2009

Making money with clouds: Revenue optimization through automated policy decisions

Yoav Kolodner
Ben-Gurion University of the Negev, kolodner@bgu.ac.il

Adir Even
Ben-Gurion University of the Negev, adireven@bgu.ac.il

INTEGRATING VALUE-DRIVEN FEEDBACK AND RECOMMENDATION MECHANISMS INTO BUSINESS INTELLIGENCE SYSTEMS

(Research in Progress)

Kolodner, Yoav, Ben Gurion University of the Negev,
P.O. Box 653, Beer-Sheva, 84105, Israel, kolodner@bgu.ac.il
Even, Adir, Ben Gurion University of the Negev,
P.O. Box 653, Beer-Sheva, 84105, Israel, adireven@bgu.ac.il

Abstract

Business intelligence (BI) systems and tools are broadly adopted in organizations today, supporting activities such as data analysis, managerial decision making, and business-performance measurement. Our research investigates the integration of feedback and recommendation mechanisms (FRM) into BI solutions. We define FRM as textual, visual, and/or graphical cues that are embedded into front-end BI tools and guide the end-user to consider using certain data subsets and analysis forms. Our working hypothesis is that the integration of FRM will improve the usability of BI tools and increase the benefits that end-users and organizations can gain from data resources. Our first research stage focuses on FRM based on assessment of previous usage and the associated value gain. We describe the development of such FRM, and the design of an experiment that will test the usability and the benefits of their integration. Our experiment incorporates value-driven usage metadata - a novel methodology for tracking and communicating the usage of data, linked to a quantitative assessment of the value gained. We describe a high-level architecture for supporting the collection, storage, and presentation of this new metadata form, and a quantitative method for assessing it.

Keywords: Business Intelligence, Data Warehouse, Decision Support Systems, Metadata.

1 INTRODUCTION

Data repositories, along with the information systems (IS) utilizing them are critical organizational resources. While in the past the primary goal of managing data was to enable business operations, recent years have witnessed a transition toward extended use of data for business analysis and decision support, as firms attempt to gain competitive advantage by developing advanced data-analysis capabilities [Davenport, 2006]. Our research investigates the integration of feedback and recommendation mechanisms (FRM) into business intelligence (BI) system, which support activities such as data analysis, managerial decision making, and business-performance measurement. We define FRM as textual, visual, and/or graphical cues that are embedded into front-end BI tools and guide the end-user to consider using certain data subsets and analysis forms. The working hypothesis of our study is that the integration of FRM into BI tools will improve their usability and increase the benefits that end-users and organizations can gain from data resources.

BI involves acquisition, interpretation, and analysis of data to support managerial decision making. The software market offers a plethora of commercial platforms for supporting BI activities. Such platforms typically offer a variety of presentation capabilities (e.g., tables, charts, statistics, and advanced analytics), rapid-development utilities, and administrative tools. BI tools permit different forms of data usage such as reports, spreadsheets, OLAP (On-Line Analytical Processing), digital dashboards, and data mining. This variety of presentation and analysis forms confers the flexibility to use the same data resource for supporting different analytic tasks and to adapt the presentation style to end-users' capabilities and skills. BI solutions often use a data warehouse (DW) as an infrastructure. The DW stores historical data about past business behavior, patterns and trends, covering a broad range of business perspectives and activities. In a typical DW, datasets are imported from internal organizational IS, such as enterprise resource planning (ERP) systems [March and Hevner, 2005], and/or from external sources, such as commercial data vendors or the Internet [West, 2000]. The imported datasets are being cleansed, transformed, consolidated, and stored in a centralized repository. This DW infrastructure is then used for creating smaller databases (also known as data marts) that can accommodate different analytical needs and thus serve as a platform for supporting BI activities.

The number of firms engaged in DW/BI implementations and the volumes of data managed in DW/BI environments have grown immensely in recent years. The increasing popularity of DW/BI can be attributed to benefits such as gaining broad business coverage, leveraging data-collection investments, and shortening implementation cycles [Counihan et al., 2002; March and Hevner, 2005]. BI systems enable analytical data usage toward supporting important decisions such as evaluation of corporate strategies [Cooper et al., 2000], optimization of financial investments [West, 2000], and customer segmentation [Even and Shankaranarayanan, 2008]. Yet, exploiting DW/BI environments is challenging both technically, due to the many components and the complexity of configuration decisions involved [Shankaranarayanan and Even, 2004], and organizationally, due to the substantial managerial support and financial resources needed [Wixom and Watson, 2001]. Moreover, DW/BI design and configuration decisions are often associated with substantial cost-benefit tradeoffs [Even et al., 2006]. So far, despite the increasing popularity of the DW and the BI concept in recent years, so far these concepts have attracted only limited academic research aimed at the challenge of increasing the effectiveness of DW/BI utilization from the end-user's perspective.

A major limitation of current BI systems is that the common end-user, in search of an answer to a business question, often finds complex DW repositories too difficult to navigate for reaching the right data, and BI tools too difficult to use for answering the question. Furthermore, it is even not uncommon for end-user to know neither the right business question to ask, nor the full range of capabilities offered by DW repositories and BI tools. This limitation exists in current BI solutions, more so with sophisticated interactive tools, which offer advanced visual and analytical capabilities for dynamic and flexible investigation of data, and less so with simple static tools, which offer "snapshot" views in forms such as pre-defined dashboards, reports, or charts. The former classes of tools are geared toward addressing the needs of the data analyst and often require a high level of expertise and an in-depth understanding of the data-resource analyzed, whereas the latter are geared toward supporting the novice user. In terms of economic tradeoffs [Even at al., 2006], sophisticated interactive BI tools offer higher benefit potential, but are costlier in terms of licensing fees and learning curves, whereas simple static BI tools are easier to implement and learn, but offer limited capabilities, and hence, lower benefit potential.

We suggest that FRM capabilities can facilitate more effective and efficient navigation by helping to reveal undiscovered potential of unused data and analysis forms, and thus add business value. In the reminder of this paper, we present the concept of integrating FRM into BI tools and highlight a few possible approaches for generating them. We focus on FRM that are based on assessments of previous usage of the data resource, and the associated value gains. To generate this form of FRM, we propose a novel methodology for tracking the use of data resources, termed as *value-driven usage metadata*, which integrates in assessments of both the frequency of use and the value gained. We describe architecture for supporting the collection, storage, and presentation of this new metadata form and a quantitative method for assessing it. We then describe the design of an experiment that will test the usability and benefits of FRM integration. To conclude, we highlight the potential contributions of the new concepts that we present – the integration of FRM into BI systems, and the collection of value-driven usage metadata - and discuss directions for future research.

2 FEEDBACK AND RECOMMENDATION MECHANISMS (FRM)

In this study, we propose to integrate FRM capabilities into BI systems in a manner that would maintain simple and easy-to-learn BI functionality, while highlighting new usage directions with a high benefit potential. We define FRM as textual, visual, or graphical cues that are embedded into BI tools, providing the end-user with feedback on the actions that s/he has taken so far, and guiding him/her to consider further actions – e.g., to use certain data subsets and/or to apply certain analysis forms. Providing recommendations is a common tool used in commercial website for enhancing the end-user's experience. Such recommendations can be generated by other users, or by automatic agents, and they have been shown to have great influence on the end-user's decisions [Adomavicius and Tuzhilin ,2005]. We suggest that similar enhancement of BI systems may have important contribution to a better usage of the BI tools, and improve the decisions made.

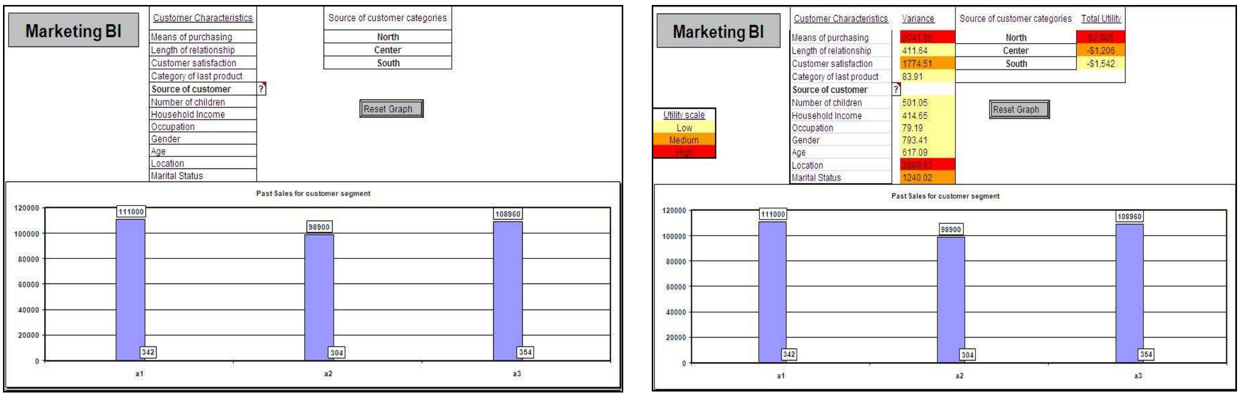

Figure 1. A Business-Intelligence Tool (Left) with FRM Added (Right).

Figure 1 offers a simplified illustration of integrating FRM capabilities into a BI tool. The original tool (on the left-hand side) lets the end-user navigate through sales data along certain dimensions (customer, location, date, etc.). This BI tool treats dimensions equally in the sense that it offers access to all dimensions and leaves navigation decisions to the user. In the FRM-enhanced version of the BI tool (on the right-hand side), navigation decisions are still left to the users, who are now provided with some additional visual cues. The cue that Figure 1 demonstrates, for example, is a color-coding that suggests giving higher navigation priority to certain dimensions. Obviously, there are other possible forms for visualizing an FRM besides color-coding, such as textual or graphical pop-up messages and side bars. Such FRM forms could indicate, in addition to the actual recommendations, the level of confidence and relevance of each recommendation based on the parameters that construct it.

We suggest that FRM, when being integrated into BI tools, can facilitate more effective and efficient navigation. This, in turn, may help revealing undiscovered potential of unused data and analysis forms thereby may increase the effectiveness of DW/BI utilization and add business value. In a preliminary assessment we have identified, at a high level, a few possible methods for constructing FRM:

(a) Value-driven usage tracking: Information resources contribute value through usage and experience. In the DW/BI context, the value can be conceptualized, for example, as an objective measure of usage success (e.g., in terms of revenue gained and/or costs saved), or subjectively via an assessment that reflects user satisfaction and willingness to pay [Ahituv, 1981]. Quantitative assessments of the value associated with the use of data resources have been applied to optimize data processes [Ballou et al., 1998], configure DW datasets [Even et al., 2007], and develop data quality metrics [Even and Shankaranarayanan, 2008]. The latter study has highlighted the importance of recognizing inequality in the value of data, suggesting that data objects (e.g., tables, attributes, and records) may vary significantly in their value contribution. Further, by evaluating a large real-life data resource, that study shows that quantitative assessments of inequality (e.g., Gini's index), have important implications for key data management decisions, such as the prioritization of data quality improvement efforts. We suggest that tracking data usage and the associated value can be used to construct FRM in BI tools, toward directing data analysis and exploration, and improving decision outcomes – this by providing users with feedback on the outcome of their own usage, as well as an opportunity to benefit from learning how other users have gained value from using the same data.

(b) Task and user characteristics: The same DW/BI environment can be used to support a plethora of business processes and tasks, each with very different data usage needs. FRM capabilities can take into account such needs, by creating either task profiles that capture specific task characteristics or by asking expert users to identify certain data elements or analysis results that are more useful and relevant for a given task. For example, referring to the illustration in Figure 1, an FRM-enhanced BI tool could direct the end-user to slice the sales data along certain dimensions (e.g., customer and product) upon designing a promotion campaign, versus focusing on other dimensions (e.g., location and date) upon optimizing distribution policies. Likewise, an FRM-enhanced tool could recommend using certain presentation styles or navigation forms that can be driven by individual preferences, e.g., presentation in textual style versus graphical visualization, or navigating data systematically and incrementally versus using heuristics. Such preferences can either be explicitly identified by asking the user to state his or her preferences or implicitly inferred via a learning process, tracking and understanding the user's interaction and identifying more-often used data and presentation formats.

(c) Data mining: FRM capabilities can also be driven by analysis of the data using algorithmic data-mining techniques. Data mining algorithms typically explore the information contained in data sources automatically and provide end-users with quantitative assessments and to enable them to extract and evaluate knowledge from this information. Data mining can suggest alternatives to decisions and actions that are about to be taken and allow users to re-consider these decisions and actions. Notably, some data-analysis tools include functionality that actively guides users through large data repositories, based on some statistical analysis of the data resource (e.g., Bissantz-DeltaMaster, http://www.bissantz.com/deltamaster). A possible drawback of using statistical analysis and data mining methods to derive FRM is the risk that the recommendations, in this case, will be based solely on the data itself, not taking into account the context in which the data is used. This limitation is addressed to an extent by the two former methods, and should be considered when further researching the use of data-mining techniques for developing FRM.

Initially, our research will focus on the first approach for generating FRM (denoted (a) above), while the exploration of other two will be deferred to later stages. This form of FRM will be based on a novel methodology for tracking the usage of data resources, described in the following section.

3 VALUE-DRIVEN USAGE METADATA

Data environments are often described as a complex manufacturing process, consisting of interconnected acquisition, processing, storage, retrieval and usage stages [Ballou et al., 1998]. These processes can be conceptualized as having two high-level stages – *data administration* versus *data consumption* (Figure 2) - each associated with different stakeholders, goals, motivations and tasks [Lee at al., 2004]. Data administration addresses technical aspects – providing the ICT capacity needed to store and process data (e.g., hardware, databases, and back-end processes), and the tools for implementing information products. Conversely, data consumption would seek to transform information products into business value, through their usage. As the typical goal of data consumers is gaining business benefits and increasing profitability, they would be more focused on the value gained by effective use of data resources and less on the technical aspects associated with managing them.

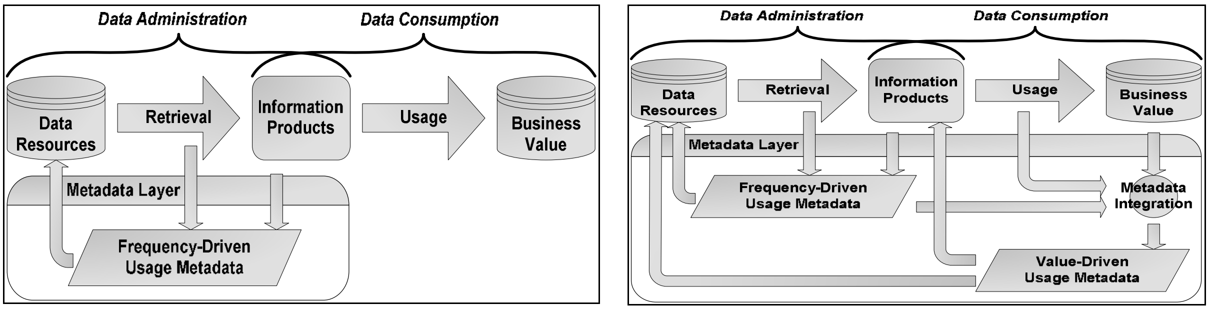

Figure 2. Frequency-Driven (left) versus Value-Driven (right) Usage Metadata

Usage metadata - tracking the usage of data objects (e.g., tables, attributes, and records) and applications - has been identified as an important form of metadata [Shankaranarayanan and Even, 2004]. Usage tracking utilities are offered by some specialized commercial solutions and, to an extent, by DBMS and BI platforms. We term the common approach implemented by today's solutions as *frequency-driven usage metadata* (the left-hand side of Figure 2). This approach is based on tracking queries and, by analyzing them, identifying the data objects being most-frequently used. The assumption underlying this approach is that frequent usage reflects higher importance. Accordingly, usage-tracking results may affect the configuration and the administration of data resources - e.g., transferring less frequently-used data to archives, and/or giving it lower quality-improvement priority.

Customers

#	Customer	Gender	Income	Children	Status	Frequency
1	Abraham	Male	High	0	Single	**1**
2	Sarah	Female	Low	1	Married	**2**
3	Isaac	Male	Medium	2	Married	**1**
4	Rebecca	Female	Low	0	Single	**1**
5	Jacob	Male	Medium	3	Married	**1**
6	Lea	Female	High	2	Married	**3**
7	Rachel	Female	Low	4	Single	**0**
Frequency		**3**	**1**	**2**	**1**	

Queries

WHERE condition	Attributes Used	Records Retrieved
Gender = 'Male' and Children > 0	Gender, Children	[3], [5]
Gender = 'Female' and Children < 3	Gender, Children	[2], [4], [6]
Gender = 'Female' and Status = 'Married'	Gender, Status	[2], [6]
Income = 'High'	Income	[1], [6]

Illustrative Example (part 1) - Assessment of Frequency-Driven Usage Metadata

To illustrate the creation of frequency-driven metadata, we use a simplified example above (extended later). The table in this example is used by marketing associates to decide which customers will be approached when promoting a new product. An associate would use a BI tool to investigate previous sale transactions, and the tool will generate queries directed to the customers tables, such as those demonstrated, to specify the subset of customers that will be targeted. Each query can be analyzed to detect which records and attributes are used to specify the selection (e.g., by parsing the WHERE clauses in the SQL statement) and, accordingly, the frequency of usage can be calculated. Frequency-driven metadata collection may provide important inputs to the data administrator, toward improving system design and prioritizing administration efforts. It is common in databases, as highlighted by the example above, that some records and attributes are accessed more frequently than others. In a larger real-world databases, this differentiation may lead to a decision to grant the more frequently used records and attributes a higher priority in terms of data quality maintenance – i.e., watch these database objects more closely, detect and correct defects, and make sure to keep them up-to-date.

While seeing the merits of collecting metadata on usage frequency for data administration, we question - does it truly address the needs of data consumers? One could argue that, to an extent, frequent usage reflects higher significance of certain data components to data consumers; hence, higher value-contribution potential. On the other hand, we suggest that frequent usage may reflect certain stagnation and a tendency to "dig into the same well" - re-using certain data subsets repetitively, while possibly ignoring unused subsets with high contribution potential. Therefore, a potential risk with basing data management decisions solely on frequency-driven metadata (e.g., due to a removal of data that is less-frequently used from the active repository into an archive) is a possible loss of opportunity to benefit from records and attributes that data consumers have neglected to use so far, which may permit new forms of data usage.

There is possibly no "clear cut" answer to this question, as it largely depends on the business context and the usage tasks. However, we suggest that important insights can be gained from tracking and considering not only actual data usage, but also the associated value gains. The benefits gained from the use of information products have been conceptualized as *utility* [Ahituv, 1980]. Utility assessments have been used to optimize the configuration of data processes and resources [Ballou et al., 1998; Even et al., 2007] – tasks typically associated with *data administration.* We suggest that, beyond the benefit offered to data administration, collecting quantitative assessment of the business-value gained as a form of metadata can improve *data consumption* as well. Business value can be measured, for example, in terms of decision outcomes (e.g., production increase, customers' purchase intent), revenues and profitability. Organizations capture such value measurements today, but rarely link them to the data resources and the decision-support tools that were used in the process of value generation.

Value-driven collection of usage metadata (Figure 2) aims at establishing such a link. To demonstrate this approach, we have successfully implemented a working prototype of a module that captures and stores value-driven usage tracking as a metadata layer. In its base, the module applies a similar approach to the one described earlier for collecting frequency-driven metadata - capturing the queries directed at a data resource, and parsing them into specific components. However, the module also collects different types of value measures (e.g., throughput, performance, and income), which are associated with a specific decision task. In certain cases, value assessments can be based on the same data resource (sale transactions, for example, can often be linked to a specific marketing campaigns that were based on a certain analysis of previous sales). In other cases – such assessments may use other information resources such as CRM and accounting systems. The module associates the value with certain decision tasks and then, through a mechanism of inference (e.g., by comparing the user-name and the time stamp), to the queries that have supported each tasks. Establishing this link between decision tasks and the underlying queries permits the creation of integrated metadata that associate business value with specific data components. An API (Application Programming Interface) can provide this form of usage metadata upon demand through function calls. Such metadata can be integrated into front-end tools, enhance the presentation, and communicate important information on the frequency of usage and on the associated value to both data consumers and administrators.

Once the link between decision tasks and queries is established, different methods can be considered for attributing value to specific data objects. For illustration, we describe here a relatively simple method, which assumes that value is attributed to the last in a sequence of queries that support a decision task. We assume that to support a certain decision, users query repetitively a certain tabular dataset with N records indexed by *[n]* and *M* attributes indexed by *[m]*. We consider *Q* queries indexed by *[q]*, each associated with a business value V^q. The binary indicator R^q_n indicates whether record *[n]* was retrieved by query *[q]* ($R^q_n = 1$), or not ($R^q_n = 0$). Similarly, R^q_m indicates whether attribute *[m]* participated in query *[q]* or not. The value of a certain query (V^q) is attributed between the participating data items, using a certain value-attribution function $V^q_{n,m}=u(V^q, R^q_n, R^q_m)$, such that $V^q = \Sigma_n \Sigma_m V^q_{n,m}$. For simplification, we use here an equal attribution of value among all participating data items. Accordingly, the overall value of a certain data items $V^{n.m}$ is given by:

(1) $V_{n,m} = \sum_{q=1..Q} V^q_{n,m} = \sum_{q=1..Q} u\left(V^q, R^q_n, R^q_m\right) = \sum_{q=1..Q} V^q \Big/ \left(\sum_{n=1..N} \sum_{m=1..M} R^q_n R^q_m\right)$, where

Q, q -	The number of queries performed, and the corresponding index, respectively
M, N -	The number of attributes (indexed *[m]*) and records (indexed *[n]*), respectively
V^q, $V^q_{n,m}$, u -	Query *[q]* value, its attribution to data item *[n,m]*, and the attribution function used
R^q_n, R^q_m -	Binary indicators of the participation *(=1)* of record *[n]* and attribute *[m]* in query *[q]*

To demonstrate the value allocation described above, we extend the previous example. We assume that each query has led to a certain promotion campaign in which a group of customers has been approached. Some customers may have responded to the campaign by making certain purchases, and the overall value attributed to a query is the total purchase amount. As illustrated, this value proxy may significantly vary among queries. We now use the allocation (Eq. 1) to assess the relative value of each data object. As illustrated by color-coding – some records, and attributes may turn out to have significantly higher value than others, and the value attribution "map" may look significantly different than the one when basing the attribution of usage frequency.

Customers

#	Customer	Gender	Income	Children	Status	…	Value
1	Abraham	Male	High	0	Single		1000
2	Sarah	Female	Low	1	Married		510
3	Isaac	Male	Medium	2	Married		50
4	Rebecca	Female	Low	0	Single		10
5	Jacob	Male	Medium	3	Married		50
6	Lea	Female	High	2	Married		1510
7	Rachel	Female	Low	4	Single		0
…							
Value		515	2000	60	500		

Value	Color Code
0	
<100	
<1000	
>=1000	

Queries

WHERE condition	Attributes Used	Records/Value	Total Value
Gender = 'Male' and Children > 0	Gender, Children	[3], [5]	100
Gender = 'Female' and Children < 3	Gender, Children	[2], [4], [6]	30
Gender = 'Female' and Status = 'Married'	Gender, Status	[2], [6]	1000
Income = 'High'	Income	[1], [6]	2000

Illustrative Example (part 2) - Assessment of Value-Driven Usage Metadata

We can potentially gain important insights by analyzing the value distribution, along with the assessment of frequency of use. For example, the *Income* attribute, which was not frequently used, is associated with the highest value, while the Children attribute, which was more frequently used, is associated with lower value. Insights as such can be transformed into valuable recommendations for a marketing associate the next time s/he plans to run a similar campaign. The decision value of each of the participants is being saved in a metadata repository. Using the API, a BI tool, which was designed to access to the value-driven usage metadata can now demonstrate value differentials and distribution, toward improving the decisions made, as demonstrated in the experimental design described next.

4 EXPERIMENT DESIGN

This section describes the design of a lab experiment, currently under final preparation stages, that tests the integration of FRM based on previous usage and the associated value.

4.1 Research Model and Working Hypotheses

The first research stage will be directed by the theoretical model shown in Figure 3. Some model variables will be measured with the test experiment tool described in the following section, while others will be assessed using a previously-tested questionnaire.

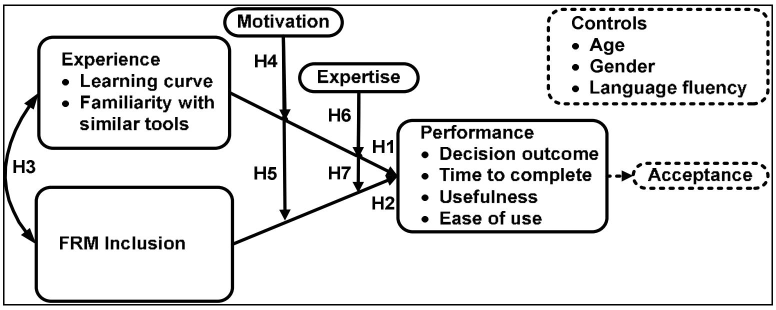

Figure 3 Research Model

Dependent Variable – Performance: The dependent variable, reflecting the ability of a user to effectively perform a task with tool support, will take one of two forms:

- Objective – actual decision outcome and time it takes to complete the decision task
- Subjective – perceived usefulness and ease of use

Previously tested models, such as TAM (technology acceptance model [Davis, 1989]), have suggested that a higher sense of usefulness and ease of use increase the likelihood of user acceptance. While this study intends to focus mainly on performance, the experiment described below will permit assessing acceptance and validating the anticipated link between performance and acceptance.

Independent Variables: The independent variables will be *Experience* and *FRM inclusion*:

a) Experience can be measured in terms of:

- Learning curve – the time a user spends using and mastering the tool
- Familiarity – the extent to which the user has previously used similar tools in the past

It is reasonable to assume that an experienced user (in terms of learning curve and/or familiarity) will perform better that a non-experienced user; hence,

- **H1: Usage Experience positively affects Performance**

b) Our key assumption is that the inclusion of an FRM will offer a major improvement in the usability of BI tools and therefore in user performance; hence,

- **H2: FRM inclusion positively affects Performance**

As discussed in the previous section, we suggest that value-driven collection of usage metadata is superior to frequency-driven collection; hence:

- **H2a: The Performance effect of FRM that are based on value-driven metadata will be superior to the effect of FRM based on frequency-driven usage metadata alone**

It is reasonable to assume a possible synergistic effect between the two independent variables, i.e., that the overall effect of Experience and FRM inclusion is higher than the effect of each alone. Hence,

- **H3: The interaction effect between Experience and FRM inclusion is positive**

Moderating Variables: It is reasonable to assume that certain user characteristics will moderate the effect of Experience and FRM inclusion on Performance. The moderating variables that will be tested are Motivation, the user's motivation to perform well, and Expertise, the extent to which the user is knowledgeable in the particular task domain. Studies (e.g., Siegel and Watts-Sussman [2003]) have shown Motivation (or involvement) and Expertise to have moderation effects on the usefulness of information resources and hence on their acceptance and adoption. Hence,

- **H4: The greater the user's Motivation, the more Experience affects Performance**
- **H5: The greater the user's Motivation, the more FRM inclusion affects Performance**
- **H6: The greater the user's Expertise, the more Experience affects Performance**
- **H7: The greater the user's Expertise, the more FRM inclusion affects Performance**

Control Variables: The experiment will control for a few additional variables - age, gender, language fluency, and possibly others.

4.2 Experiment Procedures and Tool

The model and the derived hypotheses will be tested in a laboratory setting. In the planned experiment, all participants will be asked to perform a certain decision task repetitively, aided by a BI tool. The decision outcomes, as well as the actual usage of the tool and the data resources will be tracked and measured. This will enable data collection that will allow measuring some of the variables (as described later in Table 1). In addition to tracking decision outcome and actual usage, users will be asked to complete a questionnaire, which will enable data collection on remaining variables.

Due to space limitations, we do not describe here in details all the experiment preparation procedures, but rather explain the principles that guide its design.

The decision task: the participants will act as marketing associates on behalf of a firm that offers a certain product or service to its customers (e.g., a vacation package). To decision will be aided by a large database that includes two main tables:

- ***A list of customers***, in which each associated with a given set of attributes (e.g., Income, Gender, Marital Status, and Number of Children). Based on the mix of attribute value – each customer *[t]* will be associated with a set of likelihood numbers $P_{t,z}$ $(z = 0, 1, 2,\ldots)$ of purchasing z units within a given time period, such that $\Sigma_z P_{t,z}=1$, and $Q_t = \Sigma_z Z P_{t,z}$ is the mean number of purchases.
- ***Purchases transactions,*** based on the purchase likelihoods defined per customer. A random generator will produce a large number of purchase transactions for a broad period of time.

Given access to this database, the participants will be asked to choose a customer segment that will be targeted. Approaching a customer and offering him/her a promotion has a given cost (e.g., the mail delivery fee, or the time needed for a phone call); hence, the larger is number of customers approached – the higher is the cost. Each customer is associated with certain likelihood to purchase a certain quantity of the service and, accordingly, the overall decision value is defined as:

(2) $V = \sum_{t=1..T} I_t (SQ_t - C)$, where

V - The overall decision
T - The number of customers (indexed *[t]*)
S, C- The revenue per service item sold, and the promotion cost per customer, respectively
Q_t - The expected number of item that customer *[t]* will purchase $(Q_t \geq 0)$
I_t - A decision whether to include customer *[t]* in the promotion campaign *(=1)* or not *(=0)*

A decision to include a customer may increase revenue, but at a cost. Obviously, a decision maker would prefer to include only customers for which the expected revenue is higher than the cost (i.e., only *[t]* for which $SQ_t > C$). The expected quantity Q_t is defined in advance, but not exposed to the decision maker explicitly. The user will be asked to infer which customers are likely to purchase the service by observing customers' past purchases. Moreover, the decision maker will not select specific customers, but rather will be asked to define customer segments. The segment definition will be based on a given set of customer attributes and a selection of a certain criterion per attribute (e.g., "All high-income married male customers, with 2 children or more"), where the user may choose to avoid defining selection criteria in certain attributes.

To aid the decision, each participant will be provided with a BI tool, such as the one illustrated by Figure 1. The BI tool will permit exploring past transactions, analyze purchase activity, and determine the revenue associated so far with each customer segment. The tool presents the distribution of certain measures (Revenues associated with past transactions) along certain customer attribute. The visual display imitates a decision tree. Starting at the high-level node, which reflects the entire population of customers, the user may choose an attribute (e.g., Children) along which he wishes to segment the data. For a given attribute value (e.g., Children = 2), the user may choose to segment the data along another attribute (e.g., Status), and so on. Based on the different customer segmentations that are explored by using the BI tool – the user finally selects the customer segments that will be targeted. Once the selection is made – the overall value of the selection is calculated (Eq. 2), the value is attributed among the different attributes and records (Eq. 1), and the attribution is saved in the value-driven usage metadata module. As the experiment participants keep performing the decision tasks repetitively – the value-driven usage metadata is accumulated and enhanced.

The left-hand side of Figure 1 illustrates the BI tool in its basic form, which does not include FRM. The enhance form, illustrated in the right-hand side of Figure 1, includes certain FRM enhancements – indication of the total value and the value distribution (a variance measurement) associated with the different attributes, at each node. The recommendations change dynamically, depending on the node that the user selects. The FRM enhancements are based on the usage metadata that was accumulated while participants keep performing the decision task repetitively. To help testing the hypotheses, the group of participants will be divided into a few sub-groups, and some variability will be created in the tasks that each sub-group is asked to perform (Table 1).

Hypothesis	**How the hypotheses will be tested**
H1: *Experience* affects *Performance*	• Participants will be asked to perform the same task repetitively • Some participants will provided with the same BI tool in all sessions • Participants will be asked about past experience with similar tools
H2: FRM affects *Performance*	• Initially, participants will perform the task using BI tools with no FRM • Later, some participants will be offered BI tools with FRM • Some participants will perform the task with no BI support at all
H2A: Value-Driven FRM are superior to Frequency-Driven FRM	• Some participants will be provided with FRM enhancement based on value and frequency assessments, while others will be provided with FRM enhancement based on frequency alone
H3: *Experience*-FRM interaction	• Experience/FRM – same as the above • Certain statistical regression methods permit testing interaction
H4, H5: *Motivation* moderation	• Experience/FRM – same as the above • Some participants will be offered performance-based compensation
H6, H7: *Expertise* moderation	• Experience/FRM – same as the above • Participants will be recruited from different expertise populations
Control	• The test will validate assumption that the control variables have no substantial effect in the given setting

Table 1. Hypotheses Testing

5 CONCLUSIONS

Our research investigates the integration of feedback and recommendation mechanisms (FRM) into BI tools. The working hypothesis that guides our study is that the integration of FRM into BI tools will improve their usability and increase the benefits that end-users and organizations can gain from data resources. We have described an experiment, currently under preparation, for testing the usability and the benefits of such integration in terms of improving decision-making processes. We see this experiment as a first "proof of concept" step of testing the FRM-integration idea, toward gaining insights on its usability and benefits. The controlled lab environment, in which we intend to apply the test, will permit a more precise data collection on usage patterns and value generation – what is often hard to achieve in real-world environments. Nevertheless, testing FRM integration in real-world environments would be an important follow-up step in furthering this line of research.

Another key contribution of our study, which links to the previous, is the introduction of a novel approach for usage tracking in data environments. This approach suggests that integrating quantitative assessments of usage-frequency together with the associated value gained may offer substantial benefits to data administration and consumption. Joint frequency and value assessments can help identifying unused data subsets with high value-contribution potential, may highlight flaws with repetitive use of data and, consequently, motivate new usage forms. Further, value assessment can direct design decisions, and help prioritizing data maintenance efforts. Relying on usage frequency alone might promote usage stagnation and loss of opportunity to gain new forms of benefits. Complementing frequency assessments with value assessments may help "closing the loop", in terms of providing feedback based on usage performance, and reducing the potential risks. First, value allocation gives higher weight to past usages with high contributions. Second, it can reflect variability in the importance of different subsets depending on the usage context. Lastly, it can help detecting data subsets with high contribution potential that have not been frequently used. Obviously, future extensions to our study will need to address some key limitations of this approach:
(a) Quantifying value – organization maintain performance measurements (e.g., productivity, income, and profitability) that can be possibly linked to decision tasks. However, decision performance may depend on other resources such as human knowledge and financial assets. Further, the value depends on the usage context, and value assessment for a certain type of usage tasks does not necessarily apply to others. Further, value is time-dependent, as data that can be used effectively at a certain point of time, might become obsolete later. We hasten to say that the value-allocation methodology, which we apply in this study, appears to be a better fit to operational environments in which decision tasks have a high degree of repetition, and causal relations between data usage and business performance are easier to establish. Promotion-campaign management, such as in our illustrative example would be a good representative for this type of decision-making. Financial-investment decisions would be another example for data-driven decisions, in which outcomes are measurable (e.g., the change in the value of the financial asset) and linkable to the data resources being used. Conversely, quantifying the value of decision outcomes might turn out to be more challenging in strategic decision scenarios, which are not repetitive in nature and often relay on information sources other that organizational data repositories.

(b) Linking value to specific queries - performance assessments are rarely linked explicitly to the data resources and tools used. Our preliminary prototype includes inference mechanisms for creating implicit links – e.g., based on the user name, and/or time proximity. Obviously, implicit links cannot be absolutely precise and might bias the value allocation significantly. Establishing explicit links will require stronger metadata integration between systems and, likely, redesign of data environments (e.g., joint codes that link each decision task and queries). One could question whether or not making such a high investment in redesigning data environments and BI tools would justify the benefits gained.

(c) Attributing value to specific data objects – the attribution system has critical impact on the results. Our prototype attributes value only to the last query in the sequence that generated the decision, and distributes the value equally between all the data that were retrieved. A different allocation method

may consider, for example, spreading the usage value along all queries and/or consider possible interactions among attributes – hence, unequal allocation.

Finally, we would suggest that future extensions of this study should further explore links to the research of recommender systems. Recommender systems are common in web-based user interfaces (e.g., rating systems in E-commerce sites), but less so in BI tools. Adomavicius and Tuzhilin [2005] identify the need to incorporate contextual information into the recommendation process. As value assessments depend on the context in which data is used [Even and Shankaranarayanan, 2008], we would suggest that the value-driven metadata approach may help such incorporation.

Acknowledgments

This project was supported by a grant from Microsoft R&D, Israel.

We wish to thank Prof. Nava Pliskin from the Department of Industrial Engineering and Management at Ben-Gurion University of the Negev for her advice and support.

References

Adomavicius, G., Tuzhilin A (2005), Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Transactions on Knowledge and Data Engineering, 17 (6). 734-749

Ahituv N., (1980) A Systematic Approach towards Assessing the Value of Information System," MIS Quarterly, 4:4, 61-75

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

Cooper B. L., Watson H. J., Wixom B. H., and Goodhue D. L. (2000). Data Warehousing Supports Corporate Strategy at First American Corporation, MIS Quarterly. 24 (4), 547-567

Counihan, A., Finnegan, P., and Sammon, D. (2002). Towards a Framework for Evaluating Investments in Data Warehousing, Information Systems Journal. 12. 321-338

Davenport, T.H. (2006). Competing on Analytics. Harvard Business Review. 84 (11), 99-107

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13 (3), 319-340

Even, A., Shankaranarayanan, G. and Berger, P. D. (2006). The Data-Warehouse as a Dynamic Capability: Utility-Cost Foundations and Their Implications for Economics-Driven Design. Proceedings of the Intl. Conference on Information Systems (ICIS), Dec. 2006, Milwaukee, WI

Even, A., Shankaranarayanan, G., and Berger, P. D. (2007). Economics-Driven Design for Data Management: An Application to the Design of Tabular Datasets, IEEE Transactions on Knowledge and Data Engineering. 19 (6), 818-831

Even, A., and Shankaranarayanan, G. (2008). Comparative Analysis of Data Quality and Utility Inequality Assessments. The 2008 European Conf. on Inf. Sys. (ECIS), June 2008, Galway, Ireland

Lee Y. W., Pipino, L., Strong, D. M. and Wang, R.Y. (2004). Process-Embedded Data Integrity. Journal of Database Management. 15 (1). 87-103

March, S. T. and Hevner A. R. (2005). Integrated Decision Support Systems: A Data Warehousing Perspective. Decision Support Systems 43 (3), 1031-1043

Shankaranarayanan, G., and Even, A. (2004). Managing Metadata in Data Warehouses: Pitfalls and Possibilities. Communications of the AIS 2004 (14). 247-274

Watts-Sussman, S., and Siegel, W. S. (2003). Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption. Information Systems Research 14 (1). 47-65

West L.A. Jr. (2000). Private Markets for Public Goods: Pricing Strategies of Online Database Vendors. Journal of Management Information Systems 17 (1). 59-84

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