The Customer-Centric Data Warehouse – an Architectural Approach to Meet the Challenges of Customer Orientation

Hans-Georg Kemper[1], Phil-Lip Lee[2]
[1]*University of Stuttgart, Dept. VII Business Administration and Information Systems*
Breitscheidstr. 2c, 70174 Stuttgart, Germany
Phone: (49) 711 / 121-3037, Fax: (49) 711 / 121-3197
[2]*University of Cologne, Chair of Information Systems and Information Management*
Albertus-Magnus-Platz, 50923 Cologne, Germany
Phone: (49) 221 / 470-5325, Fax: (49) 221 / 470-5383
Kemper@wi.uni-stuttgart.de, Lee@wi-im.uni-koeln.de

Abstract

This paper presents research-in-progress. An extensive customer-centric data warehouse architecture should enable both complex analytical queries as well as standard reporting queries on customer data without performance restrictions for both requirements. This paper introduces a dichotomic approach, which brings together these contradicting tasks of a data warehouse. On the one hand, it elaborates on the qualities of customer data and their implications on data structures due to their change over time. The authors will present a data concept, that is specialized in gaining a realistic image of the customer, ideally over her entire customer lifecycle. On the other hand, the paper works out the role of the operational data store (ODS) in the light of CRM: it presents how the ODS supports its counterpart, the data warehouse, in the dichotomic approach for the maintenance of high performance and low response times.

1 Introduction

This paper presents research-in-progress. The emergence of the CRM[1] concept in the recent years shows that enterprises are eager to implement the philosophy of customer orientation in their business processes. The main reason for this is that market power has shifted to the customer. Competition has gained in rigor through market saturation, new competitors and technologies, price pressure and cost containment.[1; 14] With the implementation of CRM systems, enterprises collect and combine data about their customers' preferences. And by utilizing this CRM-generated knowledge, they attempt to improve their service for the customer through individualization, e.g. by choosing the preferred channel and timing of communication. The goal is to reduce customer churn, i.e. to keep up (or even increase) customer loyalty [13] and, finally, to maximize the customers' value for the enterprise over their lifetime (e.g. Customer Lifetime Value[2]).

A fundamental condition to achieve this goal is customer information. For a successful maximization of CLV, enterprises require an integrated, holistic view on their customers, especially in mass markets. But the task to build a homogenous customer database is enormous. The long-known problem of customer information is that it is scattered all over the heterogeneous system landscape of the enterprise, whether in e.g. the ERP[3] system, the marketing database, call center applications or externally acquired customer profile data.

A first step to solve this problem has been done with the introduction of data warehousing, which implied the filtering, harmonizing, aggregation and enrichment of operational data for analytical purposes. [7] Yet, the handling of customer data, especially querying of historized data, still poses enormous problems for data warehouse architects. [18, pp.83] Customer data have special qualities, which are difficult to manage. Thus, the goal of this paper is to present an extension of existing data warehouse approaches to meet the requirements for customer orientation.

The following sections give an overview of the characteristics of customer data, the challenges they put

[1] CRM: Customer Relationship Management

[2] Definition of Customer Lifetime Value (CLV) by [11]: "…the amount, by which a person's, household's or business' revenues over time exceed, by an acceptable amount, the company's cost of attracting, selling and servicing that customer".

[3] ERP: Enterprise Resource Planning

on data warehousing and an architectural approach to meet these challenges.

2 Motivation

With the appearance of CRM, the data warehouse seems to obtain a real business justification in a number of enterprises. We support this opinion in the controversial discussion about the importance of CRM for data warehousing, but vice versa one may also say that CRM necessitates data warehousing to enable essential analytics. [4]

The introduction of customer orientation can have serious implications on a data warehouse, as it may increase enormously in complexity and volume. Reasons for this are:

- Customer data are often spread redundantly in separate applications and databases (see Section 5.1).
- Masses of transaction-oriented low-level detail data become highly relevant in the course of customer orientation (see Section 5.2).

If these implications are not considered carefully, then the success of a CRM initiative may be put at risk. Therefore, the development of a customer-centric approach to data warehousing seems to be appropriate.

The research topic customer-centric data warehouse is in the intersection of the areas CRM and data warehousing. These areas are embedded in the main theme of Business Intelligence (BI). The term BI represents the vertical integration of the different kinds of management supporting IT systems, according to [8].

3 Objective, Design and Status Quo of Research

The research project follows a constructive research design in the form of an intensive case study, which will be backed by an empirical evaluation. The goal is the conception and implementation of a customer-centric data warehouse architecture. The research approach is as follows:

Phase 1:

- *Analysis* in the research fields “CRM” and “implementation of customer orientation in businesses” by secondary research.
- Gathering and Analysis of CRM-related requirements of data warehousing in the telecommunications sector through explorative interviews and case studies.

Phase 2:

- Evaluation of risks and opportunities of CRM support by data warehousing.
- Derivation of recommendations for a customer-centric data warehouse framework.
- Development of hypotheses on customer-centric data warehousing.
- Development of a customer-centric data warehouse *architecture* in collaboration with an implementation partner from the telecommunications industry.

Phase 3:

- Development of a customer-centric data warehouse *architecture* in collaboration with an implementation partner from the telecommunications industry.
- Development of a database *prototype* of a customer-centric data warehouse, based on the elaborated architecture, if possible in collaboration with the implementation partner.

Phase 4: Evaluation of prototype and hypotheses.

At the moment of paper publication, the project will be in Phase 2. While continuing with Phase 1 “Analysis” of the research approach, the next step of this project consists of the search for an implementation partner from the telecommunications industry (refer to Section 3). Thus, the next concrete tasks are to investigate in the state of customer orientation of telecommunication providers and in the state of implementation of customer-centric issues in their data warehouses. Therefore a quantitative empirical analysis is planned to be performed.

Furthermore, the main task is to engage a dedicated telecommunications provider for the joint elaboration on the customer-centric data warehouse architecture and the building of a prototype. For this research project, the telecommunications industry has the following advantages:

1. Telecommunication providers are highly customer-oriented due to the strong competition in the telecommunications market. They are highly experienced with the CRM subject, as they have been with the first companies to implement systems that were called CRM.
2. Telecommunication providers have a large customer base, with contract and billing relationships to their customers. Thus, they can exactly assign data to customers where applicable. This may e.g. not apply for buyers of cellular phone prepaid cards.
3. The usage of telecommunication services (mainly calls) is personalized and can be monitored very well.
4. The usage frequency of telecommunication services is very high, so detailed customer profiles can be derived from the call detail records.
5. Telecommunication providers maintain data warehouses with large volumes of customer and usage data.

[4] Refer to [12] and [18, p.10].

4 Theoretical Foundation

This paper is based on the concepts of the following authors:
- The concept of the *Corporate Information Factory* by Inmon [5] points out the alignment of different data warehouse elements with different requirements, e.g. the core data warehouse for the provision of exhaustive data and the operational data store (ODS) for the real-time provision of data (see definitions on the terms *data warehouse* and *ODS* in Section 4.1).
- The concept of the *Customer-Centric Enterprise* by Imhoff et al [2] describes the embedding of a specialized ODS for CRM requirements – the Customer ODS – into the Corporate Information Factory architecture of Inmon.
- In his *General Conceptual Model for a Customer-Centric Data Warehouse,* Todman [18, pp.117ff.] analyzes, how the time dimension profoundly influences the customer dimension in the data warehouse. He has built a framework that is able to absorb these influences in a history concept and to provide a more realistic image of the customer.

In the following, special aspects of the data warehouse, the ODS and the customer dimension will be described more profoundly.

4.1 Definition and Comparison of Data Warehouse and ODS

As for the term "data warehouse", the generally accepted definition coined by Inmon [3] as "a subject-oriented, integrated, time-invariant and non-volatile collection of data" will also be used in this paper. The data warehouse structure shall be optimized for the "support of management's decision-making process" through data analysis [3], e.g. the data should be normalized to support data mining. The data warehouse is also supposed to provide detailed data as well as aggregated and derived data.

Complementary to the data warehouse, the Operational Data Store (ODS) is a hybrid data architecture to cover the requirements for both analytical and operational tasks. Inmon [4] characterizes the ODS as a "subject-oriented, integrated, volatile, current-valued data store, containing only corporate detailed data". In his Corporate Information Factory (CIF) framework, he has positioned the ODS between the operational systems and the data warehouse, which may lead to the impression, that the ODS merely provides ETL and buffer functions.

In professional circles, the general idea of the ODS has been accepted, but its position and functionalities in the original CIF framework have been critically discussed and extended in subsequent frameworks. E.g., already in 1997, Kimball [9] suggested, that the functionality of the ODS should be expanded from providing only current-valued data to providing a data history in form of a current rolling month.

The ODS can be increased in functionality and add to the performance of the data warehouse in the following points:

Point 1: The ODS provides the data interfaces to the operational systems for primary data extraction. Consequently, these interfaces need not be duplicated for the data warehouse.

Point 2: The sub-tasks of an ETL layer can be broken up and divided between the ODS and the data warehouse for a balancing of the ETL load:
- *Extraction* of data from operational systems and incremental update of ODS with light or no transformations on data. It is presumed that the ODS is updated several times a day.
- Daily *loading* of bulk data from the ODS into the data warehouse staging area, utilizing efficient database-proprietary tools of the data warehouse.
- *Transformation* of imported ODS data in the data warehouse staging area with filtering, harmonization, aggregation and enrichment of data. [7]
- Final storage of transformed data into DWH.

Point 3: A ODS may also receive summarizations and calculations from the data warehouse (refer to the CIF architecture by Inmon [5]). Thus, the DWH could be unburdened from masses of standard queries and consequently provide more resources for long, complicated analyses. At this point, we refer to a discussion in Section 7.

Point 4: In an idealized infrastructure of operational and dispositive data systems, ODS could also have the function to maintain the consistency between operational systems, as changes of data in one operational system (e.g. customer master database) can be reconstructed in other systems. But such an environment would require, that operational systems be enabled to import data from third party systems, which they usually are not.

4.2 Conceptual Customer Model by Todman

In the following, the term "customer data" will be differentiated into three categories, according to Todman [18, pp.75]:

1. *Customer Behavior Data:* comprise the transaction history and/or contact history of the customer. These histories mirror the customer's behavior towards the

enterprise. In a multi-dimensional data context, usually the sales and contact fact tables represent these histories.

2. *Changing Customer Circumstances Data:* usually are attributes in the customer master data, e.g. address, age, marital status, telephone number or bank account. This data category refers to the concept of Slowly Changing Dimensions by Kimball [10]. Alterations in these data are not as frequent as in the customer behavior data, but they have a profound impact on the semantics of queries. E.g. a change of the customer address may imply a move into another sales region, but the former sales volumes of this customer must not be transferred to the new sales region. Otherwise this would distort the sales totals per sales region. Therefore, a historization of customer-related dimensional data and their changes is regarded as necessary.

3. *Derived Customer Segment Data:* are data which have been calculated from attributes of the above data categories and which serve to segment the customer base, e.g. the Customer Lifetime Value, the Propensity to Churn or the Cancellation Probability. If an enterprise targets to monitor the development of its customers over their lifecycle, it has to track these derived data over time. But this kind of data need not necessarily be derived only from data within the enterprise, as third-party vendors offer data of individuals or geographical areas with demographical and marketing-relevant information e.g. purchasing power. These external data can be integrated into the customer database and be utilized as guiding or planning values.

In transaction systems, Changing Customer Circumstances Data and Derived Customer Segment Data usually are modeled as dimension attributes of the entity type CUSTOMER. This is being done for performance reasons, and it is a practical approach. But for the purpose of customer analysis *over time periods,* an enterprise also needs to track how the values of these data categories develop over time.

This leads to a conceptual customer model (as depicted in Figure 1. [18, p.77]), which explodes the former single entity type CUSTOMER into four sub-entity types. The three data categories are in an n:1 – relationship with the original CUSTOMER entity type. With "validity periods" as unique identifiers, it is possible to reconstruct the state of a customer attribute for any time period.

This approach considers the above-mentioned requirements for customer analysis over time periods. A disadvantage is that queries on this data structure become very complex, due to the necessity to historize dimensional data.

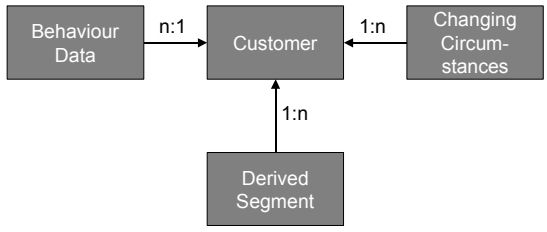

Figure 1: General conceptual model for a customer-centric data warehouse by Todman

5 Challenges of Customer Data

Besides the problems the time dimension induces on the customer dimension, there are further

qualities of customer data that constitute serious challenges to CRM and data warehousing:

- Customer data in different operational data sources cause data inconsistencies.
- Exponential growth of customer data decreases the efficiency of data warehousing.

5.1 Implications of Multi-Channel / Multiple Customer Touch point Environments

When implementing a customer-oriented perspective, enterprises aim to align their activities with customer parameters that are derived from the collected customer data. These data are generated by sales over multiple sales channels (e.g. direct and indirect sales) and by contacts over multiple customer touch points (e.g. phone, fax, mail, e-mail). Thus, the data are scattered in different databases and applications within the enterprise, such as call center, databases of marketing, sales and service, accounting and ERP systems. An implication is that, in most cases, this partial information is not integrated and therefore highly inconsistent. If these data are not carefully processed before loading into the data warehouse, then the quality of the data warehouse is put at risk.

A serious consequence is that the enterprise lacks a "single view" on their customer, i.e. an integrated and consistent set of customer data. For example, different company departments may not know they are serving the same customers, or doubles in a customer database contain contradicting attribute values.

These phenomena are commonplace, and often they are the reason when customers are being addressed and

served inadequately at customer touch points. The results become apparent in the form of higher costs (e.g. multiple mailings to customers) or in lost sales (e.g. churning customers who are annoyed of multiple mailings).

5.2 Implications of Exponential Growth of Customer Data

To be able to analyze and understand their behavior, customers ideally have to be tracked over their complete lifecycles, from acquisition to retirement. But this would require enterprises to collect and store all customer-related transaction data on the highest possible detail level, with the effect that the amount of analyzable data is predestined to grow exponentially. Taking an example from the telecommunications industry, about 200 million calls are made in the US daily, with a call detail record (CDR) size of 50 bytes per call. A telecommunication provider with a market share of ca. 10 per cent would have to provide over 1 Terabyte of storage capacity to analyze CDRs for 3 years. [16, p.8] Besides this there are also legal issues to data growth, e.g. in Germany, telecommunication providers are obliged to store all call detail records (CDR) for the minimum of 80 days. [15; 17]

The sheer data volume has negative effects on the performance and the response time of the data warehouse. But the importance of these measures depends on the kind of analysis task that is executed on the data warehouse.

1. Exhaustive CRM analyses e.g. for the segmentation of the customer base may need to access many data in the data warehouses, but they are not expected and required to provide results within seconds, whereas
2. Operational CRM applications may need to access only few derived customer segment values from the data warehouse. But they are required to deliver them in near-real-time, as they are utilized in customer touch points such as call centers or web extranet portals.

As can be seen, the system requirements *data amount* and *response time* are conflictive criteria for a data warehouse. But both criteria have their justification and need to be covered by an appropriate data warehouse concept, which will be presented in the following.

6 Customer-centric Data Warehouse Architecture

Depending on the CRM operations that should be executed on the data warehouse, the architecture should fulfill basic requirements, which are specified in the following table:

	Analytical CRM	Operational CRM
Usage of Customer Relevant Data	Extensive analyses, complex queries expected	Standard reporting queries
Require-ments	Detail data	Near-real-time query responses, pre-aggregated data

For these fundamentally different kinds of data warehouse utilization, the following data warehouse architecture aims to fulfill the oppositional, even contradicting requirements. It is regarded as an extension on the architectures of [5] and [2].

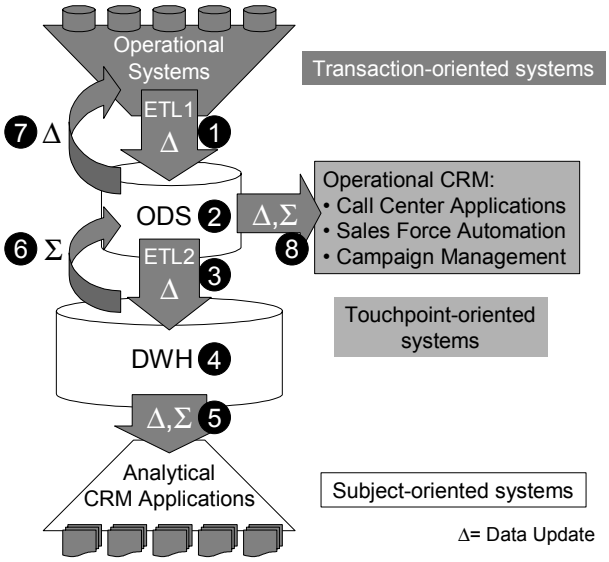

Figure 2: Customer-centric Data Warehouse Architecture

Explanation of the data flow through the architecture along the marked points:

1. The ODS extracts operational data from the transaction systems via the data interfaces several times per day. In this first-level ETL step, the data are merely copied or only lightly transformed, so that the high performance of the ODS is not diminished. An important issue is the time stamping of incoming data for building the data history later. In the customer-centric context, this especially applies to customer data.

2. The ODS contains a short history of young, low-level detail data. Depending on the computer capabilities, the ODS might contain e.g. one month's

data, which can be managed in form of a current rolling month[5].

3. In a second-level ETL step, data is periodically transferred from the ODS into the data warehouse: either the ODS pushes the data into the data warehouse, or – referring to Section 4.1 – the data warehouse pulls the data from the ODS. This depends on the position of the staging area in the architecture and the capabilities of the database-immanent data management tools. In this ETL step, the filtering, harmonization, aggregation and enrichment of the data are executed and claim a large portion of the data warehouse resources.

4. The data warehouse contains both detailed data histories and calculated values, such as the *derived customer data*. The General Conceptual Model by Todman [18] is considered as a suitable fundament to structure the core of the customer-centric data warehouse architecture.

5. The data warehouse provides updated detail data and derived data to analytical applications, in the customer-centric context e.g. to analytical CRM.

6. This is an important characteristic of this architecture from a performance point of view: The data warehouse may also write back summarized data into ODS. Thus, a large number of standard queries – which would otherwise block resources for more complex queries – can be shifted from the data warehouse to the ODS.

7. In an idealized architecture, a ODS may also write back data updates to operational systems and, thus, may maintain the consistency between operational systems. Example: a change of a customer attribute in one transaction system triggers the ODS to re-perform this change automatically in other relevant transaction systems. This would require an idealized, highly integrated infrastructure of operational and dispositive systems. A solution to this matter would pose another interesting field of research.

8. This second important characteristic of this architecture refers to the former item 6: The ODS provides both updated data and derived data to operational CRM applications. This approach may contribute to the performance of the data warehouse, as it can be freed from a large volumes of standard queries, which can now be executed on the ODS. In this architecture, operational CRM applications are intentionally not counted to operational systems, for they usually provide the data interfaces to standard databases and are designed to be customized to existing data infrastructures. Thus, we do not regard them so much as transaction-oriented in this architecture, despite their operational nature. To differentiate them from transaction systems, we rather characterize operational CRM applications as *oriented-oriented*, as they are utilized at customer touch points.

[5] Refer to Section 4.1 in this text and [9].

7 Discussion

Referring to Section 4.1:*Definition and Comparison of Data Warehouse and ODS*, we assume that the ODS in the customer-centric architecture receives aggregated and derived data from the data warehouse. This may also apply to the case of a dedicated CRM data mart, so a justified question about the necessity of a dedicated ODS in a customer-centric architecture can be raised. Our answer is that it depends on the relative position of the CRM data mart to the data warehouse.

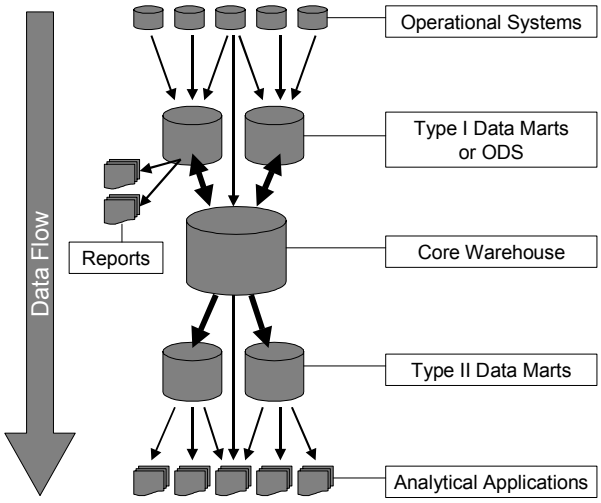

Figure 3: Hub-and-spoke architecture of a Data Warehouse

Inmon defines a data mart as "a collection of data tailored to the DSS processing needs of a particular department". [5, p.110] It may be added that they represent a controlled-redundant subset of the data warehouse. In a hub-and-spoke architecture [6] – in which the data warehouse represents the hub and the data marts represent the spokes (see Figure 3) –, two types of data marts can be made out:

- Type I data marts have data interfaces to the operational systems and are populated with their data. Data marts transform these data to the requirements of the mart owners and serve also as data providers for the core data warehouse. They may also receive aggregated and calculated data from the data warehouse.

- Type II data marts are more commonly used in data warehouse architectures: they are controlled-redundant subsets of the data warehouse, populated with processed data by the data warehouse. Data in the data marts are often transformed in a second–level ETL step to the requirements of reports or analytical applications.

Comparing the characteristics of the data mart types and the ODS with regard to the customer-centric architecture, it can be said that only
- a Type I data mart,
- which is dedicated to contain customer relevant data,
- which has high update frequency similar to an ODS and
- which serves as a data provider for a data warehouse

might actually function similar to a Customer ODS. But as Type I data marts are not as commonly used as Type II data marts, the ODS seems to be justified as a self-contained element in the customer-centric data warehouse architecture.

A Type II data mart is per se not qualified to be redefined as an ODS, as it does not import data from the operational systems.

8 Summary, Conclusion, Contribution to Research

This paper has elaborated on the implications of customer-orientation on data warehousing and presented an extended data warehouse architecture dedicated for CRM. It was one goal to show that it is not sufficient to change organization and to buy new software when implementing CRM and customer orientation in an enterprise. When the customer gets into the center of the corporate interest, there are serious implications on the data which management relies on for decision support. These implications should be absorbed with a customer-centric data warehouse model.

The aim of the presented architecture was not only to show *how* to provide for consistent customer data, but also to give an idea how to improve query performance by balancing the query load between the data warehouse and the ODS.

It is not necessarily a goal of this research to create a new process model on the implementation of customer orientation in data warehouses, but the well-known incremental development model is recommended as an appropriate approach: A CRM implementation project does not need to initiate a full-size enterprise data warehouse, as a CRM data mart may be sufficient for the start. It can even grow to a large data warehouse, but the crucial point to this is that the predisposition of the data mart – its structure – should be defined and fixed beforehand. The structure should be conforming to the overall context of a data warehouse still to come, otherwise the advantages of enterprise-wide data integration will be lost and an integrated, consistent view on the customer becomes impossible.

9 References

[1] Hildebrand, V.G.: Kundenbindung mit Online-Marketing. In: Link, J.(ed.): Wettbewerbsvorteile durch Online-Marketing. Springer, Berlin et al 1998, pp.53-73.

[2] Imhoff, C., Loftis, L; Geiger, J.G.: *Building the Customer-centric Enterprise*. John Wiley & Sons Inc., New York/ NY, 2001

[3] Inmon, W.H.: *Building the Data Warehouse*. John Wiley & Sons Inc., New York/ NY, 1992

[4] Inmon, W.H.: *Building the Operational Data Store, 2nd Ed*. John Wiley & Sons Inc., New York/ NY, 1996

[5] Inmon, W.H.: *Corporate Information Factory*. John Wiley & Sons Inc., New York/ NY, 2000

[6] Inmon, W.H.: Wherefore Warehouse ? In: Byte.com, January 1998. http://www.byte.com/art/9801/sec6/art1.htm, last accessed on May 31st, 2002.

[7] Kemper, H.G.: "Conceptual Architecture of Data Warehouses – a transformation-oriented view." In: Proceedings of the 2000 Americas Conference on Information Systems (AMCIS), Aug 10-13, 2000. H. Michael Chung, Long Beach/CA, pp.113-118.

[8] Kemper, H.G.; Lee, P.: Business Intelligence - Integrierte Lösungen für eine erfolgreiche Unternehmenssteuerung. In: Computerwoche Iss. 44, 2001, Nov. 2nd , 2001, pp.54.

[9] Kimball, R.: Relocating the ODS. In: DBMS online – www.dbmsmag.com, December 1997. http://www.dbmsmag.com/9712d05.html, last accessed on May, 28th, 2002.

[10] Kimball, R.: *The Data Warehouse Toolkit – Practical Techniques for Building Dimensional Data Warehouse*. John Wiley & Sons Inc., New York/ NY, 1996.

[11] Kotler, P., Armstrong, G.: Principles of Marketing. Englewood Cliffs/NJ 1996

[12] Neumann, J.: *Data Warehousing – Die Infrastruktur des CRM*. Saracus Consulting, May 21st, 2001. http://www.competence-site.de/crm.nsf/ 9297EDC526C740D3C1256A5 30066C514/$File/dwh_infrastrukturcrm__saracus.pdf, last accessed on May 14th, 2002

[13] Peter, S.I.: Kundenbindung als Marketingziel – Identifikation und Analyse zentraler Determinanten. Gabler, Wiesbaden/Germany 1997, pp.53.

[14] Piller, F.T.: Mit Mass Customization zu echtem Beziehungsmarketing. In: Harvard Business Manager, Iss. 6, 1998, pp.103-107.

[15] RegTP (Regulierungsbehörde für Telekommunikation und Post der Bundesrepublik Deutschland): Verfügungsblatt 168/1999 im Amtsblatt 23/1999, Teil 3, p.6. http://www.regtp.de/imperia/md/content/tech_reg_t/tkv/5.pdf. Last accessed on Sept. 25th, 2002

[16] Sand Technology Inc.: Sand Customer Insight Solution Telekommunikation, Whitepaper, http://www.sandtechnology.de, last accessed on Sept. 25th, 2002

[17] Bundesrepublik Deutschland: Telekommunikationsdienstunternehmen - Datenschutzverordnung – TDSV, BGBl. (Bundesgesetzblatt) I 1996, p. 982-987, July 19th 1996, §6 Abs. 3.

[18] Todman, C.: Designing a Data Warehouse Supporting Customer Relationship Management. Prentice Hall, Upper Saddle River/ NJ, 2001