

Organization of data warehousing in large service companies - A matrix approach based on data ownership and competence centers

Robert Winter and Markus Meyer

Institute of Information Management, University of St. Gallen
Robert.Winter@unisg.ch

Abstract

Most work on data warehousing is dominated by architectural and data modeling issues. Although often key to the success of data warehousing projects, organizational issues are rarely covered. Based on project experiences in several large service companies, organizational requirements for data warehousing are derived. As a foundation for the development of organizational structures and organizational rules for data warehousing, the data ownership concept is specified. Based on that concept, a two-dimensional organizational structure is presented that allows to combine infrastructural competencies and content competencies. The proposed concept was implemented in a large Swiss bank.

Introduction

Data warehousing overcomes various problems that result from the need to connect large numbers of decision support systems to large numbers of operational systems by providing a hub for subject-based, historical, consistent, and non-volatile information. By connecting decision support systems and operational systems to a (logically) centralized hub, the number of interfaces can be reduced dramatically and information quality can be guaranteed more effectively, to name just two important benefits.

Due to the nature of data warehousing projects as large systems development projects, the discussion of organizational issues is usually restricted to the systems development process (e.g. Kimball, Reeves, Ross and Thornthwaite 1998) or even to data organization. Organizational structures and organizational rules that are needed to guarantee sustainable operations of data warehousing in large companies are hard to find. This gap is particularly surprising because several studies point out that organization-related issues are among the most critical success factors for data warehouse projects (Watson and Haley 1998, p. 34; Finnegan and Sammon 1999, p. 191): “Most [enterprise data warehousing] projects fail for political and organisational reasons, rather than for technical ones” (OVUM 1998, p. 1).

The development of organizational structures and organizational rules is covered by many integrated approaches to systems development (e.g. Scheer 1992; Österle 1995). Since all these approaches, however, are process-centered, they cannot be directly applied to data warehousing which is data-centered. As a foundation for developing the organization of data warehousing, the concept of data ownership has to be derived from traditional, process-oriented ownership concepts.

The derivation of the data ownership concept in section 3 is based on a short discussion of organizational challenges of data warehousing in large service companies (section 2). Based on the proposed data ownership concept, a two-dimensional organizational structure is presented that allows to combine infrastructural competencies and content competencies in section 4. The implementation of the proposed concept in a large Swiss bank is summarized in section 5. Section 6 comprises conclusions and directions for further research.

Most of the work reported in this paper has been initially developed by Meyer (2000) and has been integrated into a broader business engineering context by Meyer and Winter (2000). The concept has been validated by implementing the proposed organizational structures and rules in selected business units of a large Swiss bank. Experience from additional implementations in other large service companies and a generalization of the concept within a general information logistics infrastructure (Winter 2001) are ongoing research projects.

Organizational challenges of data warehousing

An analysis of data warehousing projects in large Swiss and German service companies (Meyer 2000, for project details refer to the competence center intranet included in the references) shows that the following issues can be regarded as the most important organizational challenges of data warehousing:

- **Alignment with regard to company goals:** Different units involved in a data warehouse project usually have different - maybe even conflicting - goals. Decision makers want their specific information requirements to be covered flexibly and in realtime. Operations units want to manage their daily business efficiently and securely. Information management wants to build a common platform that effectively decouples as many decision support systems and operational systems as possible. In an early phase of a data warehouse projects, the involved units have to agree on common project goals and have to solve obvious or latent conflicts between goals.
- **Clear responsibilities for data:** In contrast to business processes which have usually been assigned to a process owner, it is often not clear whether certain data are owned by information processing units, data management units, or non-IS business units. Have high-quality data to be delivered by some service unit, or are decision makers responsible for collecting and cleansing of data? How should responsibilities be assigned for data that go through a multi-step derivation sequence from operational systems through various staging areas, the enterprise-wide data warehouse, various aggregation/selections stages, data marts, and finally data pools of decision support systems? The concept of data ownership is intended to create a sound foundation for the definition of roles, responsibilities, and data warehousing processes.
- **Data quality management:** Data quality relates not only to the correctness of data, but also to the way data are provided and used (English 1999, p. 22). Since operational systems are intended to support business transactions and not to support business decisions, data quality can be sufficient (for information processing units) and insufficient (for non-IS business units) at the same time. For a specific decision, often not all necessary data that decision makers need (or believe to need) can be provided. But quality requirements for decision support often conflict operational systems management tools (e.g. time-consuming integrity checks). For an effective data quality management, business specialists (who only are able to assess data quality from a usage perspective) have to be made responsible for sourcing and transformation even if operational systems are affected.

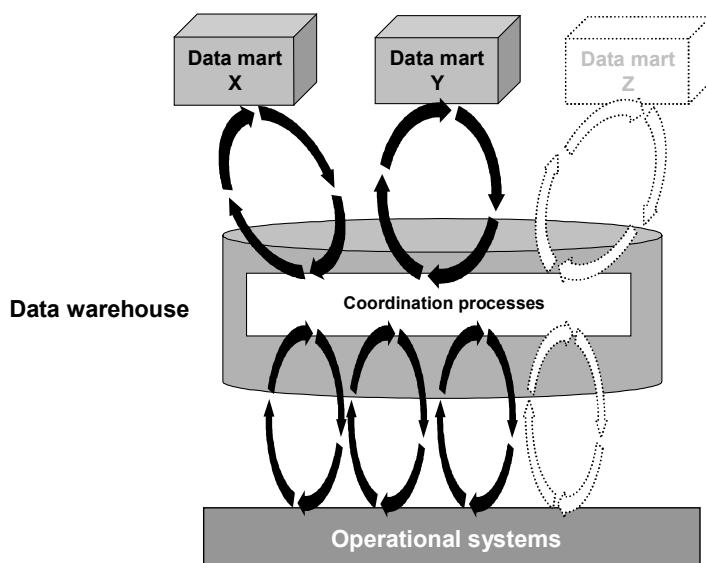


Figure 1. Multi-level coordination in a multi-layer data warehousing infrastructure

- **Integration management:** Better coordination between business processes and management processes allows data warehousing to be implemented more effectively because the right, critical information can be provided more accurately. Integration management, therefore, has to identify performance indicators that are both useful for specifying effective management processes and can be provided accurately and timely by operational systems.
- **Multi-level structure:** A physically centralized, enterprise-wide data integration layer can only be implemented in small companies. For a large company, several layers of data integration between operational systems and decision support systems have to be implemented. The most common solution is to differentiate at least an enterprise data warehouse layer and a data mart layer. If n layers exist, coordination processes between data-providing systems (and responsibilities) and data-consuming systems (and responsibilities) have to be implemented on $n-1$ levels (see figure 1).
- **Sustainability:** Traditional development projects can be characterized by restricted resources, restricted running time, and uniqueness. In contrast, data warehousing is a permanent process (Gardner 1998, p. 54): After an initial development phase, not only stable operations and reliable data supply must be provided, but also continuous improvements and adjustments are needed to reflect changing decision support needs and modifications of operational systems. For a sustainable implementation of data warehousing, dedicated permanent roles and structures must be created.

Data ownership in large organizations

Data warehousing is intended to support decision makers. Hence, business units - and not support units like data management or information processing - must specify information needs and must sponsor data warehousing projects (Kimball, Reeves, Ross and Thornthwaite 1998, pp. 43ff.; Finnegan and Sammon 1999, p. 186; Watson and Haley 1998, pp. 35ff.). As a consequence, only business units – and not support units like data management or information processing – can be data owners. Data ownership has three aspects:

- **Contents:** The data owner is responsible for the correct modeling, documentation, and quality control of data including the maintenance of appropriate meta data.
- **Methods:** The data owner is responsible for specifying correct extraction, transformation, and aggregation methods to derive relevant information from operational data.
- **Development and supply:** The data owner is responsible for the development and operations of an infrastructure that provides information for decision support purposes.

This interpretation of data ownership is more comprehensive than Devlin's (1997, pp. 279ff.) or Kimball et al.'s (1998, p.70) definitions to reflect a holistic, active responsibility that business units should accept to gain effective control over an increasingly important corporate resource.

Identification and responsibilities of data owners

In mature organizations, ownership of operational applications should be consistent with business process ownership. As a consequence, only business units – and not support units like data management or information processing – can 'own' data because they 'own' the operational applications that manage respective data. Only those business units should supply operational data to the data warehouse that have sufficient knowledge about the relevant systems and data. Responsibility for business processes should not only be complemented by responsibility for respective operational systems, but also by responsibility for respective data warehousing supply processes. Hence data owners on the data warehouse layer are identified by analyzing business process / operational application ownership.

In contrast to the data warehouse layer, the identification of data owners on the data mart layer is derived not from business process ownership, but from information needs of the respective decision makers. In many cases, information needs go across several business processes. As an example, decision makers from marketing units or from risk management units need data from all transactions of certain customers across all product lines of a bank. Data owners on the data mart layer receive data from various data owners on the data warehouse layer. Being end users in the information supply chain, data owners on the data mart layer are the true sponsors of data warehousing activities and use this role to guarantee that the entire data warehousing process is aligned with their information requirements.

Figure 2 illustrates various types of data ownership: Business units A and B own operational data, data warehouse data, and data mart data for their respective business processes. Business unit C owns operational data that it loads into the data warehouse, but runs no

decision support systems and, as a consequence, owns no data mart data. Business unit D owns no operational and no data warehouse data, but runs decision support systems so that it owns data mart data. While business unit C is only a 'data supplier' and business unit D is only a 'data user', business units A and B have both roles.

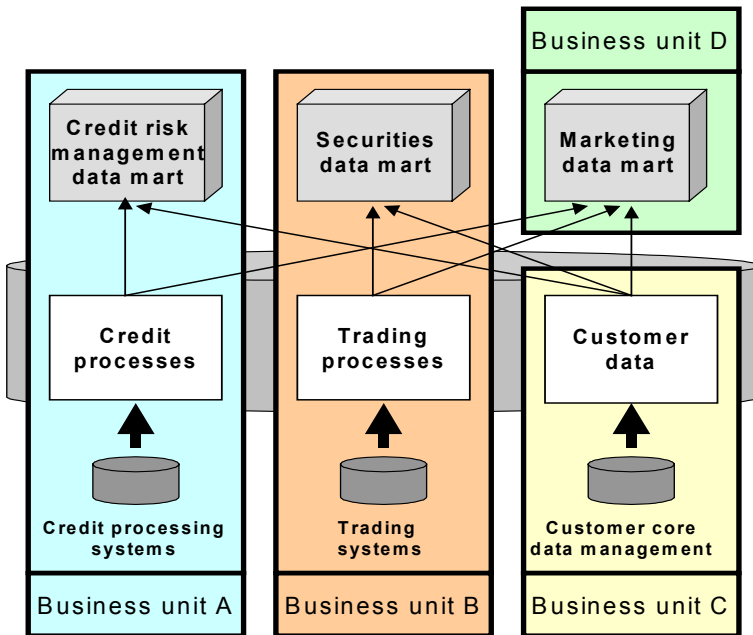


Fig. 2: Different types of data ownership

Data owners on the data warehouse layer are responsible for interfaces to operational systems. As a consequence, ownership candidates must

- be able to interpret information needs of data mart projects correctly,
- be able to derive data requirements for operational data sources correctly from information needs, and
- have sufficient knowledge of operational systems to identify and utilize appropriate data sources.

Two-dimensional organizational structure

The foundation for the design of organizational structures is the analysis of relevant activities. Basically, data warehousing comprises activities related to business processes, coordination activities, and technical activities. Due to different skill sets, these activity clusters are usually aggregated into different organizational units:

- **Activities related to business processes:** Sourcing activities are needed to extracting certain information from various sources, thereby implementing data ownership on the data warehouse or the data mart layer. Sourcing activities comprise the entire process of analyzing, designing, implementing, and running data supply processes. For such tasks, teams comprising business specialists and systems specialists are necessary to replace the traditional, project-oriented separation of (business) sponsor and (IT) contractor.
- **Coordination activities:** In particular on the data warehouse layer, coordination activities are extremely important: By appropriate measures, reuse synergies should be exploited and data from different sources should be presented in an integrated manner. Architectural activities enforce modeling standards (e.g. for historical data, for references, for dimension hierarchies, for meta data specification). Data mart channeling activities assign priorities to concurrent data mart development projects and help to avoid redundant development activities for data marts that use identical data warehouse data. On the data mart layer, coordination activities are only needed to integrate decision makers' information requirements.
- **Technical activities:** In addition to business-related activities, data warehousing includes a large number of technical activities: Systems must be specified, designed, implemented, maintained, and run. Sourcing activities must be supported by providing

technical information regarding formats, refresh cycles, backups, archiving, etc. Coordination activities must be supported by documenting appropriate meta data.

Organizational units

Based on the above activity analysis, the following organizational units were created:

- **Sourcing teams:** Sourcing teams analyze information needs, identify appropriate data sources, maintain data models and meta data, develop, implement, and run respective sourcing processes. Since significant knowledge of specific business or decision support processes is required, different business units create different sourcing teams.
- **Data mart channeling:** Channeling units consolidate information needs in different data mart projects and coordinate the data warehouse development process with data mart development projects.
- **Data architecture:** Data architects maintain data models and modeling standards on the data warehouse layer or the data mart layer.
- **End user coordination:** Not unlike a communications unit, professionals in this unit support end-users by distributing knowledge about what information is available in the data warehouse and data marts. In addition, information requirements analyses are supported, project proposals are supported by integrating information needs, and tool utilization is supported.
- **Infrastructural units:** These units run the respective systems on the data warehouse or data mart layer. Software tools for developers and decision makers are evaluated and run. Other organizational units are supported in software tool utilization.

Sourcing teams are structured according to business activities, i.e. reflect the company's business process structure. All other organizational units are structured according to technical and/or coordination activities, i.e. create a competence center structure that is usually orthogonal to business unit structure.

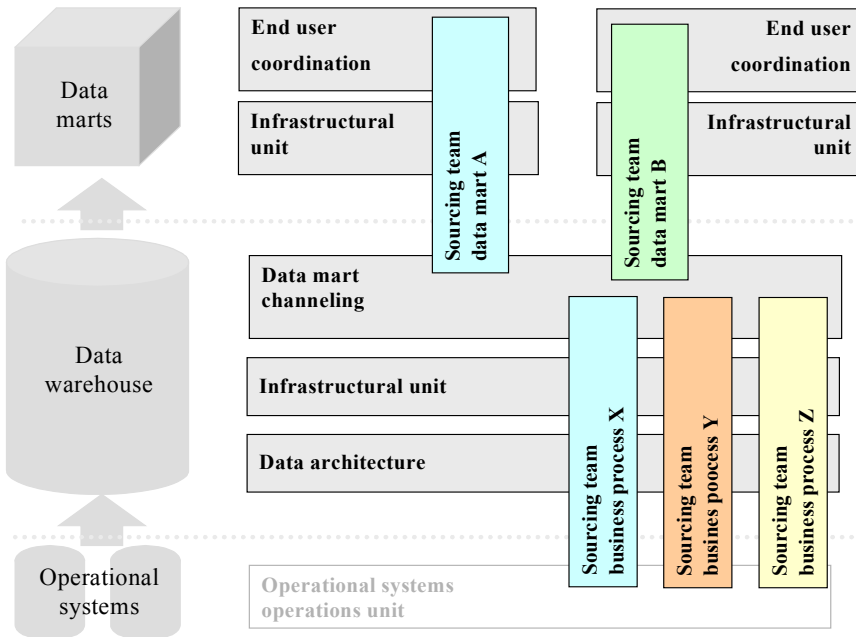


Fig. 3. Matrix structure for data warehousing organization

Matrix structure

If sourcing teams that perform process-related activities are assigned to a vertical dimension and all other organizational units that perform technical or coordinative activities are assigned to a horizontal dimension, a matrix structure results that is found in many organizations where both functional specialization and management of shared resources is important. The matrix structure is illustrated

by figure 3. In order to guarantee effectiveness, however, it has to be specified whether vertical, business-related units or horizontal, infrastructure-related units have priority.

The importance of data ownership and the strong link between business process ownership and data ownership leads to the decision that vertical, business-oriented units have priority over horizontal, infrastructure-related units.

- The **primary organizational dimension** are sourcing teams that holistically perform data supply activities for their respective business units. These permanent, cross-disciplinary teams are responsible for the entire information supply process and implement data ownership.
- The **secondary organizational dimension** is comprising channeling teams, coordination units, and infrastructure units that coordinate sourcing activities and create synergies by providing a common technical basis for data warehousing. These tasks are organized by means of permanent competence centers.

Implementation at a large Swiss bank

The proposed organizational structure has been implemented in selected business units of a large Swiss bank. Sourcing teams for payment services and custody services, customer information management, product data management, and other core data management areas have been set up as primary organizational units to perform vertical integration tasks. As secondary organizational units which perform horizontal integration tasks, a data mart channeling competence center, a data architecture competence center, and a technical architecture unit have been created. While business representatives of sourcing units on the data warehouse layer have been assigned to centralized units that are associated with the respective business unit to consistently implement data ownership, business representatives of sourcing teams on the data mart layer as well as members of end user coordination teams have been directly assigned to respective business units, i.e. no centralized units were created. Technical representatives of sourcing teams on the data warehouse layer as well as representatives of horizontal units on that layer were assigned to the centralized IS unit of the bank. In contrast, technical representatives of sourcing teams on the data mart layer as well as members of infrastructure units on that layer were assigned to decentralized IS units associated with respective business units. Within the centralized IS unit of the bank, data warehouse development was positioned parallel to the development / maintenance of operational systems. The implemented organizational structure is illustrated by figure 4.

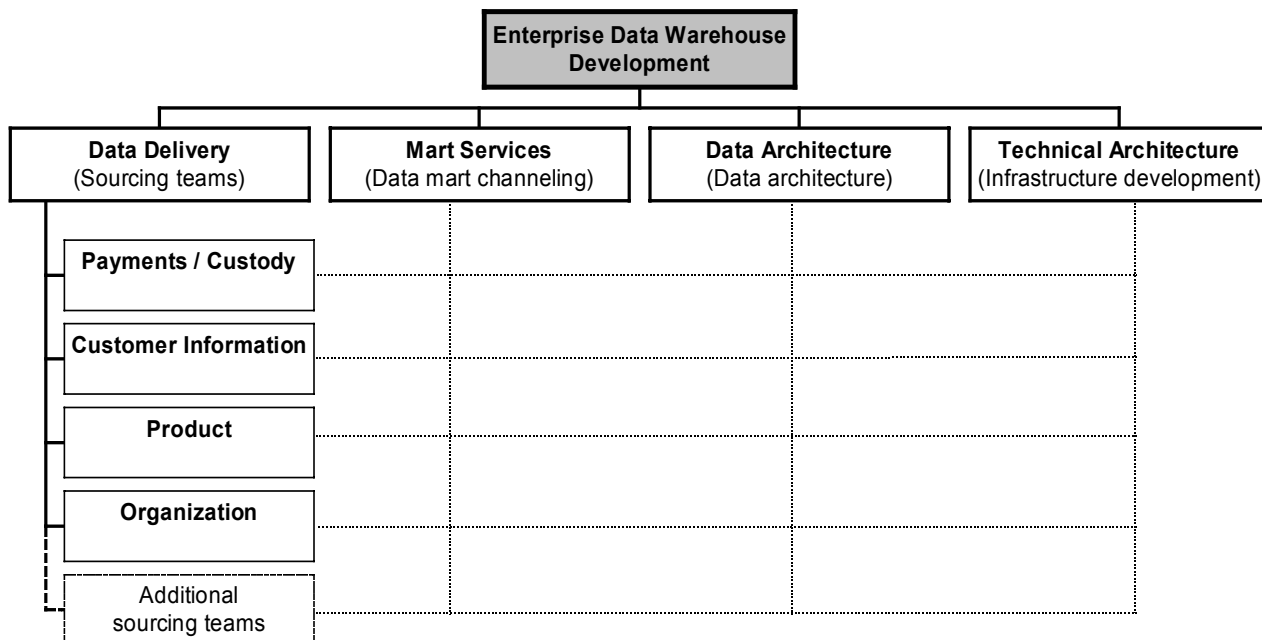


Fig. 4. Organizational structure of data warehousing at a large Swiss bank

The two-dimensional organizational structure based on data ownership (= business responsibility) and supported by competence centers (= technical responsibilities) proved to be successful so that a roll-out to other business units of that bank and the creation of similar organizational structures by other partner companies of the competence center Data Warehousing are under consideration (details can be obtained at the competence center's web site <http://datawarehouse.iwi.unisg.ch>).

Conclusions and further research

In all large companies surveyed by the data warehousing competence center, data warehousing is regarded as a permanent business activity. Based on the proposed, extended data ownership concept, business units become responsible not only for implementing their business processes by appropriate operational systems, but also for providing appropriate data for decision support purposes.

To allow for an effective data quality management, process-oriented sourcing teams which are associated with data owning business units should be created. In those teams, business specialists and technical specialists collaborate closely and can be held responsible to implement sourcing processes from data sources in operational systems all the way to information utilization in decision support applications (given that the respective business units is both data supplier and data user, see section 3).

Since vertical integration alone would lead to redundant data structures and infrastructures, competence centers should be created to implement horizontal coordination. Enforcement of standards, utilization of methods, reuse of infrastructures etc., however, are regarded as secondary organizational goals compared with a tight integration of sourcing activities with business responsibilities.

Like every other organizational design, the proposed concept is closely related to a specific company with more or less unique properties and strategies. We observe, however, similar considerations in many other large, decentralized service companies. If optimization of resource utilization dominates the ability for delivering fast, flexible, and high-quality information support in a specific setting, 'horizontal' units will be regarded as more important as 'vertical' sourcing teams. We expect, however, the concept of extended data ownership, the taxonomy of organizational units, and the derivation of the matrix structure to be applicable in such settings, too. Continuing research is needed to validate the concept by analyzing additional companies and industries. In addition, it has to be analyzed whether operational usage of information integration infrastructures (e.g. operational data stores) leads to similar organizational concepts.

References

- Competence Center Data Warehousing, <http://datawarehouse.iwi.unisg.ch>
- Devlin, B.: Data warehouse: from architecture to implementation; Reading 1997.
- English, L. P.: Improving data warehouse and business information quality methods for reducing costs and increasing profits; New York 1999.
- Finnegan, P.; Sammon, D.: Foundations on an organisational prerequisites model for data warehousing; in: Pries-Heje, J.; Ciborra, C.; Kautz, K.; Valor, J.; Christiaanse, E.; Avison, D.; Heje, C. (Eds.): Proceedings of the 7th European Conference on Information Systems; Copenhagen 1999, pp. 182-193.
- Gardner, S.R.: Building the data warehouse; Communications of the ACM, 41 (1998) 9, pp. 52-60.
- Kimball, R.; Reeves, L.; Ross, M.; Thornthwaite, W.: The data warehouse lifecycle toolkit - Expert methods for designing, developing and deploying data warehouses; New York 1998.
- Meyer, M.: Organizational design of enterprise-wide data warehousing – Roles, responsibilities and processes in a Swiss universal bank; Dissertation no. 2424, University of St. Gallen 2000 (in German).
- Meyer, M.; Winter, R.: Organization of enterprise-wide data warehousing; in: Jung, R.; Winter, R. (Eds.): Data Warehousing 2000 – Methods, applications, strategies; Physica: Heidelberg 2000, pp. 309-331 (in German)
- Österle, H.: Business in the Information Age - Heading for New Processes; Springer: New York 1995.
- OVUM Evaluates: Data warehouse tools and strategies; London 1998.
- Scheer, A.-W.: Architecture of integrated information systems; Springer: New York 1992.
- Watson, H. J., Haley, B. J.: Managerial considerations; Communications of the ACM, 41 (1998) 9, pp. 32-37.
- Winter, R.: The current and future role of data warehousing in corporate application architecture, in: Proc. HICSS 34; IEEE 2001.