Eliciting Information Requirements for Data Warehouses,

Vijay Raghavan
Northern Kentucky University

Follow this and additional works at: http://aisel.aisnet.org/amcis1999

Recommended Citation
http://aisel.aisnet.org/amcis1999/262

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISel). It has been accepted for inclusion in AMCIS 1999 Proceedings by an authorized administrator of AIS Electronic Library (AISel). For more information, please contact elibrary@aisnet.org.
Eliciting Information Requirements for Data Warehouses
Dr. Vijay V. Raghavan, Associate Professor of Information Systems, College of Business, Northern Kentucky University, Highland Heights, KY 41099. E-mail: raghavan@nku.edu

Introduction
Information Requirements Determination (IRD) has generally been recognized as a critical step in the software development process. The determination of information requirements is one of the most crucial stages in the software design and development process (Montezemi 1988). It is during this stage that the greatest degree of interaction occurs between the analyst and the user (Lauer et al. 1992). Information technology practitioners have generally recognized how requirements determination is different in data warehouses and very large databases. (Dacha, 1998). This study takes explores whether current knowledge of information requirements determination (IRD) can apply to the context of data warehouses and very large databases.

Characteristics of Datawarehousing
Data warehouse provides a platform of integrated, historical data from which to do analysis. Data warehouse provides the facility for integration in a world of unintegrated application systems. Data warehouse is achieved in an evolutionary, one-step-at-a-time fashion. Data warehouse organizes and stores the data needed for informational, analytical processing over a long historical time perspective. Major characteristics of a datawarehouse include subject-orientation, integration, time-variation and non-volatility. A detailed explanation of these characteristics can be found in (Inman, 98). Of these four characteristics, the time-variation component has greater implications for how the data is collected and stored. This characteristic makes the data in the warehouse as very different from data found in an operational environment. In the operational environment data is accurate as of the moment of access. In other words, in the operational environment when you access a unit of data, you expect that it will reflect accurate values as of the moment of access.

Datawarehouse data represents data over a long time horizon typically from five to ten years. The time horizon represented for the operational environment is much shorter. Applications that must perform well and must be available for transaction processing should carry a certain minimum data if they are to have any degree of flexibility at all. Therefore operational applications have a short time horizon, as a matter of sound application design. This means that every one of the key structures in a data warehouse contains - implicitly or explicitly - an element of time, such as day, week, or month. This element of time is almost always at the bottom of the concatenated key found in a data warehouse. On occasions, the element of time will exist implicitly, such as the case where an entire file is duplicated at the end of the month, or the quarter. A the third way that time variance appears is that data warehouse data, once correctly recorded, cannot be updated. Data warehouse data is essentially a long series of snapshots. Of course if the snapshot of data has been taken incorrectly, then snapshots can be changed. But assuming that snapshots are made properly, they are not altered once made. In some cases it may be unethical or even illegal for the snapshots in the data warehouse to be altered. Operational data, being accurate as of the moment of access, can be updated as the need arises.

Differences In Requirements Definition for Datawarehousing:
These differences cited here are fundamental in nature may have implications on the systems design aspects of a data warehouse. Piper (1998) summarizes certain guidelines for requirements development for general applications. The following guidelines based on Piper and other populist writings on datawarehouses are used to contrast IRD for data warehouses from conventional data stores.

Guideline 1: Lock the Business processes during development cycle to ensure that the baseline information is accurate the final product will meet all requirements of functionality. Data warehousing, by definition, will incorporate numerous queries and it may be impossible to foresee all the queries. Compared to a system to be implemented in a traditional operational environment data warehousing applications may include queries into data store that are extremely infrequent but nevertheless important. Hence the participants in the requirements development process may have to have more of a business orientation rather than simple query orientation. This also means that the process used for requirements gathering must facilitate a holistic thinking of a business environment rather than simple operational queries.

Guideline 2: Reduce the development cycle time by delivering incremental increases in functionality and quality. Data modeling for data warehousing is fundamentally different from data modeling for operational databases. The time-variant characteristic of the data warehouses has resulted in new paradigms such
as multidimensional modeling and star schema. Changing the data model is a difficult task in any environment as the applications may have to be modified to deal with a new data model. It is especially difficult in a data warehouse environment for two reasons: (1) the data model is much more complex, and (2) the performance of queries is affected to a greater degree since data warehouse deals with large set of data. The bottom line is that the traditional wisdom of providing incremental functionality is more difficult to implement in a data warehousing environment.

**Guideline 3:** Involve users during the requirements definition stage. User attitudes to systems and performance are influenced by their perceptions how involved they are in the requirements definition process (Hunton J.E. 1996). Hunton found that perceptions of control over the development process, satisfaction with the outcome, and objective measures of performance increased with different levels of participation. He encourages researchers to integrate the psychology of procedural justice into the study of IS user participation. It appears that the mere scope of a data-warehousing project associated with the increased penalties for failure in such large venture as a data warehousing project will mean user participation may have to be given greater emphasis. Number of users involved during the requirements definition stage may have to be greater in a data warehousing project as the goal of data warehousing is to support time-related queries as well. This by definition includes users from the data-to-operating environment as well as users from decision-environment.

**Guideline 4:** Be aware that team members may sometime withhold information from the stakeholders. Ginan P.J, Cooprider, J.G., and Faraj S (1998) found that team members participating in requirements development sometimes withhold information from stakeholder thus undermining the requirements elicitation process. Operational data stores may have multiple users of the same data thus providing greater opportunity to cross verify the data model. Data warehouses provide lesser opportunities to validate the data model developed. Dyche (1998) offers some specific guidelines for requirements development for data warehousing.

**Guideline 5:** Make sure management and end users understand the complexity and time required for data sourcing. Sourcing data from operational systems is the bugaboo of large, cross-functional corporate warehouses. Management should understand that planning for data sourcing is never an exact science and can affect the project timeline, while end users should be prepared for potential delays or modifications to expected data.

**Guideline 6:** Make sure that the business discovery results are predicated on availability of certain data. If possible, plan on a data verification activity. Since data warehouse data is more cross functional, there is a greater the likelihood for misinterpretation of data requirements. One way avoid misrepresentation of data is to include data samples when discussing sources as well as in building the data model.

**Guideline 7:** Make sure that specific business areas have budget to support enhancements and fund applications. This not only sets realistic expectations but also makes people think twice about “need to know” analysis.

**Guideline 8:** Recruit an executive sponsor from the business side. IT is the only common thread between the disparate organizations slated to use a corporate warehouse. IT managers and staff should be key participants in the business discovery process, but because IT cannot prioritize or differentiate between the business problems of, say, marketing organizations, the executive sponsor should come from the business side. This guideline highlights the need to have the business orientation guiding the requirements definition. While it is sufficient to have a user specify the use of data in an operational setting, only a sponsor from the business side can fully understand the implications of providing for data that are not normally used but will be needed in future.

**Guideline 9:** Differentiate between wish lists and tangible benefits. Just because someone has a new way of exploiting all that data does not mean that it is legitimate or will help everyone. This guideline simply highlights the need to scope the data warehouse project even at the information requirements stage.

**Conclusion**

The traditional guidelines for requirements definition will have to approached with a different perspective in eliciting requirements for a data warehouse. Marakas and Elam (1998) have recommended that disciplined questioning strategies for interviewers can be provided through structured training. Since data warehousing is fundamentally different from operational data stores conventional guidelines will have to revisited, and if necessary, redefined to suit the characteristics of a data warehouse. The goal of this research is to revisit the traditional guidelines and propose a conceptual model categorizing issues of requirements elicitation of data warehouses that would help in understanding how the process differs from requirements definition for operational data stores.

**References available upon request from author**