Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2014 Proceedings

IS THERE STILL A NEED FOR MULTIDIMENSIONAL DATA MODELS?

Tom Hänel
TU Bergakademie Freiberg, Freiberg, Saxony, Germany, tom.haenel@bwl.tu-freiberg.de

Michael Schulz
Philipps University Marburg, Marburg, Hesse, Germany, michael.schulz@wiwi.uni-marburg.de

Tom Hänel and Michael Schulz, 2014, "IS THERE STILL A NEED FOR MULTIDIMENSIONAL DATA MODELS?", Proceedings of the European Conference on Information Systems (ECIS) 2014, Tel Aviv, Israel, June 9-11, 2014, ISBN 978-0-9915567-0-0
http://aisel.aisnet.org/ecis2014/proceedings/track04/5

IS THERE STILL A NEED FOR MULTIDIMENSIONAL DATA MODELS?

Complete Research

Hänel, Tom, TU Bergakademie Freiberg, Freiberg, Germany, tom.haenel@bwl.tu-freiberg.de

Schulz, Michael, Philipps University Marburg, Marburg, Germany, michael.schulz@wiwi.uni-marburg.de

Abstract

Organizational and technical changes challenge standards of data warehouse design and initiate a redesign of contemporary Business Intelligence and Analytics environments. As a result, the use of multidimensional models for performance oriented reasons is not necessarily taken for granted. Simple data models or operational structures emerge as a basis for complex analyses. The paper therefore conducts a laboratory experiment to examine from a non-technical perspective the influence of different data modeling types on the representational information quality of end users. A comparison is made between the multidimensional model and the transactional model respectively the flat file model. The experiment involves 78 participants and aims to compare perceived and observed representational information quality aspects of ad hoc analyses regarding the data modeling type. The results indicate a higher observed quality for multidimensional modeled data, while different types of data models do not influence the end user perception of the representational information quality.

Keywords: Business Intelligence, Business Analytics, Data models, Experiment.

1 Introduction

Business Intelligence (BI) systems process data they receive by an integration of various data sources to provide business insights for decision making (Hamilton, 2009). An important aspect of the data integration is the data modeling, which eventually affects the analysis opportunities for end users (Ballard et al. 2006). Data models need to illustrate business domains according to user requirements in order to make accurate decisions (Moody, 2005). Multidimensional models have thereby gained acceptance in BI contexts (Chaudhuri et al., 2011). The data therefore have to be prepared and they are typically stored in an integrated data store, namely the data warehouse (Sen and Sinha, 2005). The modeling and transformation of data in data warehouse contexts is costly, so that alternatives are under discussion due to current technical and organizational changes of BI environments. Such changes results in an extension of analytical functions and user groups. Business users recently are able to access manifold information sources and analytical options (McAfee and Brynjolfsson, 2012). This leads to a discussion about changes on the architectures of analytic tools and questions distinct data modeling types. For instance, a replacement of multidimensional models by simpler data structures seems to be a beneficial strategy to avoid additional efforts, especially to establish and maintain data transformation processes. Therefore, this paper's goal is to investigate whether the quality of an analysis will suffer from changing the underlying data model. Thereby, this research examines the extent to which a data modeling type influences the representation of information perceived by an end user performing analysis tasks as well as the actual impact observed by impartial measures. The research question is whether multidimensional data models are still needed and beneficial for a decision making in context of ad hoc analysis.

The discussion about the representational information quality on decision tasks was mainly completed twenty years ago, but the consideration of differences in data modeling requires an investigation from a new perspective. We consider multidimensional models, transactional models and flat file models, which are commonly used data modeling types in context of BI and Analytics (Sen and Sinha, 2005; Cios et al., 2007). Previous studies investigate multidimensional and transactional models with a dominant focus on the database design perspective (Corral et al., 2006; Dowling et al., 2001; Schuff et al. 2011; Schuff et al. 2005). Existing usability studies are rather technical driven than application oriented (Vujošević et al., 2012) so that they are less conclusive to estimate the impact of a certain data model on the decision quality of end users from a non-technical perspective. However, especially an end user needs to understand new approaches in order to assess a potential usefulness of advanced analytic forms. This is in particular necessary to substantiate subsequent implementations. The paper contributes therefore to a discussion about data models and new analytical trends. This discussion will gain insights regarding the understandability and the expressiveness of certain data models in an ad hoc analysis context. The paper's arguments address practitioners and researchers from a database design and also focus on the non-technical perspective of end users.

Section 2 discusses technological and organizational reasons for analytical advances of BI to show the current relevance of the topic. In the following, we point out the relevance for analysis of the considered data models. The paper refers to laboratory experimentation, because this approach allows quality evaluations by independent participants (Moody, 2005). The research model is presented in Section 3. This includes a formulation of hypotheses in terms of an observed and a perceived representational information quality. Section 4 introduces the participants and the data collection procedure of the experiment. We describe the software environment and the constraints that guided our experimental design. In Section 5, the collected data are analyzed so that the explanatory power of the research model can be evaluated. The paper concludes with a discussion of the results and an outlook on further research activities.

2 Status Quo

2.1 Problem refinement

Since the last decade, the term *BI* is under scrutiny and the term *Business Analytics* is becoming more important. Statistical analysis is emphasized more in Business Analytics than in BI, but a separate consideration of the terms is not beneficial (Davenport, 2012). Therefore, we follow the arguments of Chen et al. (2012) to use BI and Analytics (BI&A) as a unified term.

The debate on a common term indicates a continuous change of business analysis over time as well as a progressive emergence of new requirements. An increasing amount of data generated by a large number of different data sources is available to organizations intending to use the data to learn more about their business (McAfee and Brynjolfsson, 2012) and to create competitive advantages. The focus of analysis has shifted partially from the support of strategic and tactical decisions to operational issues (Böhringer et al., 2010). This requires detailed and recent information and also leads to new user groups with different demands and not only to traditional users of BI&A tools. If standard reports do not cover such information requirements, users are recently able to create their own analysis. Related trends like self-service BI are driven by an accelerating change of market demands for data analysis and provisioning (Evelson, 2012). Technical restrictions diminish due to advances in hardware, like large main memory, parallelized hardware platforms, solid state disks (van der Lans, 2012) and faster networks. New software architectures are discussed for analysis environments, in particular in-memory databases (Plattner, 2009) and cloud computing (Baars and Kemper, 2010).

As a consequence of the arguments mentioned above the research problem is that organizational and technical changes challenge existing standards of data warehouse design and initiate a redesign of existing BI&A environments. Van der Lans (2012) for example, proposes a data virtualization to unify current data from various sources. Plattner (2009) and Loos et al. (2011) discuss a possible combination of transactional and analytical databases in that context. Considering further big data analytics, organizations have to deal with data that is in general not structured according to traditional data models. Faced by such proceedings, performance oriented reasons based on software and hardware issues for special data modeling of analytical environments (Codd et al., 1993) seem to be insufficient. In line with this, research findings point out that technical considerations of BI&A software are rather secondary from an end user's perspective. The intention to use an analytical system depends to a higher degree on the suitability of information to fulfill existing information requirements (Popovic and Jaklic, 2010). Especially representational information quality is recognized as one of the most important factors to create a source for BI&A (Nelson et al., 2005).

The goal of our research activities is therefore to examine how different data models affect the representational information quality in analytical tools from the non-technical viewpoint of end users. This non-technical perspective assumes that users do not necessarily need experiences with data modeling. They access data structures via a user interface to answer questions based on the information presented. In particular, we want to investigate a perceived representational information quality in relation to an observed representational information quality. The perceived quality is a subjective assessment of an end user about the information presented. The observed quality concerns impartial measurable indicators of an analysis result. This distinction allows us to examine whether the actual unbiased quality of the analysis corresponds to the evaluation of the participants or if there is a discrepancy. The outcome of this investigation renders assistance to justify the creation of special data models for analytical activities.

2.2 Data models and their relevance for analysis

We focus on the creation of ad hoc reports in order to examine the refined problem by an experiment so that we are able to obtain reliable results with a reasonable anticipated number of participants. Multidimensional models are popular to support the analysis of managerial issues. Therefore, Online Analytical Processing (OLAP) provides operations in order to filter, aggregate, pivot, rollup or drilldown (Chaudhuri et al., 2011). The concept was introduced by Codd et al. (1993) and originates from the inadequateness of Online Transaction Processing (OLTP) systems to represent decision relevant information according to the analysis needs of business users. The reason is that OLTP systems are intended to support operational applications commonly conceptualized by normalized transactional models (Datta and Thomas, 1999). However, advances in technology like in-memory (Plattner, 2009) and the increasing discussion of big data analytics (Russom, 2011) have shifted the analysis focus beyond the OLAP concept. Arguments considering advantages in processing time do not have much relevance anymore. Direct access to transactional databases supersedes a need for transformation processes and a redundant storage in OLAP and OLTP systems (Plattner, 2009). Benefits of such a strategy are lower costs for data processing through the omission of data cleansing and structuring, which leads to time savings and a more up to date decision making.

Multidimensional and transactional models are common design techniques in data analysis contexts (Sen and Sinha, 2005). However, the creation of such a structure can be very complex and changes of displayed information are time consuming. This justifies the question whether a comparable representational information quality can be achieved by using simpler structures like flat files ignoring structured relationships of data (Cios et al., 2007). In context of big data, software vendors (e.g. SAS, 2013) already offer similar products without fixed table structures. Flat file models can be created easily, even if the data is stored in different sources.

Transactional and flat file models represent solely metadata so that actual data values are only visible in corresponding reports. A multidimensional model allows for a representation of data and metadata on the model level. This structure assumes a hierarchical data organization so that measures can be summarized at all hierarchy levels by different rules (Boehnlein and Ulbrich-vom Ende, 1999). A hierarchy value is an attribute or a combination of various attributes, e.g. first name and surname. Such information items are usually not combined in transactional models and flat file models. The multidimensional model provides the highest degree of structuring from the perspective of an end user. Transactional models are also very structured, but usually from a more technical point of view. The structure of a flat file model can only be derived at the level of a single record. Figure 1 shows a simple example to visualize the different modeling approaches from an end user perspective.

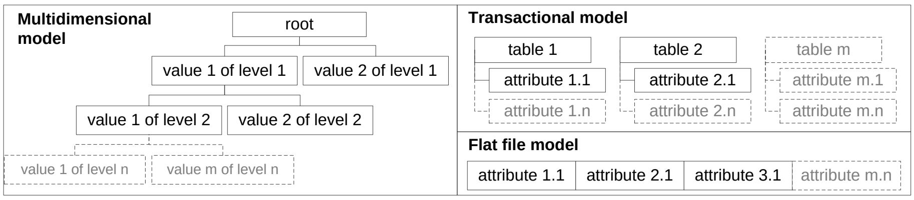

Figure 1. Examples of data model representation.

2.3 Related studies

The majority of the previous studies focus on design issues of transactional and multidimensional models. Flat file models are not considered by any previous study. There are experiments dealing with a content reproduction depending on a data modeling type (Dowling et al., 2001; Corral et al., 2006). The findings confirm beneficial effects of multidimensional structures for a data model comprehension. Jones and Song (2005) explore positive impacts on modeling time and correctness by dint of dimensional design patterns. Further studies compare the understandability of transactional and multidimensional models in context of content identification and modification (Schuff et al., 2011; Schuff et al., 2005). A preliminary study reveals that inexperienced users understand multidimensional models easier, while the results did not achieve significance (Schuff et al., 2005). However, an additional experiment neglects a better user understanding in context of multidimensional models (Schuff et al., 2011). The findings suggest the consideration of technical aspects for choosing either multidimensional or transactional models.

The most recent study investigates the usability of multidimensional and transactional models in an analysis oriented context (Vujošević et al., 2012). The focus shifts to the use of data models to facilitate decision support in a business context. The study reveals higher accuracy values, faster completion times, as well as a better holistic view and subjective perception in favor of the usability of multidimensionally modeled data. However, the underlying software prototype reflects rather a technical perspective and provides no analysis interface for an end user.

Technical limitations have left no room to investigate the impact of a specific prepared data structure on aspects of information representation. This constrained studies on the representation of an underlying problem. Overall, we reviewed about 100 articles on the representational information quality in decision-making situations. More than 80 percent were published in the 1980s and 1990s. Previous experiments focused on a comparison of tabular and graphical views (e.g. Anderson and Mueller, 2011; Ghani and Lusk, 1982; Vessey, 1991), a differentiation of graphical representations (e.g. Jarvenpaa, 1989; Tan and Benbasat, 1993; Dull and Tegarden, 1999) or an analysis of advanced forms of presentation (e.g. Bharati and Chaudhury, 2004; Cao et al., 2009).

All these studies mentioned above address an end user perspective, which means that their research design can guide our research intention, too. Five of the selected experimental studies are limited to observed values (Anderson and Mueller, 2011; Dull and Tegarden, 1999; Ghani and Lusk, 1982; Jarvenpaa, 1989; Tan and Benbasat, 1993), while two contributions took an end user perception into account (Bharati and Chaudhury, 2004; Cao et al., 2009). Next to various observed values, an accuracy measure was always considered. Four experiments recorded the time for problem solving (Cao et al., 2009; Ghani and Lusk, 1982; Jarvenpaa, 1989; Tan and Benbasat, 1993).

3 Research Design

The independent construct of our research model is the data modeling type (DMT) that influences the quality of an ad hoc analysis. We differentiate between a perceived representational information quality of an end user and an observed representational information quality (cf. Figure 2).

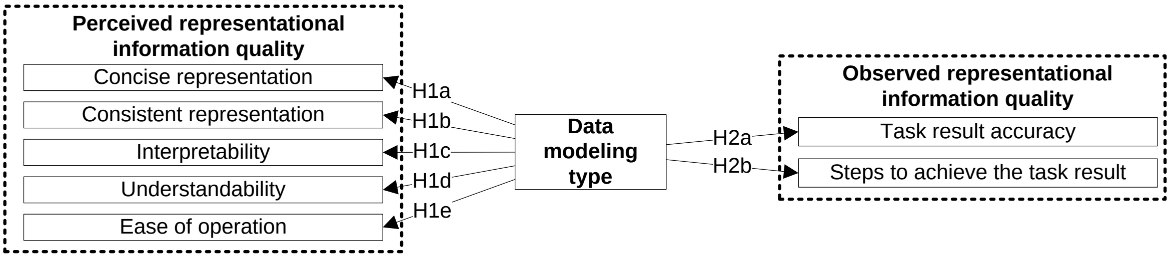

Figure 2. Research model.

The perceived representational information quality considers to what extent a DMT leads to information according to end user needs. The information quality a system produces is distinguishable into intrinsic, contextual, representational and accessibility information quality (Lee et al., 2002). In order to achieve comparable results, different DMTs have to consider a consistent business context, content and accessibility for any end user. Therefore, the representational information quality is suitable for an investigation of the perceived analysis quality influenced by a given DMT. This means in particular how a data structure provides the information to the end users.

Lee et al. (2002) distinguishes representational information quality into concise and consistent representation, interpretability, understandability as well as ease of operation. Concise representation concerns how compact the information is provided. A representation should encapsulate the main aspects precisely to the point in order to avoid overwhelming and unnecessary information. A further quality attribute is consistent representation. Information should be presented in a coherent and invariant format. Appropriate units, definitions or labels have to characterize information in terms of interpretability. Another representational information quality aspect is understandability. This considers unambiguousness and comprehensibility of information. Finally, ease of operation means that information needs to be managed and manipulated easily. We formulate the following hypotheses in context of the perceived analysis quality by the end users:

- H1a: The DMT affects the concise representation of information perceived by an end user.
- H1b: The DMT affects the consistent representation of information perceived by an end user.
- H1c: The DMT affects the interpretability of information perceived by an end user.
- H1d: The DMT affects the understandability of information perceived by an end user.
- H1e: The DMT affects the ease of operation of information perceived by an end user.

Apart from the perceived quality, we also need to introduce measures for an observed quality. According to the findings of Vujošević et al. (2012), the task result accuracy and completion time are suitable measures in context of ad hoc analyses. These measures were also used in other experiments in the field of representational information quality. However, the documentation of times for a task performance is associated with difficulties. A time measurement by the end user is not impartially verifiable and therefore a potential source of error. It is furthermore plausible that users are influenced by time tracking, e.g. by pressure or distraction. Such an approach excludes as well the possibility to conduct experiments online and implies additional efforts. A system side time recording neglects breaks of users and interruptions during the task performance. Hence, we use the steps an end user needs to achieve the task result. This measure is expressed by the number of task related clicks, which are logged during the task performance. Consequently, we formulate two hypotheses for the observed analysis quality:

- H2a: The DMT affects the observed task result accuracy of an ad hoc analysis.
- H2b: The DMT affects the observed steps of an end user to achieve the task result of an ad hoc analysis.

A support of the hypotheses indicates differences of the observed and perceived representational information quality for different DMTs. We expect an increasing representational information quality with an increasing degree of structuring. This would be in line with existing studies of Popovic and Jaklic (2010) or Vujošević et al. (2012). A missing support suggests that different DMTs do not influence an end user in any way. This would advance the findings of Schuff et al. (2011) from the data base design to an end user perspective.

4 Experimental Design

We decided to use an experimental design of independent measures (McLeod, 2007), in which each design option needs a separate group of participants. Three groups are required in our study, since we intend to compare transactional, multidimensional and flat file structures. 20 participants are considered as a minimum number for each group (Hair et al., 2006). The experiment was conducted online from June until July 2013. We asked 600 undergraduate students of business administration from Germany to participate. Experts are familiar with multidimensional structures, because this has evolved into a de facto standard for analysis (Chaudhuri et al., 2011). Thus they may be affected by previous experiences. Ghani and Lusk (1982) demonstrate a decision makers' preference for an initially used representation of information, so that efforts are required to learn new representation patterns or to adapt heuristics. Therefore, we performed the experiment among students without prior knowledge in the use of BI&A systems to achieve homogeneous results from unbiased participants.

The students were requested via e-mail to participate voluntarily in the experiment. One reminder was sent after 14 days to those who did not respond to the first e-mail. The participation allowed the students to take part at a raffle of vouchers. The students were informed that the experiment intends to study the representational information quality of certain data structures and not to evaluate personnel skills in dealing with software. All results were saved anonymously. The survey was accessed by 197 persons. 93 students participated, which is a response rate of 15.5 percent. Nine participants stopped processing before enough information for evaluation was collected. Incomplete responses were eliminated. Response sets with an unusual short processing time and apparent fill out patterns are also not considered in the evaluation. The resulting data set includes 78 records, in equal parts male and female participants. The average age is 24.

Each participant was confronted with a software user interface that was specifically programmed for this experiment. It includes a toolbar as well as a data source and a working area (cf. Figure 3). No vendor-specific characteristics were displayed so that the participants could understand the system quickly. The software served the needs to understand efforts and usability aspects in data modeling. Therefore, technical considerations played only a minor role in the choice of the data model. The software modeled a reporting tool of a sales scenario. The working area allowed a creation of reports in a simple grid form. The participants needed to click on the attributes or hierarchy elements in the data source. The working area displayed grouped attributes and aggregated measures. A selection of a particular expression within the working area set a report filter. The filter expression appeared in the toolbar. The toolbar had two buttons to clear the working area and to insert/delete a total column sum. Any column could be sorted in ascending or descending order by clicking on the column name.

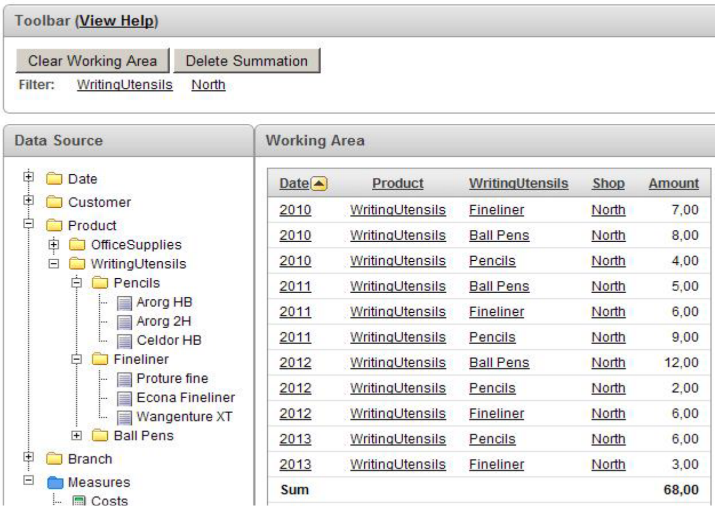

Figure 3. User interface for the multidimensional structure.

The specific characteristics of the DMT options were represented by the data source, which was the only difference of the three user interfaces. The left hand side of figure 3 is thereby an example for the structure of the multidimensional model. Figure 4 illustrates and compares additionally the structures of the three different DMTs.

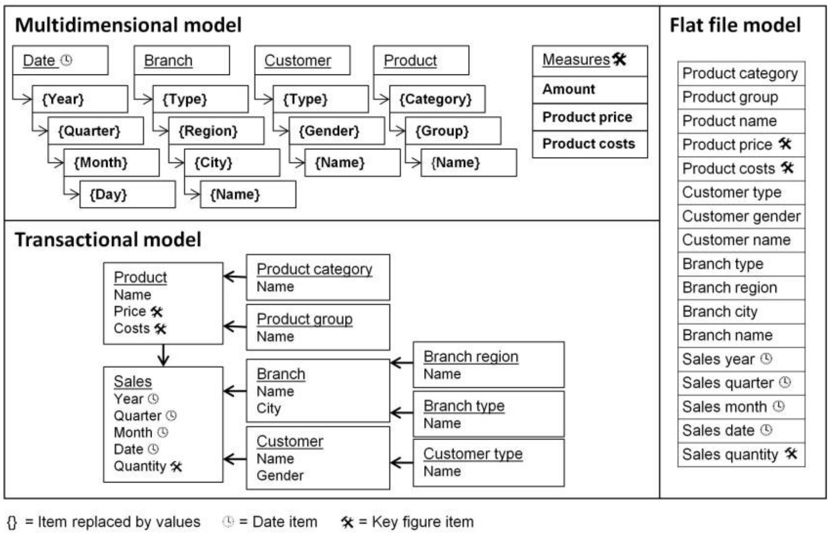

Figure 4. Data structures used in the experiment.

All DMT options used the same sample data set and data granularity. We chose a sales scenario because this is a typical data warehouse issue (see, for example Kimball et al., 1998). The number of attributes, metrics, and data sets were chosen so that they could be understood by the participants in a reasonable time. In the multidimensional case, the data source included hierarchies for time, customers, products, geography and a folder for measures. The transactional data source consisted of several folders with the attributes of entities and relationship tables, while the flat file structure only listed the attributes and metrics. We made simplifying modeling decisions to measure only the differences in the handling of the data structures. We chose field names that identify the contents clearly and did not use calculated measures. The date field was divided into year, quarter, month and day. The fields were thematically grouped and no technical keys were displayed. If appropriate, individual attributes were combined into values.

A participant was randomly assigned to a DMT option. Using a specific DMT, the participants had to solve ten tasks in the course of the experiment. The tasks could be solved clearly and represented typical questions to a BI&A system in the context of a sales scenario. One task asked for example: *How many franchise stores are in Hamburg?* Another task was formulated as: *Which female customers of customer type B bought pencils on 2012-01-23?* The examples show that the tasks have differed in their complexity. A higher complexity was created by longer and nested tasks, by more attributes and metrics that had to be used to solve the task and by the need to use filters and summations. The tasks were equivalent for each DMT. The result for each task had to be written in a text box in the task area. Tasks could be left out. After answering all the tasks the participants were asked to assess the information content of the data structure (not of the reports).

5 Findings

In line with our research question and the singularity that the multidimensional model has a wide awareness, we compare this data modeling type to the flat file model and to the transactional model. The results support H2b and partly H2a. The observed quality measures behave in favor of the multidimensional model. The hypothesis H1b remains unsupported. H1a, H1c, H1d and H1e are supported only partly, so that an influence of the differing DMTs on the perceived representational information quality cannot be certainly confirmed from a statistical point of view. The detailed results are described and analyzed in the following.

5.1 Impacts on observed representational information quality

The participants had to solve ten tasks that can be divided into four result groups. We measured the observed quality of each task by a system side record of steps needed by a user to achieve the result and a comparison against the correct result. A correct result is marked with one point, while zero point is the rating for incorrect, incomplete or nonexistent results.

We calculated the means of the observed quality measures for each task across the users (cf. Table 1). The users of the multidimensional model achieved the best results on average viewing all tasks. The accuracy is however only statistically significant in comparison to the flat file model. In contrast, the single tasks lead to a differentiated consideration.

Task 1 and 10 asked for a specific list of attributes (customers, products). Thereby, the number of steps to achieve the result is significantly less for the multidimensional model in comparison to the other models. Users needed on average half as many steps in both tasks and achieved a clearly higher accuracy in task 1. The transactional model lead to the most correct answers in task 10, but the difference is not statistically significant in comparison with the multidimensional model.

Task 2 and 3 required the counting of stores in a city and the number of a group of customers. Users of the multidimensional model needed again the fewest steps to achieve their results. The participants using this model had also the most correct results. However, this mean is not significant in task 3.

The volume of sold products was the subject of three tasks (no. 4, 5 and 7). Only in 2 tasks, a significant result could be achieved for the number of steps. The participants using the flat file model needed the fewest number in task 4 compared to the other data modeling types. The user of the multidimensional model achieved the result faster in task 7.

In three tasks (no. 6, 8 and 9), the participants had to determine customers, sales regions and stores for a maximum or a minimum value of sold products. Statistically significant differences between the models were found in task 8 and 9 by comparison of the multidimensional and the transactional model. Thereby, users of the transactional model achieved more correct result in task 8 and a worse accuracy in task 9. The flat file model allowed the fewest steps for users to the result also in task 9.

Task	Result group	Observed measure	Multidimens. model	Comparison to flat file model		Comparison to transactional model	
			Mean	Mean	Sig. (2-tailed)	mean	Sig. (2-tailed)
1	Attribute listing	Steps	**5.00**	10.58	0.000	11.35	0.001
		Accuracy	**0.96**	0.73	0.021	0.77	0.043
2	Attribute counting	Steps	**2.65**	8.88	0.000	9.88	0.000
		Accuracy	**0.92**	0.69	0.035	0.46	0.000
3	Attribute counting	Steps	**3.42**	7.96	0.000	9.00	0.000
		Accuracy	**0.88**	0.85	0.692	0.73	0.166
4	Quantitative value selection	Steps	5.69	**3.15**	0.093	3.88	0.223
		Accuracy	0.96	**1.00**	0.327	0.96	1.000
5	Quantitative value selection	Steps	9.04	**8.08**	0.518	8.46	0.694
		Accuracy	0.73	0.62	0.385	**0.81**	0.520
6	Attribute selection according to a quantitative value	Steps	**10.58**	11.15	0.828	11.65	0.678
		Accuracy	0.62	0.62	1.000	**0.77**	0.238
7	Quantitative value selection	Steps	**5.23**	8.73	0.000	9.42	0.000
		Accuracy	0.92	0.81	0.231	0.85	0.395
8	Attribute selection according to a quantitative value	Steps	9.08	9.65	0.733	10.73	0.313
		Accuracy	0.81	0.88	0.452	**1.00**	0.022
9	Attribute selection according to a quantitative value	Steps	8.88	**6.00**	0.020	6.50	0.045
		Accuracy	**1.00**	0.92	0.155	0.88	0.077
10	Attribute listing	Steps	**6.00**	13.38	0.000	11.54	0.000
		Accuracy	0.92	0.85	0.395	**0.96**	0.561
	Average of all tasks	Steps	**6.56**	8.76	0.001	9.24	0.000
		Accuracy	**0.87**	0.80	0.027	0.82	0.145

*Table 1. List of tasks and mean comparison for the observed quality measures. The best result for each row is shown in **bold**, non-significant values are shown grayed.*

Table 2 summarizes the properties of the final reports, i.e. the representations which were displayed before the participant entered the results of the ten tasks. The users of the multidimensional model required fewer columns and fewer filters on average in comparison to the other two models. A summation was most frequently used in the multidimensional model. Between the flat file model and the transactional model, only small differences were measured.

	Multidimensional Model	Flat file model	Transactional model
Columns used on average	1,98	3,08	3,11
Reports in which one or more filters were used	8,08%	77,69%	78,46%
Reports in which one or more summations were used	14,61%	6,54%	7,69%

Table 2. Properties of the final reports.

5.2 Impacts on perceived representational information quality

We used items according to Lee et al. (2002) to compare the perceived representational information quality of the data models. Thereby, we have not considered such items which were determined in Lee et al. (2002) by reverse coding techniques. A seven-point Likert scale was used to rate each item statement including response options from ‘strongly disagree’ (1) to ‘strongly agree’ (7). The Likert scale design is chosen based on measurement instruments of information quality (Lee et al., 2002) and conceptual models (Maes and Poels, 2007). Table 3 shows the mean comparison of the rated item statements and the statistical significance. Thereby significant results arise only for less than half of the used items. The differences of the means are smaller than in case of the observed quality measures. The perceived representational information quality of the users tends always to an over-average or high degree for each data modeling type. The lowest mean is 4.8 and the highest averages 6.3.

	Item	**Multidim. model**	**Comparison to flat file model**		**Comparison to transactional model**	
		Mean	Mean	Sig. (2-tailed)	mean	Sig. (2-tailed)
Concise representation	This information is formatted compactly in the database.	5.8	5.5	0.284	**6.1**	0.202
	This information is presented concisely in the database.	5.7	5.8	0.911	**6.0**	0.263
	This information is presented in a compact form in the database.	**6.1**	5.5	0.068	5.7	0.115
	The representation of this information is compact and concise in the database.	**5.9**	5.3	0.068	6.0	0.786
	Average concise representation	5.9	5.5	0.133	**6.0**	0.682
Consistent representation	This information is consistently presented in the same format in the database.	5.9	5.7	0.655	**6.0**	0.745
	This information is presented consistently in the database.	5.9	5.6	0.472	**6.3**	0.241
	This information is represented in a consistent format in the database.	5.5	5.9	0.326	**6.0**	0.161
	Average consistent representation	5.8	5.7	0.931	**6.1**	0.199
Understandability	This information of the database is easy to understand.	5.4	5.7	0.519	**6.1**	0.046
	This information of the database is easy to comprehend.	5.2	5.3	0.866	**5.7**	0.183
	The meaning of this information is easy to understand.	5.6	**5.9**	0.228	**5.9**	0.268
	Average Understandability	5.4	5.6	0.507	**5.9**	0.092
Interpretability	It is easy to interpret what this information means.	5.2	**5.8**	0.061	5.7	0.133
	This information is easily interpretable.	5.4	**6.1**	0.053	5.9	0.175
	The measurement units for this information are clear.	4.9	**5.1**	0.618	**5.1**	0.691
	Average Interpretability	5.2	**5.7**	0.053	5.6	0.206
Ease of operation	This information is easy to manipulate to meet the user’s needs.	5.0	5.3	0.447	**5.8**	0.034
	This information is easy to aggregate.	4.8	5.2	0.467	**5.7**	0.014
	This information is easy to combine with other information.	5.5	6.0	0.355	**6.1**	0.152
	Average ease of operation	5.1	5.5	0.351	**5.9**	0.023

Table 3. List of items and mean comparison for the perceived quality measures. The best result for each row is shown in ***bold****, non-significant values are shown grayed.*

The concise and consistent representation is perceived at the highest average for the transactional model, while the averages are not statistically significant. The single means range from 5.7 to 6.1 in context of concise representation. However, only two items are statistically significant. The compact form and representation of information is thereby perceived more concisely by users of the multidimensional model than by users of the flat file model.

In terms of consistent representation, the user perception ranges from 5.5 to 6.3. The means of all items are highest for the transactional model, but none of the values has statistic significance. The values of the multidimensional model are in between those of the flat file and the transactional model in the context that the information is presented in the same format and consistently. In terms of the representation in a consistent format, the multidimensional model has the lowest mean compared to the other models.

The transactional model is perceived as most understandable and thereby the average and the single mean for ease of understanding are significant in comparison to the multidimensional model. The interpretability is perceived significantly higher by users of the flat file model in comparison to the multidimensional model. Only the clearness of the measurement units achieves no statistical significance. The transactional model is perceived as easiest to operate. The values are significant excluding the aspect information combination.

The values range from 5.2 to 6.1 in context of understandability, from 4.9 to 6.1 in terms of interpretability and from 4.8 to 6.1 regarding the ease of operation. Thereby, the multidimensional model is perceived as least understandable, interpretable and easy to operate compared to the other models. However, statistical significance is achieved only by comparing the multidimensional model to either the flat file or the transactional model. A statistically significant comparison of the perceived representational quality regarding to the transactional and the flat file model is not possible.

5.3 Discussion

The goal of the experiment is to find out whether different data models affect the analysis quality of end users. The need of classical multidimensional models is questioned. Thereby, the experiment investigates if models that can be made more readily available provide comparable or better insights into data with the same or a better level of intuitiveness. We consider common data modeling types by a comparison of a multidimensional model with a flat file model and a transactional model. The focus is on the representational information quality. The experiment addresses the effectiveness and the quality in the processing of tasks and the perceived assessment of the participants.

The experiment was conducted among students who do not have experience in data analysis. With this group of participants, it was possible to carry out the quality independent of existing experiences. For a long-term orientation of the BI&A strategy, this is more important than the short-term advantage of models that are known by the use in the past. In recent years, new user groups are able to use BI&A systems to solve complex problems. This leads to new data requirements, so that data structures need to be extended in favor of user-friendly and analysis-oriented data structures. However, models with less structure are increasingly used for the analysis yet.

Data modelers cannot directly influence the subsequent usage of data models by a non-technical end user, but an incorrect and inconsistent sharing of data has to be avoided. Decision-makers should be able to use a representation of data without adaptations to specific needs (Larcker and Lessig, 1980). The hierarchies in a multidimensional model represent the logical data structure. This helps to identify relevant information quickly. In transactional models, the relationships are defined, but they are not represented in a form that is understandable to a non-technical end user. In flat file models, the relationships are only available at the level of a single record. This increases the risk that relevant information is not considered and analysis options are not recognized (Corral et al., 2006). Therefore, less structured data sources from the perspective of an end user require a high level of domain knowledge.

Multidimensional models usually entail costs for data processing. There is a potential to save these costs of data transformation for analysis purposes in case of flat file models and transactional models.

However, the experiment indicates a higher reporting complexity and higher probability to make wrong decisions in case of using these models (H2a, H2b). This is associated with a danger to eat up the potential savings of data processing or even to incur losses.

In contrast to previous studies, our approach is responsive to end users, who actually perform ad hoc analyses with a certain user interface. Although we consider another perspective, the results confirm that the multidimensional model is more appropriate for analysis purposes than the transactional model. This is in line with four previous studies (Corral et al., 2006; Dowling et al., 2001; Jones and Song, 2005; Vujošević et al., 2012). Flat file models also have weaknesses compared to the multidimensional model.

Vessey (1991) argues that the problem representation has to be suitable for the task to be fulfilled. This does not seem to apply to the representation of the data source: In our experiment, the different numbers of steps to get a result and the task accuracy cannot be attributed to a task type or the complexity of a task. However, the observable quality measures indicate advantages of the multidimensional model considering all tasks together. The results regarding the properties of the final reports show that the reports that have been built using the multidimensional model contain overall fewer elements and have a lower complexity. This is one reason why fewer mistakes are made using the multidimensional model.

We have not expected that only a few differences are perceived by the participants of the experiment (H1a – H1e). Bharati and Chaudhury (2004) also find out that a change in the representation does not result in a change of end user satisfaction. A possible explanation is that the participants are not experienced to judge whether an achieved result is correct or not.

5.4 Limitations and future research

One experiment can hardly consider all aspects, which can be usefully tested in the context of representational information quality of different data models. This experiment discusses a problem area that can be investigated by further studies. Therefore we have conducted a small-scaled experiment to provide a basis for a meaningful interpretation. The participants were not overwhelmed by a large information source and they were able to perform the tasks in a timely manner. Further investigations can enhance our experimental design to examine the effects of different information loads in conjunction with the various data structures.

We chose a simple model to assign the results clearly to their source. The participants of the experiment are students, who are in the same phase of their studies. This is an attempt to keep individual differences as small as possible. Ghani and Lusk (1982) have considered this issue in more detail by typology of the personality of the participants. In future studies, such aspects should be considered.

A laboratory experiment has limitations to prove a general practicality. Therefore, we cannot bring forward the argument that decision makers benefit more from the use of multidimensional models in comparison to flat files or transactional models. However, our results point out that the model used for the data analysis is important for the quality of a derived decision. Thereby, the laboratory experiment leads us to the opportunity to control the confounding variable of unequally distributed experiences of end users.

In general, students are not familiar with the preparation of business reports. Therefore, we chose the tasks so that a substantial understanding is very simple and cognitive differences do not have any relevance. The focus is on the preparation of the reports. However, the inexperience of the participants is not negligible. Further research activities should examine how the results will change when experienced users are asked. This includes situations, in which problem domains are already known beforehand or decision tasks are used that require not only simple extractions of information but decisions under uncertainty.

6 Conclusion

Companies have more data available than ever before (McAfee and Brynjolfsson, 2012). However, this is only an advantage for analysis activities if the data is presented in a structure so that correct information can be drawn easily. The paper's contribution takes the usefulness of data structures and the viewpoint of end users of a BI&A system into account in order to enhance a discussion about organizational and technical aspects in context of integrating new analytical systems. We have conducted an experiment that compares the representational information quality of multidimensional models, transactional models, and flat file models.

The observed representational information quality was measured by the task result accuracy and the steps to achieve the task results. Although newer data models have relevance for specific issues (e.g. analyzing data without internal structure), we have shown that the omission of technical restrictions does not mean that a particular analysis-oriented data preparation and a dimensionally oriented representation is no longer necessary. The less structured data are made available, the more possibilities exist for the users to prepare reports and to gain analysis results. This increases the complexity of report preparation and hence of the search for information. Companies must decide whether to tolerate higher costs during the processing of the data or in the course of the preparation of data for analysis purposes.

An important finding of the experiment is that end users are commonly not able to assess the representational information quality of a data source. This shows that end users of BI&A systems cannot make a decision for the appropriateness of the underlying data model. An end user's perception seems to be not suitable to select a certain analysis system. The results of a laboratory experiment cannot be readily transferred to the practice and the findings cannot be generalized without limitations. However, there is evidence that the experiment provides useful implications and that multidimensional data models are still needed. We have formed a basis, which includes further research aspects.

References

Anderson, J. C. and Mueller, J. M. (2011). The Effects of Experience and Data Presentation Format on an Auditing Judgment. Journal of Applied Business Research, 21(1).

Baars, H. and Kemper, H.G. (2010). Business Intelligence in the Cloud? In Proceedings of the Pacific Asia Conference on Information Systems, Paper 145, Taiwan, Taipeh.

Ballard, C., Farrell, D. M., Gupta, A., Mazuela, C. and Vohnik, S. (2006). Dimensional Modeling: In a Business Intelligence Environment. International Business Machines Corporation. http://www.redbooks.ibm.com/redbooks/pdfs/sg247138.pdf, Accessed 12/03/2014.

Bharati, P. and Chaudhury, A. (2004). An Empirical Investigation of Decision-Making Satisfaction in Web-based Decision Support Systems. Decision Support Systems, 37(2), 187-197.

Boehnlein, M. and Ulbrich-vom Ende, A. (1999). Deriving Initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems. In Proceedings of the 2nd ACM International Workshop on Data Warehousing and OLAP, 15-21.

Böhringer, M., Gluchowski, P., Kurze, C. and Schieder, C. (2010). A Business Intelligence Perspective on the Future Internet. In Proceedings of the Americas Conference on Information Systems, Paper 267, Peru, Lima.

Cao, Y., Theune, M. and Nijholt, A. (2009). Decision Making with a Time Limit: The Effects of Presentation Modality and Structure. In European Conference on Cognitive Ergonomics: Designing beyond the Product - Understanding Activity and User Experience in Ubiquitous Environments (p. 8). VTT Technical Research Centre of Finland.

Chaudhuri, S., Dayal, U. and Narasayya, V. (2011). An Overview of Business Intelligence Technology. Communications of the ACM, 54 (8), 88-98.

Chen, H., Chiang, R.H. and Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36 (4), 1165-1188.

Cios, K.J., Pedrycz, W., Swiniarski, R.W. and Kurgan, L.A. (2007). Data Mining: A Knowledge Discovery Approach. Springer, New York.

Codd, E.F., Codd, S.B. and Salley, C.T. (1993). Providing OLAP to User-Analysts: An IT Mandate. Codd & Associates, Ann Arbor.

Corral, K., Schuff, D. and St. Louis, R.D. (2006). The Impact of Alternative Diagrams on the Accuracy of Recall: A Comparison of Star-Schema Diagrams and Entity-Relationship Diagrams. Decision Support Systems, 42 (1), 450-468.

Datta, A. and Thomas, H. (1999). The Cube Data Model: A Conceptual Model and Algebra for On-Line Analytical Processing in Data Warehouses. Decision Support Systems, 27 (3), 289-301.

Davenport, T.H. (2012). Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data. FT Press, Upper Saddle River.

Dowling, K., Schuff, D. and St. Louis, R.D. (2001). Dimensional Data Models versus Entity Relationship Models: Does it Make a Difference to End-Users? In Proceedings of the Americas Conference on Information Systems, Paper 80, USA, Boston.

Dull, R. B. and Tegarden, D. P. (1999). A Comparison of Three Visual Representations of Complex Multidimensional Accounting Information. Journal of Information Systems, 13(2), 117-131.

Evelson, B. (2012). The Forrester Wave™: Self-Service Business Intelligence Platforms. http://www.forrester.com/The+Forrester+Wave+SelfService+Business+Intelligence+Platforms+Q2+2012/fulltext/-/E-RES71902?objectid=RES71902, Accessed 12/03/2014.

Ghani, J. and Lusk, E. J. (1982). The Impact of a Change in Information Representation and a Change in the Amount of Information on Decision Performance. Human Systems Management, 3(4), 270-278.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006). Multivariate Data Analysis. 6th edition. Prentice Hall, Upper Saddle River.

Hamilton, J. (2009). A Customer Centric Approach to Front-End-Business Intelligence Deployment. E-Business Review, 9 (1), 58-68.

Jarvenpaa, S. L. (1989). The Effect of Task Demands and Graphical Format on Information Processing Strategies. Management Science, 35 (3), 285-303.

Jones, M.E. and Song, I.Y. (2005). Dimensional Modeling: Identifying, Classifying & Applying Patterns. In Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP, Germany, Bremen.

Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W. (1998). The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses, John Wiley & Sons, New York.

Larcker, D. F. and Lessig, V. P. (1980). Perceived Usefulness of Information: A Psychometric Examination. Decision Sciences, 11(1), 121-134.

Lee, Y.W., Strong, D.M., Kahn, B.K. and Wang, R.Y. (2002). AIMQ: A Methodology for Information Quality Assessment. Information and Management, 40 (2), 133-146.

Loos, P., Lechtenbörger, J., Vossen, G., Zeier, A., Krüger, J., Müller, J., Lehner, W., Kossmann, D., Fabian, B., Günther, O. and Winter, R. (2011). In-memory Databases in Business Information Systems. Business & Information Systems Engineering, 3 (6), 389-395.

Maes A. and Poels, G. (2007). Evaluating quality of conceptual modelling scripts based on user perceptions. Data & Knowledge Engineering, 63 (1), 701-724.

McAfee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harvard Business Review, 90 (10), 60-66.

McLeod, S.A. (2007). Experimental Design - Simply Psychology. http://www.simplypsychology.org/experimental-designs.html. Accessed 12/03/2014.

Moody, D.L. (2005). Theoretical and Practical Issues in Evaluating the Quality of Conceptual Models: Current State and Future Directions. Data & Knowledge Engineering, 55 (1), 243–276.

Nelson, R.R., Todd, P.A. and Wixom, B.H. (2005). Antecedents of Information and System Quality: An Empirical Examination within the Context of Data Warehousing. Journal of Management Information Systems, 21 (4), 199-235.

Plattner, H. (2009). A Common Database Approach for OLTP and OLAP Using an In-Memory Column Database. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, USA, Providence.

Popovic, A. and Jaklic, J. (2010). Benefits of Business Intelligence System Implementation: An Empirical Analysis of the Impact of Business Intelligence System Maturity on Information Quality. In Proceedings of the European, Mediterranean & Middle Eastern Conference on Information Systems (EMCIS), Late Breaking Papers, Abu Dhabi.

Russom, P. (2011). Big Data Analytics. TDWI Research, Renton.

SAS (2013). Visual Analytics: Visually Explore Your Data for Better, Faster Insights, Create Reports and Share Results to the Web and Mobile Devices. http://www.sas.com/resources/factsheet/sas-visual-analytics-factsheet.pdf. Accessed 12/03/2014.

Schuff, D., Corral, K. and Turetken, O. (2011). Comparing the Understandability of Alternative Data Warehouse Schemas: An Empirical Study. Decision Support Systems, 52 (1), 9-20.

Schuff, D., Turetken, O. and Corral, K. (2005). Comparing the Effect of Alternative Data Warehouse Schemas on End User Comprehension Level. In International Conference on Information Systems: Third Annual Pre-ICIS SIGDSS Workshop on Decision Support Systems, USA, Las Vegas.

Sen, A. and Sinha, A.P. (2005). A Comparison of Data Warehousing Methodologies. Communications of the ACM, 48 (3), 79-84.

Tan, J. K. and Benbasat, I. (1993). The Effectiveness of Graphical Presentation for Information Extraction: A Cumulative Experimental Approach. Decision Sciences, 24 (1), 167-191.

van der Lans, R. (2012). Data Virtualization for Business Intelligence Systems: Revolutionizing Data Integration for Data Warehouses. Morgan Kaufmann, Waltham.

Vessey, I. (1991). Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature. Decision Sciences, 22(2), 219-240.

Vujošević, D., Kovačević, I., Suknović, M. and Lalić, N. (2012). A Comparison of the Usability of Performing Ad hoc Querying on Dimensionally Modeled Data Versus Operationally Modeled Data. Decision Support Systems, 54 (1), 185–197.