DATA QUALITY INFORMATION AND DECISION-MAKING: SEMANTICS, USABILITY, AND IMPACT

Rosanne J. Price[1,2]
Graeme Shanks[2]
[1]Clayton School of IT, Monash University, Australia;
[2]Department of Information Systems, University of Melbourne, Australia

Email: [rosanne.price@infotech.monash.edu.au, gshanks@unimelb.edu.au]

Abstract

There has been some evidence that providing decision-makers with metadata about the quality of the data used to make a decision may impact decision outcomes. However, there has been little emphasis on the semantics and usability of such metadata in research to date, with consequent implications for experimental validity, generalizability, and reliability. In this paper, we describe how these issues are addressed through the novel application, respectively, of the information quality framework InfoQual and the interaction design technique of contextual inquiry to experimental design. A usability study based on contextual inquiry highlighted possible design issues in previous experimental designs and provided recommendations to guide the design of current experiments. In distinct contrast to previous studies, there was no evidence that data quality metadata was used when it was included, but there was still some evidence of decreased decision consensus and efficiency.

Keywords: Data quality tagging, Decision-making, Contextual inquiry, Experimental soundness

1 INTRODUCTION

In the context of data warehouses and decision-support systems, decision-making commonly involves reliance on heterogeneous data sources that are remote from the user and possibly external to the organization or individual. Decision-makers routinely use data whose context is unfamiliar and may vary depending on the individual data type and source. In particular, it is unlikely that the set of data used to support decision-making is of uniform quality. The quality of the data used to make a multi-criteria-based decision potentially impacts the effectiveness of that decision. It has therefore been proposed (Chengular-Smith, Ballou and Pazer 1999) that information about data quality (DQ) be provided to decision-makers in the form of metadata, called *DQ tags*.

The use of DQ tags could potentially impact decision outcomes such as the decision-making efficiency (e.g. if decision-makers take time to consider data quality), the resultant decision choice (e.g. if criteria that would otherwise be considered are disregarded because of their low quality ratings), or the decision-maker's confidence in that decision. For example, when using an on-line system to search for a rental property; users may choose to ignore floor-space or treat it as a low priority if it is known to be unreliable compared to other criteria. Since the use of DQ tags is associated with significant overheads with respect to tag creation, storage, and maintenance; the adoption of DQ tagging as a business practice needs to be justified by a clear demonstration of its efficacy. The use of DQ tags by decision-makers may depend on factors such as decision-maker characteristics (eg. experience), the decision-making strategy employed, or the complexity of the decision-task. In order to predict the decision-making contexts most likely to benefit from the use of DQ tagging, it is therefore important to understand how such factors influence the use of DQ tags.

For example, decision-making strategies (Payne, Bettman and Johnson 1993) can be classified based on whether multiple attributes of a single alternative or multiple alternatives for a single attribute are considered first (*alternative* or *attribute* based respectively) and whether desirable values for one attribute can or cannot compensate for undesirable values in another attribute (i.e. *compensatory* or *cutoff* respectively). An *Additive* decision-making strategy is therefore classified as *alternative* and *compensatory* based because it involves assigning a desirability score to each attribute value for a single alternative, calculating a summed scores for each alternative, and then choosing the alternative with the highest sum—even though that alternative may have very undesirable values (ie. low desirability scores) for some attributes. In contrast, the *Elimination by Attributes (EBA)* strategy involves elimination of alternatives not meeting the minimum requirements for the value of the attribute most important to the decision-maker. This process is repeated for other attributes in descending order of priority until only one alternative remains. Similarly, different levels of task complexity can be defined in terms of the number of decision criteria and alternatives and the ease of differentiating between alternatives. Thus research into the effects of DQ tagging on decision-making in different circumstances constitutes a pre-requisite step to any proposed implementation of DQ tags.

1.1 Previous Work

Previous investigators (Chengular-Smith et al. 1999; Fisher, Chengular-Smith and Ballou 2003; Shanks and Tansley 2002) have reported that DQ tags impact decision outcomes under certain circumstances, although there is conflicting evidence as to which specific circumstances (see Shanks and Tansley 2002, p12). These studies have in common the use of attribute-level tagging, the definition of two levels of task complexity (simple and complex), and the premise that a significant difference in the preferred decision-choice with and without DQ tags implies that the tags were used in the decision-making process when available. DQ tag *usage* is thus defined with respect to changes in preferred decision choice in the literature. We note, however, that there may potentially be some limited consideration of DQ tags by participants that affects decision consensus or efficiency even if decision choice is not significantly affected (ie. reported in the literature as "no evidence of tag use").

There is agreement (Chengular-Smith et al. 1999; Fisher et al. 2003, Shanks and Tansley 2002) that increased task complexity is associated with reduced DQ tag usage, explained in terms of information overload. Shanks and Tansley (2002) reported DQ tag use only with an EBA but not an additive strategy, whereas Chengular-Smith et al. (1999) found DQ tag use to be more prevalent when DQ information was presented in a manner convenient for use with an additive strategy. In general, DQ tagging experiments to date indicate that DQ tag usage increases with experience (see Fisher et al. 2003 p182)—with no use of tags by freshmen undergraduate students (Fisher et al. 2003), use of tags for simple but not complex tasks for senior undergraduate students (Chengular-Smith et al. 1999; Shanks and Tansley 2002), and use of tags for both simple and complex tasks for professionals (ie. those with at least one year's prior work experience) (Fisher et al. 2003). All previous studies reported instances of decreased consensus in decision choice when DQ tags were used (ie. when there was a significant difference in preferred decision choice); however, there was never any decrease in consensus observed for those cases where DQ tags were ignored (ie. present but not used).

Chengular-Smith et al. (1999) and Fisher et al. (2003) have limitations with respect to the data sample size used and the control of the decision-making strategy used. These studies relied on small paper-based data sets of less than eight alternatives. This is in marked contrast to the large data sets characterizing on-line decision-making, with obvious implications for the *generalizability* (ie. applicability to the real-world and to contexts other than that considered in the experiment) of the experimental results to on-line decision-making and the credibility of the experimental context to participants. If participants do not feel involved in the experimental task, they may not be affected by (and their feedback may thus not be a response to) the experimental treatments, potentially impacting the *reliability* (ie. repeatability) and *validity* (ie. satisfactorily establishing that observed results are based on manipulations of dependent variables) of the research (see Neuman 2006, p265).

With respect to decision-making strategy, Fisher et al. (2003 p183) acknowledges in "that in our study we did not constrain the subjects to using a particular decision-making strategy"and that "the posttask questionnaire...found that most subjects used a combination of strategies”. In Chengular-Smith et al. (1999), the differentiation between compensatory and non-compensatory decision-strategies in experimental treatments relied primarily on the inclusion or omission of cut-off scores in the DQ information presented to participants. Decision alternatives were not presented to participants in an order corresponding to the ranking resulting from a given strategy; nor were summed scores for each alternative given for the additive decision-strategy. Consequently, observed decision outcomes that were attributed solely to the use of tags could actually have depended (partly or completely) on the strategy or strategies used—as was shown by Shanks and Tansley (2002). This study addressed both concerns of scale and strategy through the use of separate on-line interfaces, each with a different built-in decision-making strategy, to access an electronic database with 100 alternatives. Furthermore, in order to provide better support for validity, the same decision domain and attribute-based DQ tag granularity was used for both simple and complex tasks; whereas that was not true in Chengular-Smith et al. (1999) and Fisher et al. (2003). However, a limitation common to all DQ tagging studies to date is that there has not been a focus on the semantics or usability of DQ tags.

Tag semantics (i.e. meaning) relate to the specific DQ characteristic or criterion (e.g. *security* or *precision*) whose value is represented by the DQ tag. The only guide to the meaning of the DQ tag used in previous experiments is its label (reliability in Chengular-Smith et al. (1999); Fisher et al. (2003) and accuracy in Shanks and Tansley (2002)), without any further explanation given participants. In fact, a DQ tag could potentially be based on a number of different DQ criteria discussed in the literature (Eppler 2001, Wand and Wang 1996, Wang and Strong 1996). If the semantics of the DQ tags used are not explicitly defined and specified to participants, there might not be agreement between the interpretations of different experimental subjects or between their interpretations and that of the researcher. Such a problem could easily go undetected in a pilot study. Individual subjects may say that the meaning of the DQ tags used in experimental materials is clear because they each have their own internal—even if erroneous or inconsistent—interpretations. This could lead to random error in when or how DQ tags are used that impacts experimental reliability. Specific cases of such an occurrence in practice are documented in Price and Shanks (2008a 2009), where data labelled as being of poor quality using the term *accuracy* were used for decision-making by participants who interpreted the DQ tag semantics as numerical precision and ignored by participants who interpreted the semantics as correctness (ie. the correspondence of the database to the real-world).

The only explicit reference to usability in the DQ tagging literature is the use of pilot tests in Fisher et al. (2003) and Shanks and Tansley (2002), a technique whose limitations were illustrated above and are further discussed in Price and Shanks (2008a, 2009). Since the use of DQ tags is not common in current business practice, we lack real-world precedents or widely-understood conventions available to guide the researcher in designing or the user in understanding DQ tags. The only explicit consideration of different DQ tag representations to date is experimental treatments using two-category ordinal (ie. *below average, above average*) versus continuous (ie. an integer between 1 and 100) representations of DQ information in Chengular-Smith et al. (1999). Other possibilities for representing DQ values (eg. using ranges rather than single points, using graphics) and other aspects of DQ tag representation such as tag nomenclature or documentation have not been explicitly considered.

Previous papers by the authors (Price and Shanks 2008a, 2009) have argued for the explicit specification of DQ semantics and the explicit consideration of usability when designing DQ tagging experiments in order to improve support for experimental *soundness* (i.e. the generalizability, reliability, and validity). Tag semantics, including derivation rules (the method used to calculate tag values), must be specified explicitly to ensure that the DQ information provided is meaningful and the experimental context is credible. The need to consider alternative types of DQ tags based on the different aspects of DQ described in the literature is highlighted. Price and Shanks (2008a, 2009) further describe the novel use of contextual inquiry interviews (Beyer and Holtzblatt 1998; Holtzblatt and Jones 2000) in a usability study. This study resulted in a set of design recommendations based on decision-makers’ judgements of what DQ information was understandable and relevant to decision-making in practice. The current paper describes DQ tagging experiments that incorporate these design recommendations and that explicitly specify DQ tag semantics, thus addressing the limitations of prior research with respect to design and usability. Section 2 discusses DQ tag semantics and cost-based considerations of DQ tag design. Section 3 gives an overview of the usability study and consequent design recommendations in sufficient detail to understand the rationale for the current experimental design. Section 4 presents the experimental design and research hypotheses. Results and discussion follow in Section 5 and the conclusion in Section 6.

2 DQ TAG SEMANTICS AND COST-CONSIDERATIONS

The semantics and derivation of DQ tags for the experiment are explicitly specified in the experimental materials given to participants. DQ tag semantics could potentially be based on metadata either indirectly or directly related to DQ. Data characteristics such as *source* or *processing history* do not directly describe DQ but are frequently employed by users as a basis for judging the likely quality (e.g. trustworthiness) of data, as described in Even, Shankaranarayanan and Watts (2006). The semantics of tags directly related to DQ are based on a previously defined information quality framework called *InfoQual* (Price and Shanks 2005), selected by virtue of its sound theoretical foundation and comprehensive coverage of different aspects of DQ. Different types of DQ tags can be defined based on *InfoQual*'s three DQ categories and their criteria. The first two categories (i.e. data *conformance* to rules and real-world *correspondence*) are relatively objective in nature since they are inherently based on the data set itself rather than the individual user or context of use. In contrast, the third category (i.e. *usefulness*) is necessarily subjective since it is based on context-specific information consumer (i.e. user) views (see Price and Shanks 2005 for a detailed discussion). Therefore; DQ tags based on subjective quality measures must be associated with additional contextual information (e.g. activity or task, organizational or geographic context, user profile) to be meaningfully interpreted or used. Within each of these three categories, *InfoQual* defines a set of individual DQ criteria further detailing DQ aspects. For example, The *usefulness* category includes criteria such as *security* and *timeliness*; whereas the *correspondence* category includes criteria such as *completeness, correctness*, and *consistency* (of the real-world representation in the database).

Since the creation, storage, and maintenance of tags incurs additional costs that offset potential benefits; it is desirable to restrict the scope to those choices that are likely to be the most practical in terms of simplicity, cost, and use. In the following discussion, we assume that a single decision-making *criterion* is represented as a database *attribute* or equivalently, in a relational database, as a *column* in a relational table; thus the three terms are used interchangeably. Considered in detail in Price and Shanks (2009) and summarized here, cost considerations thus suggest that DQ tags be:

- Objective: ie. have semantics based on the relatively objective aspects of DQ described earlier (i.e. source, process-history, conformance, correspondence) so that there is no need to store any of the additional contextual information required for subjective DQ.
- Column-based: ie. specified in terms of a single DQ measurement for all the values of a single column. Data quality tags can potentially be specified at different levels of granularity (eg. relation, column, row, field within the relational model) with the obvious trade-off that overheads and information value increase at finer tagging granularities. Column-level tagging is a natural compromise in the context of multi-criteria decision making, since the underlying cognitive processes involve evaluation of decision alternatives in terms of relevant criteria that are represented as columns. This is the coarsest (and thus cheapest) level of granularity likely to influence decision-making and could be easily supported by extending column-level information in relational data dictionaries.
- Consolidated: composite tags used when possible to logically combine related DQ criteria in order to limit storage overheads and reduce semantic complexity, eg. previous research (Price and Shanks 2008b) shows that users find it difficult to distinguish between individual criteria in the

data *correspondence* category of *InfoQual,* preferring to combine them in a single summarized category-level concept instead—thus supporting the use of a single composite tag in this case.

These choices serve to limit the amount of information (and thus overhead costs) required for DQ tags. However, cost is not the only consideration for DQ tag design. As discussed in Section 1, it is important to address usability concerns, especially given their potential impact on experimental soundness and the lack of common business conventions that could serve to guide experimental design or participant understanding of DQ tags. Usability questions relate to which aspects of DQ and which possible DQ tag representations (including incorporation of such tags into the decision-making artefact) are the most understandable and relevant to decision-makers. Explicit consideration of such questions can facilitate the efficient use of research resources (eg. subjects, funding) by identifying which types of DQ tags are of interest to decision-makers and thus should be included in experiments. The next section overviews a usability study conducted for this purpose.

3 USABILITY STUDY

In order to improve the usability and thus better support experimental soundness, contextual inquiry interviews can be used to solicit user opinions on DQ tag design and use in the work environment rather than in a contrived laboratory setting. Beyer and K. Holtzblatt (1998) and Holtzblatt and Jones (2000) describe how work processes and artefacts serve as reminders enabling decision-users to articulate their opinions in more detail and in reference to actual business practice (eg. rather than to the internal coherence of an experiment). Contextual inquiry is the interrogatory component of contextual design, described in Beyer and K. Holtzblatt (1998) and Holtzblatt and Jones (2000). This technique involves on-site interviews of decision-makers while they demonstrate their real decision-making tasks. We summarize the goals, design, and resulting recommendations of the contextual-inquiry-based usability study reported in detail in Price and Shanks (2008a, 2009).

The goals of the usability study were (1) to observe relevant types of decision-making in practice and (2) to find DQ tag semantics and representations and decision-making interfaces that decision-makers consider to be both understandable and likely to improve their decision effectiveness or confidence.

Following the contextual inquiry guidelines from Beyer and K. Holtzblatt (1998) and Holtzblatt and Jones (2000) for a single work role (ie. decision-maker), two investigators conducted one-hour audio-taped interviews with nine different decision-makers representing a diverse set of organization (eg. sizes), data (eg. types and sources), decision (eg. domain), and technological (eg. software) contexts. Interviewees were asked to demonstrate a multi-criteria, data-intensive, and on-line decision-making task and to explain their experience in response to interviewer questions throughout the interview. In order to address specific usability questions related to DQ tag and experimental design, an additional and novel segment was added after the standard contextual inquiry interview. This ordering ensures that the additional segment will not bias the initial demonstration of current decision-making practice.

In the additional segment, decision-makers were first asked to reflect on possible ways to improve the demonstrated decision-making task using DQ tags and then asked to review alternative DQ tag representations and a proposed decision-making interface incorporating tags based on the study in Shanks and Tansley (2002). Interviewees were asked what types of DQ information (ie. DQ tags based on which specific DQ criteria) would be useful and which representations they preferred for their work and for our experimental context. Transcribed interviews were analyzed by recording individual investigator interpretations on post-it notes and then reconciling and structuring those interpretations as a team. The final result of this process was a table summarizing interviewee responses on a structured list of topics and a set of design recommendations based on this analysis. The value of these recommendations is supported by the general agreement between interviewed decision-makers despite their diverse contexts and despite the two different domains (work and experimental) considered.

The only type of DQ information considered to be of general interest at the attribute-based level was the degree of data *correspondence* to represented real-world values. Based on clear respondent preferences, the recommended representation of such information would be as follows:

- use the term *accuracy* for tag nomenclature,
- use a traffic light to graphically represent range-based tag values based on both color and position,
- include explicit documentation of tag semantics and derivation on-line, using brief explanations in pop-up boxes, and in any paper-based explanatory or instructions materials given to participants.

Furthermore, with respect to the proposed experimental decision-making software artefact (ie. on-line decision-making interface) it was recommended that:

- other interface elements (eg. attributes, value units) be documented using pop-up boxes, and
- desirability scores (see next paragraph for discussion) currently included in the decision-making interface be omitted as they were regarded as unnecessary and confusing.

All previous DQ tagging experiments (Chengular-Smith et al. 1999; Fisher et al. 2003, Shanks and Tansley 2002) have displayed desirability scores. Such scores allow the relative desirability of different attribute values and alternatives to be compared despite differences in measurement units (eg. dollars rent versus minutes commute time for rental properties) and directionality (prefer a lower rent but more bedrooms). With respect to DQ tag design, most studies have used a single numerical figure to represent DQ information. None has used a graphical symbol or color to represent DQ information or has explicitly specified DQ tag semantics in the experimental materials used by participants. This is in direct contrast to the expressed preferences of the majority of interviewed decision-makers.

4 RESEARCH DESIGN

A laboratory experiment is used to examine the impact of DQ tagging on decision outcomes for different decision-making strategies. The research model is shown below in Figure 1, illustrating the causal relationships between theoretical constructs (in ovals) using solid lines and the measurement relationships between theoretical constructs and their empirical indicators (in rectangles) using dotted lines. In common with Shanks and Tansley (2002), the focus is on multi-criteria, data-intensive and on-line decision-making. In general, our experimental methodology and design is based on theirs. However, central to our work—and distinguishing it from previous DQ tagging research—is the explicit emphasis on usability and semantics in designing DQ tags and incorporating them into a decision-making artefact. The experimental design is modified in accordance with the semantic-, cost-, and usability-based recommendations outlined in Sections 2 and 3.

The independent variables were decision-strategy and DQ tagging, both of which had two levels. Additive and EBA strategies are selected based on their contrasting properties (i.e. compensatory and alternative-based versus cut-off—therefore non-compensatory—and attribute-based respectively). DQ tags are either present or absent. The result is four separate experimental treatments: additive with DQ tags, additive without DQ tags, EBA with DQ tags, or EBA without DQ tags.

The dependent variables, potentially impacted by DQ tags, are decision complacency, consensus, efficiency, and confidence As in previous DQ tagging studies (Chengular-Smith et al. 1999; Fisher et al. 2003, Shanks and Tansley 2002), complacency and consensus relate to the impact of DQ tags on decision choice. Complacency refers to the degree to which decision-makers ignore DQ information. Thus a non-complacent outcome means that there is a significant change in the preferred decision choice when DQ tags are present as compared to that made when DQ tags are absent. The preferred decision choice is defined as that made by the plurality of participants in the treatment group. Consensus refers to the level of agreement on decision choice between decision-makers: we are interested in whether the level of agreement with DQ tags is the same as that without DQ tags. Essentially, we are asking whether there is any difference in the number of participants making the preferred decision choice with and without tags, even though the preferred choice may not be the same for the two different groups. In accordance with the research model in Shanks and Tansley (2002), we investigate whether the presence of DQ tags significantly changes the time required to make the decision or the decision-maker's confidence in the decision choice made.

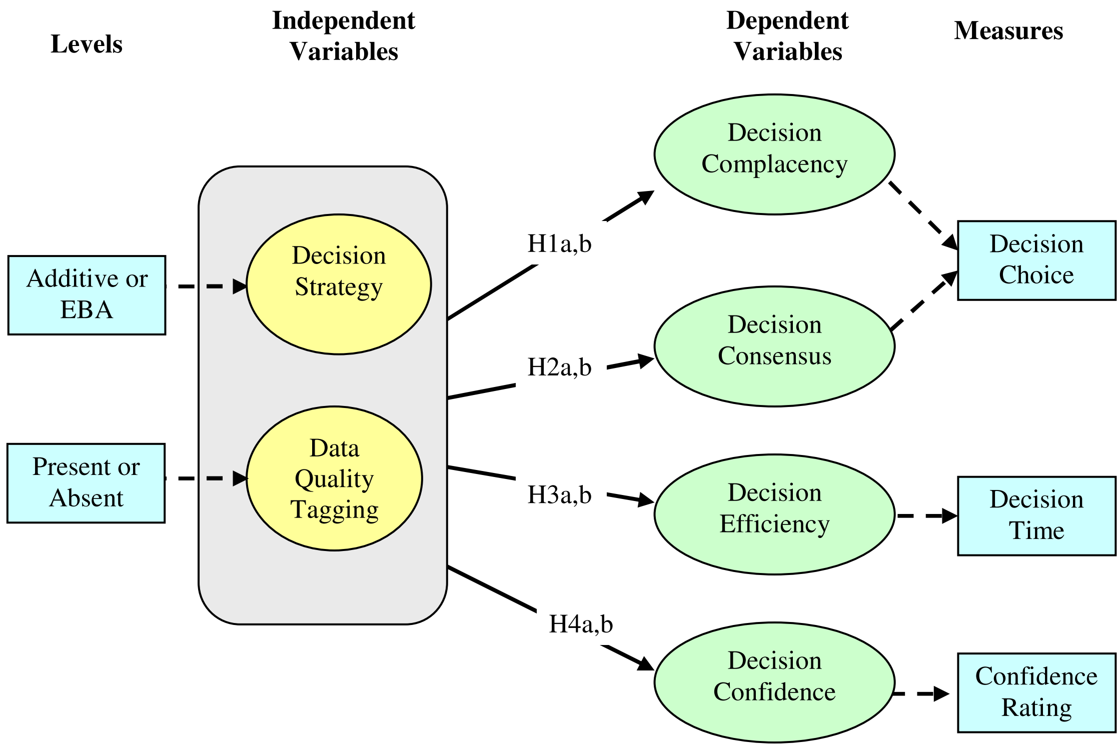

Figure 1. Research Model.

Based on these variables, the case for the potential benefit of DQ tags would best be supported if the experiment shows that decision-makers are not complacent and have increased (or at least the same) level of consensus, efficiency, and confidence with tags. Such results require the rejection of the corresponding null hypotheses, formulated as follows: decision-makers are complacent (*H1*), and there is no difference in decision consensus (*H2*), efficiency (*H3*), or confidence (*H4*) with DQ tags for either the additive (*a*) or EBA (*b*) decision-making strategies. A chi-squared statistic is used to test the first two hypotheses. Depending on whether the data is normally distributed or not, either an independent samples t-test or a Mann-Whitney test is used for the last two hypotheses.

Given prior research evidence that DQ tag usage generally increases with decision-maker experience (see Section 1) and considering available resources, participants were limited to postgraduate university students enrolled in a masters or PhD degree. Such students are likely to have had more decision-making and professional experience than undergraduates. In fact of the 62 participants, 25 (ie. 44%) had prior work experience and 10 (ie. 16%) had prior managerial experience.

The decision-making task used in the experiment involves selection of a preferred rental apartment based on the weekly rental cost, number of bedrooms, floor space, commute time, and parking facilities. The task involved recording decision start and finish times, nominating a confidence level using a 5-point likert scale ranging from *very low* to *very high*, and providing a brief explanation of the decision. In particular, they were asked to write down which attributes—if any—they ignored in their search and why. Surveys of the target participant population of postgraduate university students showed that they typically had a good understanding of the domain and were frequent users of actual on-line rental property selection applications. The rental-property application domain, the set of attributes used (both their description and number), the treatment group sample sizes, and the DQ tag values are selected in order to be consistent with the simple task used in previous DQ tagging research

(Chengular-Smith et al. 1999; Fisher et al. 2003, Shanks and Tansley 2002). In common with the rental property decision task from all previous studies, the attribute *commute time* has the lowest DQ value. Given prior research evidence that DQ tag usage is more likely in a simple rather than a complex decision-making task—especially for university students, we limit our consideration to the former. In the context of the decision-making task, the above hypotheses are operationalized as no significant difference in the preferred apartment (*H1*), the number of decision-makers selecting the preferred apartment (*H2*), the decision time (*H3*), or the nominated decision confidence (*H4*).

As in Shanks and Tansley (2002), we adopt a relational database-type interface and use Microsoft Access software for development – both well-understood and widely used. Issues of scale and decision-making strategy (each potentially affecting the impact of DQ tagging) are similarly addressed through the use of separate on-line interfaces for each experimental treatment, each with 100 alternatives and a specific built-in decision-strategy and some including DQ tags. A set of instructions and an answer sheet were also developed for each different interface. In common with previous DQ tagging experiments, the alternatives in the databases are designed so that one apartment is clearly the most desirable without DQ information but is less desirable if DQ information is considered. The EBA interface with tags is shown in Figure 2.

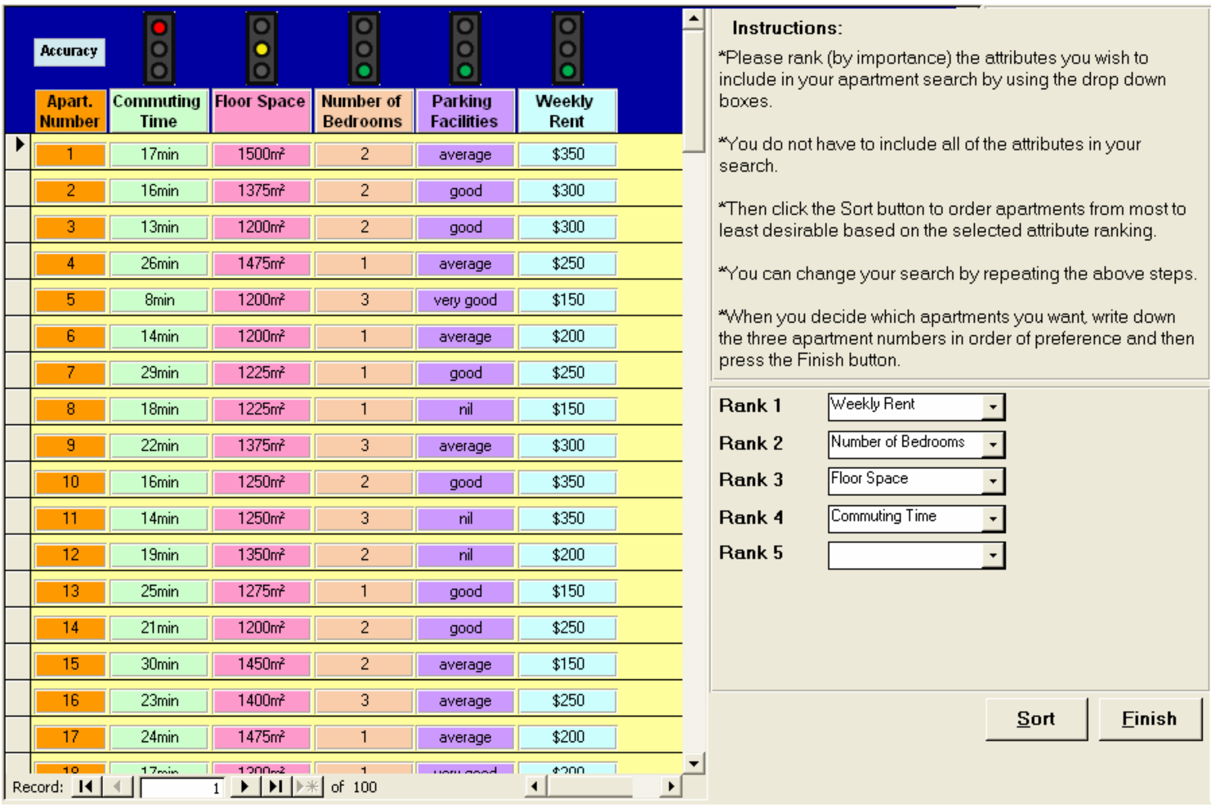

Figure 2. Interface for Elimination-by-attribute (EBA) decision strategy with DQ tags

In line with the recommendations resulting from the usability study described in Section 3, calculated desirability scores are not displayed and concise explanations of interface elements are given in pop-up windows. Further to this purpose and in contrast with previous experiments, the current experiments use DQ tags with (1) semantics and derivation explicitly specified (in both on-line and paper-based experimental materials), (2) semantics based on data correspondence (to the real-world, as defined in *InfoQual* in Section 2), and (3) range-based values represented by a traffic light symbol labelled *accuracy*. Furthermore, illustrative examples are included in explanatory notes given participants to clarify that the intended meaning of this term is *correctness* (based on real-world correspondence) rather than *numerical precision* (see Section 1 for documented evidence of such confusion).

As per semantic- and cost-based considerations highlighted in Section 2, the DQ tags used in the experiment are attribute-based (as in all previous DQ tagging research except the complex task in Fisher et al. (2003) Appendix A, p184, where field-based DQ tags were used), based on relatively objective rather than subjective DQ information, and are consolidated. As explained in Section 2, user preferences evident from prior research (Price and Shanks 2008b) support the consolidation of correspondence-based DQ information (described by the set of related *InfoQual* criteria in the *correspondence* category) into a single DQ tag.

The experimental design was initially piloted individually with 5 postgraduate students and 1 professional. They were asked to verbalize their thoughts during the experiment, resulting in minor modifications to the screen display and instruction wording and repair of one bug. A further pilot study with 14 postgraduate students writing down their comments found the materials clear and did not result in any new suggestions, indicative of saturation.

Participants were randomly assigned to one of 4 treatment groups. After being introduced to the experimental task, decision-makers are asked to search for, select, and rank in order of priority their preferred apartment. Decision-makers select and—for the EBA strategy—rank those attributes they wish to include in the decision-making process. Desirability scores are automatically calculated and alternatives sorted in order of decreasing desirability based on the attributes selected and the specific decision-making strategy built into the interface. Participants can then repeatedly change their attribute selections and request a new sort. For the treatments involving DQ tags, completed answer sheets were checked before participants left the laboratory to see if the attribute with the lowest DQ value—commute time—was used in the search. If so, the participant was asked whether he or she understood the meaning of the DQ tags and—if so—why they used it despite its poor quality.

5 RESULTS AND DISCUSSION

A chi-squared test checks for differences between an observed and the expected frequency distribution, where expected frequencies are derived from groups with no DQ tags. This test is non-parametric and therefore relatively free of underlying assumptions (Pallant 2001). Yates' Correction for Continuity is used as appropriate for a 2x2 chi-squared table. Results for analysis of decision complacency (*H1a, H1b*) and consensus (*H2a, H2b*) are summarized in Table 1. We can see that there is no significant difference ($p>.05$) in either for either decision-making strategy; therefore, it is not possible to reject the null hypotheses.

	Decision Strategy	
	Additive	Elimination by Attributes
Complacency	$\chi^2 = 1.874$	$\chi^2 = .212$
	$p = .171$ (H1a)	$p = .645$ (H1b)
Consensus	$\chi^2 = 1.874$	$\chi^2 = .212$
	$p = .171$ (H2a)	$p = .645$ (H2b)

Table 1. Analysis of Complacency and Consensus

For either decision-making strategy, Table 2 shows that the preferred apartment is the same with and without tags; therefore, the chi-squared statistic is the same for complacency and consensus. For each treatment group, Table 2 also shows the total number and percentage of participants selecting any other than the preferred apartment under the label *Other*. Information about the plurality of participants next in size compared to that of the participants selecting the preferred apartment is given under the label *Alternate*. The size of each treatment group is specified under the label *Total.* We can see that in all cases (ie. treatments), the plurality of participants selecting the preferred apartment is much larger (ranging from 18 to 73% larger depending on the treatment) than any other plurality; thus the less sensitive non-parametric chi-squared statistic does not show a significant difference in consensus. However, for both decision-strategies, the percentage of participants selecting the preferred apartment decreases (from 80% to 50% with additive and from 50% to 35% with EBA) and the range of different apartments selected increases (from 4 to 10 with additive and from 3 to 8 different apartments with EBA) when DQ information is present (compared to when it is not). This suggests that there is in fact an overall decline in consensus with tags that is not detected by the chi-square test.

		Number of Participants Selecting Apartment (% in treatment group selecting from set of specific apartments listed)	
		No Tags	Tags
Additive	Preferred	12 (80% for apt 70)	8 (50% for apt 70)
	Other	3 (20% for apt 77,83 or 98)	8 (50% for apt 5,33,37,47,49,62,77,83 or 98)
	Alternate	1 each (7% each for apt 77, 83 and 98)	2 (12% for apt 33)
	Total	15	16
EBA	Preferred	7 (50% for apt 5)	6 (35% for apt 5)
	Other	7 (50% for apt 33 or 66)	11 (65% for apt 16,33,38,44,66,98 or 100)
	Alternate	4 (29% for apt 33)	3 (17% for apt 66)
	Total	14	17

Table 2. Number of Participants Selecting Preferred and Other Apartments

Based on participant responses to the exit query described in Section 4, only 4 (1 using the EBA and 3 using the Additive decision-strategy) of the 33 participants given DQ information ignored the low quality *commute-time* attribute. Of those, 2 didn't care about commute time and 2 ignored it because of its low quality. Of the 29 participants not given DQ information, 1 (using the Additive strategy) ignored *commute-time* because they didn't care about it. Of the 29 participants given DQ information but using *commute-time* in their search, only 1 did not understand the meaning of the tags. All the others said they considered *commute-time* to be so important that they included it despite its low quality. Petrol prices, environment, and traffic were all given as reasons for the attribute's importance. The importance given this attribute is further highlighted by written comments from participants.

Since a visual inspection of relevant histograms, the Kolmogorov-Smirnov test, and the Shapiro-Wilks test revealed that both decision efficiency and confidence rating demonstrated significant variations from normal distribution for both decision-strategies, the non-parametric Mann-Whitney test was used for data analysis (*mean rank* and *level of significance* shown for each treatment group). The *mean* and *standard deviation* are also shown for each treatment group. Results for analysis of decision efficiency (*H3a, H3b*) and confidence (*H4a, H4b*) are summarized in Table 3. It can be seen that the only significant result is for decision efficiency using the additive strategy, where the presence of DQ information is associated with increased decision-time.

		Decision Strategy					
		Additive			Elimination by Attributes		
		Mean rank	Mean	SD	Mean rank	Mean	SD
Time	No Tags	12.67	4.14	1.92	14.18	5.29	2.81
	Tags	19.13	6.38	2.60	17.50	6.47	3.28
		p = .046* (H3a)			p = .308 (H3b)		
Confidence	No Tags	14.40	2.00	.76	14.29	2.00	.56
	Tags	17.50	2.31	.80	17.41	2.24	.56
		p = .305 (H4a)			p = .247 (H4b)		

Table 3. Analysis of Time and Confidence

6 CONCLUSION

The results for decision time and confidence were consistent with the findings reported in Shanks and Tansley (2002). Decision confidence remained *high* throughout all treatments. The decision time was increased with DQ tags for the additive but not the EBA decision-strategy, however, the *p* value was not highly significant (p=.046). Shanks and Tansley (2002) suggest that the increased time required for the additive strategy could be because the impact of an individual attribute on the sort sequence is less obvious given the compensatory nature of the strategy.

With respect to decision complacency and consensus, the results of this study are significantly different from that of previous DQ tagging research (Chengular-Smith et al. 1999; Fisher et al. 2003, Shanks and Tansley 2002) in that (1) there was no evidence of DQ tag use with respect to changing the preferred decision-choice and (2) there was some indication (based on Table 2 rather than the relevant chi-square test) of an overall decline in consensus even when DQ tags were not used. Participant comments reveal that they considered *commute-time* much too important to ignore, even when available DQ information indicated it was of low quality. The reasons given for its importance were generally petrol prices, environment, or traffic. We note that these factors have become of much more urgent concern in the years since previous DQ tagging experiments were conducted, with the massive increase in petrol prices and environmental awareness and continuing increase in traffic volumes in the last few years. Thus the attribute of lowest quality in the rental property scenario (used in all DQ tagging experiments to date) is likely to have become much more critical from the participants' perspective as compared to earlier DQ tagging studies, which may help explain the difference in results. One of the decision-makers interviewed in the usability study expressed his opinion that DQ information was only of potential use for moderately important decision-criteria, since decision-makers would include critical criteria and exclude marginal criteria from their decision-making process in any case. The current findings provide support for the assertion regarding critically important criteria and suggest that the relative importance of different attributes to the decision should be considered in experimental design, ie. when selecting DQ tag values to use in experiments.

The previous studies could be said to provide limited support for the possible utility of DQ tags in specific circumstances, although with caveats regarding the associated risk of increased decision-time and reduced consensus when DQ tags are used. However, since there were indications of decreased decision efficiency and consensus even without any evidence of DQ tag use, the findings reported in this paper clearly contraindicate the adoption of DQ tags in general business contexts.

Further work is required to explore the impact of DQ tags on decision-making. Additional experiments looking at the effect of changed DQ tag values will strengthen our understanding of whether the impact of tags depends on how important the tagged attribute is to the decision. Cognitive process tracing studies will help understand the impact of DQ tags on decision-making processes and help explain experimental results.

References

Beyer, H. and K. Holzblatt (1998) Contextual Design: Defining Customer-Centered Systems, San Francisco, Morgan Kaufmann Publishers, Inc.

Chengular-Smith, I.N., Ballou, D. and H.L. Pazer (1999) The Impact of Data Quality Information on Decision Making: An Exploratory Analysis, IEEE Transactions on Knowledge and Data Engineering, 11:6, p. 853-864.

Eppler, M.J. (2001) The Concept of Information Quality: An Interdisciplinary Evaluation of Recent Information Quality Frameworks, Studies in Communication Sciences, 1, p167-182.

Even, A., Shankaranarayanan, G. and S. Watts (2006) Enhancing Decision Making with Process Metadata: Theoretical Framework, Research Tool, and Exploratory Examination, Proc. of the 39th Hawaii International Conference on System Sciences (HICSS2006), Hawaii, 1-10.

Fisher, C., Chengular-Smith, I.N. and D. Ballou (2003) The Impact of Experience and Time on the Use of Data Quality Information in Decision Making, Information Systems Research, 14:2, p170-188.

Holtzblatt, K. and S. Jones (2000) Conducting and Analyzing a Contextual Interview (Excerpt), *Readings in Human-Computer Interaction: Towards the Year 2000*, San Francisco, Morgan Kaufmann Publishers, Inc., p241-253.

Neuman, W. L. (2006) Social Research Methods: Qualitative and Quantitative Approaches, 6th edition, Allyn & Bacon

Pallant, J. (2001) SPSS Survival Manual A step by step guide to data analysis using SPSS for Windows (Version 10), Allen & Unwin, Crows Nest, NSW, Australia.

Payne, J., Bettman, J. and Johnson, E. (1993) The Adaptive Decision Maker, Cambridge Univ. Press

Price, R. and G. Shanks. (2009) Representing Data Quality Information Usably, Clayton School of Information Technology, Monash University, Technical report 2009/243, p1-19.

Price, R. and G. Shanks. (2008a) Data Quality Tags and Decision-making: Improving the Design and Validity of Experimental Studies. In: IFIP TC8/WG8.3 Working Group's International Conference on Collaborative Decision-Making (CDM'08), Toulouse, p233-244.

Price, R., D. Neiger, and G. Shanks. (2008b) Developing a Measurement Instrument for Subjective Aspects of Information Quality, Communications of the Association for Information Systems (CAIS), 22:3, Jan. 2008, p49-74.

Price, R. and G. Shanks. (2005) A Semiotic Information Quality Framework: Development and Comparative Analysis, Journal of Information Technology (JIT), 20:2, p88-102. (2005).

Shanks, G. and E. Tansley (2002) Data Quality Tagging and Decision Outcomes: An Experimental Study, *Proc. IFIP Working Group 8.3 Conference on Decision Making and Decision Support in the Internet Age*, Cork, July, p399-410.

Wand, Y. and R. Wang (1996) Anchoring Data Quality Dimensions in Ontological Foundations, Communications of the ACM, 39:11, p86-95.

Wang, R. and D.M. Strong (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, 12:4, p5-34.