed to explain the micro-mechanisms underlying this transformation process. Additional research is necessary to understand this transformation.

Other important questions concern the nature of the decision-making task – which tasks are most amenable to metadata support? How does the nature of the task determine which presentation formats are most effective for communicating metadata information? Finally, we need to understand the way that the two types of metadata interact with each other and the extent to which this is a function of the particular presentation format used.

Prior research on process metadata examines it from the perspective of a technological solution for data quality problems. For example, it can help detect the root causes of quality failures, improve the design and monitoring of data processing and enhance capabilities for error detection and correction. However, these issues are all the concern of the data quality manager rather than the consumer of the data. This perspective tends to assume that data quality problems can be eliminated. Yet there is increasing evidence that organizations cannot completely eliminate such problems, particularly as the quantity and complexity of organizational data grows [30]. To the extent that organizations cannot always provide their users with the highest quality data, users increasingly need to assess the quality of the data they use in their work. And they do this assessment in the context of particular decision-making tasks. Thus it is important to understand how individuals assess data quality in context, and the effects of this process on decision outcomes. Metadata offers a means to integrate intrinsic and contextual aspects of data quality assessment, offering the possibility of better data process management and data quality assessment.

As organizational data continues to proliferate, data quality hazards will introduce growing challenges in organizations. Metadata offers technical, managerial and psychological avenues for addressing these challenges and represents an important research domain. This exploratory study contributes to this domain by offering insights into factors that affect individuals' performance during analytical, data-driven decision-making.

REFERENCES

[1] Bailey, J.E. and Pearson, S.W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science* 29(5) 530-545.

[2] Ballou, D.P. and Pazer, H.L. (1995), Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff. *Information Systems Research* 6(1) 51-72.

[3] Ballou, D.P., Wang, R.Y., Pazer, H. and Tayi, G.K. (1998). Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science.* 44(4) 462-484.

[4] Bandura A. (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.

[5] Baron, R.M., Kenny, D.A. (1986). The moderator-mediator variable distinction on social psychological research: Conceptual, strategic and Statistical Considerations. *J. Person. and Soc. Psych.* 51 1173-1182.

[6] Campbell, D. J., (1988). Task Complexity: A Review and Analysis. *Academy of Management Review* 13(1) 40-52.

[7] Chengalur-Smith I., Ballou, D. P., and Pazer, H. L (1999). The Impact of Data Quality Information on Decision-making: An Exploratory Study. *IEEE Transactions on Knowledge and Data Engineering.* 11(6) 853-864.

[8] Eckerson, W.W., (2003). Achieving business success through a commitment to high quality data. *TDWI Report Series*, Data Warehousing Institute, Seattle, WA, Http://www.dw-institute.com.

[9] Eisenhardt, K.M. and Zbaracki, M.J., (1992). Strategic Decision-making. Strategic Management Journal. 13 13-37.

[10] English, L. P. (1999). Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. John Wiley and Sons Inc, New York, NY.

[11] Fisher, C. W., Chengalur-Smith I., and Ballou D. P. (2003). The Impact of Experience and Time on the Use of Data Quality Information in Decision Making. *Information Systems Research*. 14(2) 170-188.

[12] Ford, C.M and Gioia, D.A. (2000). Factors Influencing Creativity in the Domain of Managerial Decision Making. *Journal of Management* 26(4) 705-732.

[13] Gendron, M., Shanks, G., and Alampi, J. (2004). Next Steps in Understanding Information Quality and Its Effect on Decision Making and Organizational Effectiveness., in the *Proceedings of 2004 IFIP International Conference on Decision Support Systems (DSS2004), (Ed. Widmeyer, G.), Prato, Tuscany, July 2004.*

[14] Hufford D. (1996) *Data Warehouse Quality.* DM Review, January.

[15] Hernandez, M. A. and Stolfo, S. J. (1998). Real-world Data is Dirty: Data Cleansing and the Merge/Purge Problem. *Journal of Data Mining and Knowledge Discovery* 1(2).

[16] Jarke, M., Lenzerini, M., Vassiliou, Y. and Vassiliadis, P. (2000). *Fundamentals of Data Warehouses.* Springer-Verlag, Heidelberg , Germany.

[17] Kahn, B. K., Strong, D. M., and Wang, R. Y. (2002). Information Quality Benchmarks: Product and Service Performance. *Communications of ACM* 45(4) 184-192.

[18] Klein, B.D., Goodhue, D.L. and Davis, G.B. (1997). Can Humans Detect Errors in Data? Impact of Base Rates, Incentives and Goals. *MIS Quarterly* 21(2) 169-194.

[19] Kimball R. and Ross M. (1998). *The Data Warehouse Lifecycle Toolkit*. Wiley Computer Publishing, New York NY.

[20] Mackay, J. M. and Elam, J. J. (1992). A Comparative Study of How Experts and Novices Use a Decision Aid to Solve Problems in Complex Knowledge Domains. *Information Systems Research* 3(2) 150-172.

[21] Marco D. (2000). *Building and Managing the Meta Data Repository: A Full Lifecycle Guide.* Willey and Sons, Inc., New York, NY.

[22] Mardsen, J.R, Pakath, R. and Wibowo, K. (2002). Decision Making under Time Pressure with Different Information Sources and Performance-based Financial Incentives. *Decision Support Systems* 34 75-97.

[23] Nevo D., Benbasat I. and Wand Y. (2003). Exploring Meta-Knowledge for Knowledge Management Systems: A Delphi Study. *Proceedings of the Twenty-fourth International Conference on Information Systems* 439-449.

[24] Nutt, P. C. (1984). Types of Organizational Decision Processing. *Administrative Science Quarterly* 29 414-450.

[25] Nutt, P. C. (1998). Framing Strategic Decisions. *Organization Science* 9(2) 195-216.

[26] Nutt, P. C. (1998). How Decision Makers Evaluate Alternatives and the Influence of Complexity. *Management Science* 44(9) 1148-1166.

[27] Nutt, P. C. (2002). Making Strategic Choices. *Journal of Managements Studies* 39(1) 67-93.

[28] Parssian, A, Sarkar, S., & Jacob, V.S (2004). Assessing Data Quality for Information Products – Impact of Selection, Projection, and Cartesian Product. *Management Science* 50(7) 967-982.

[29] Payne, J. W., Bettman, J. R. and Johnson, E. J. (1993). The Adaptive Decision Maker. *Cambridge University Press*, Cambridge, UK.

[30] Redman, T.C. (1996). *Data Quality for the Information Age.* Artech House, Boston, MA.

[31] Shankaranarayanan, G., Ziad, M., and Wang, R. Y. (2003). Managing Data Quality in Dynamic Decision Environments: An Information Product Approach. *Journal of Database Management* 14(4) 14-32.

[32] Shankaranarayanan, G. and Watts, S. (2003b). A Relevant Believable Approach for Data Quality Assessment. *Proceedings of the International Conference on Information Quality (IQ 2003)* Boston, MA.

[33] Shankaranarayanan, G. and Even A. (2006). Metadata: A Great Promise or a Sisphean Torture. *Communications of the ACM*. Vol. 49, No. 2, February, 88-94

[34] Simon, H. A., (1947). *Administrative Behavior.* The Free Press, New-York, NY.

[35] Speier, C. and Morris, M. G. (2003). The Influence of Query Interface Design On Decision-Making Performance. MIS Quarterly 27(3) 397-423.

[36] Strong D.M, Yang, W.L. and Wang, R.Y., (1997). Data Quality in Context. *Communications of the ACM* 40(5) 103-110.

[37] Thompson, J.D. (1967). *Organizations in Action.* McGraw-Hill, New York, NY.

[38] Wang, R. Y., Kon, H. B., and Madnick, S. E. (1993). Data Quality Requirements Analysis and Modeling. *Proceedings of the 9th International Conference on Data Engineering*, IEEE Computer Society Press, 670-677.

[39] Wang, R.Y. and Strong D.M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12(4) 5-34.

[40] Wang, R.Y. (1998). A Product Perspective on Total Quality Management. *Communications of the ACM* 41(2) 58-65.

[41] Wang, R. Y., Yang, L. W., Pipino, L. and Strong, D. M. (1998). Manage your Information as a Product. *Sloan Management Review* 39(4) 95-105.

[42] Wixom, B.H., Watson, H.J. (2001). An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly* 25(1) 17-41.

[43] Yang, W.L. and Strong, D. M. (2003). Knowing-Why about Data Processes and Data Quality. *Journal of Management Information Systems* 20(3) 13-39.

AUTHORS' BIOGRAPHIES

G. Shankaranarayanan is an Assistant Professor in the Information Systems Department in Boston University School of Management. His research areas include schema evolution in databases, heterogeneous and distributed databases, data modeling requirements and methods, and structures for and the management of metadata. Specific topics in metadata include metadata implications for data warehouses, metadata management for knowledge management systems/architectures, metadata management for data quality, and metadata models for mobile data services. His work has appeared in Decision Support Systems, Journal of Database Management, and in the Communications of the ACM.

Adir Even is a doctoral student at the Boston University School of Management, Information Systems department. He studies the contribution of data and systems to value-gain and profitability from the theoretical and the practical perspective, and explore possible implications for system architecture, data warehousing, data quality management, and organizational knowledge. He holds a MBA degree from the Tel-Aviv University School of Management, a M.Sc. in Computer Engineering from the Technion of Israel, and has substantial software-industry experience as a senior software-development manager, responsible for large-scale data warehousing and DSS implementations.

Stephanie Watts is an Assistant Professor of Information Systems at the Boston University School of Management. She was previously on the faculty of the Weatherhead School of Case Western Reserve University. Her research focuses on the various roles that information technology plays in organizations, with a focus on knowledge sharing. She has published academic papers in such journals as Information Systems Research, Organization Science, and Information and Management.