Data Warehousing for Bavarian Out-Patient Public Health Care

Katharina Wirtz[1], Martin Tauscher[1], Maria Zwerenz[1], Axel Munte[2]

[1] Vertrag und Strategie
Kassenärztliche Vereinigung Bayern
Elsenheimerstr. 39, 80687 München
[Katharina.Wirtz, Martin.Tauscher, Maria.Zwerenz]@kvb.de

[2] Vorstandsvorsitzender
Kassenärztliche Vereinigung Bayern
Elsenheimerstr. 39, 80687 München
Axel.Munte@kvb.de

Abstract: The KVB Data Warehouse enables the Bavarian Association of Statutory Health Insurance (SHI) Physicians to track public health data in a centralised way. In a sophisticated process, healthcare statistics are retrieved and users can perform multidimensional analysis. Given its unmatched breadth and depth of data there is no comparable public health information system in Germany. It contains the data of all Bavarian SHI services and diagnoses since 1997. The following case study describes the Data Warehouse (DWH), its context and scope.

Abbreviations

DWH	Data Warehouse
HITV	Health Insurance Treatment Voucher
ICD	International Classification of Diseases
KBV	Kassenärztliche Bundesvereinigung, National Association of Statutory Health Insurance Physicians
KVB	Kassenärztliche Vereinigung Bayern, Bavarian Association of Statutory Health Insurance Physicians
SHI	Statutory Health Insurance

1 Introduction

The concept of data warehousing has gained increasing attention over the past 15 years, because of the need to retrieve information from operational data. The success of the DWH concept is based on a simple yet powerful idea: duplicating the operative data by organizing it in a simple structure optimal for retrieval purposes – called star schema

[Wk2002]. This concept has gained wide acceptance in the data warehousing community.

It is, however, difficult to restructure complex operative health care data into this simple star schema, because the relationship between money paid in by the insurance organisations and money taken out as physicians fees is indirect and complicated. The Bavarian Association of Statutory Health Insurance Physicians has found one solution for this modelling problem and designed its DWH System, which is a significant example of an information system in the field of public health.

To comprehend this DWH System it is necessary to gain an understanding of the function and processes of the association, which are described in chapter 2 and 3. The DWH System itself is best described by its architecture and data model (chapter 4) and its scope (chapter 5). The article ends with a discussion on the specific contribution of the described DWH solution to health economics.

2 The KVB

The Kassenärztliche Vereinigung Bayern (KVB) is the Bavarian Association of Statutory Health Insurance (SHI) Physicians, providing services for SHI physicians and psychotherapists as well as for patients. There are 17 such regional association in Germany with their umbrella association in Berlin, the National Association of Statutory Health Insurance Physicians. Among the 17, the KVB is the largest association with more than 22.000 members. As a public organisation it has the following missions:

- Accounting for out-patient medical services
- Guaranteeing provision of services:
  As regulated in Social Code Book V, the provision of services for SHI patients in Bavaria has to be accessible, sufficient, appropriate and efficient.
- Assuring provision of medical care:
  The KVB takes responsibility for quality and efficiency. It audits the accounting of their members and provides training programs for them.
- Representing its members' interests, e.g. in contract negotiations with health insurance organisations

3 The KVB – accounting process

The data of the DWH is mainly extracted from KVB accounting systems. Apart from the accounting data there are pharmaceutical data from data processing centres for pharmacies and master data about services and diagnoses from the National Association of Statutory Health Insurance Physicians and other public health institutions. To understand the informational content of the DWH you need to have a basic grasp of the accounting process which is described in the following top-level context-diagram .

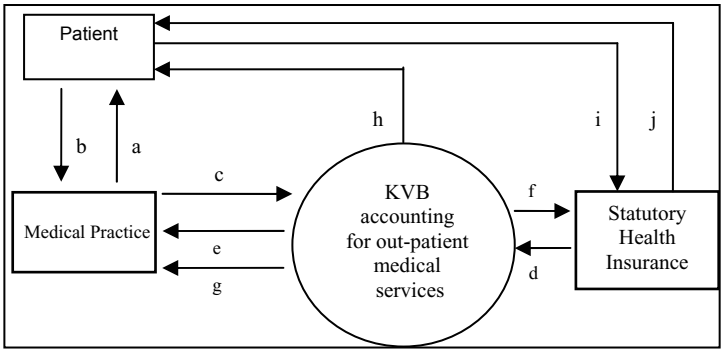

Figure 1: Context diagram accounting process

a) The patient visits the medical practice to receive treatment and
b) has his health insurance card scanned in.
c) The medical practice forwards the information of the health insurance treatment voucher of all SHI-services to the KVB. The data are mainly transferred by electronic data interchange via diskettes or internet. The remaining vouchers are carried physically to the KVB offices and scanned in. The vouchers contain data about the services, diagnoses and patients.
d) The insurance organisations pay a per capita fee for each insured member (no fee is paid for co-insured immediate family) plus a certain amount for extra-budgetary services as agreed with the KVB. The lump-sum payment is app. 75% of overall insurance remuneration. The KBV and the central associations of the insurance organisations define fee schedule items for all SHI-services as well as conditions to guarantee the quality of the services based on statutory provisions. The KVB audits compliance with medical service accounting regulations.
e) As a consequence of the per capita fee, the amount to be allocated to the SHI-services is quoted and points for fee schedule items are used for accounting most services, the so-called budgetary services. If the volume of services provided in a quarter is high, then the point values are correspondingly low and vice versa, so that the budget is in any case sufficient. In that way the morbidity risk for the budgetary services is carried by the physicians, since the remuneration for these service is quoted but not the frequency of the service requests. App. 25% of the services are accounted extra budgetary and charged in Euro. In this case the insurance organisations carry the morbidity risk, since they pay each service provided with a fixed value. After auditing and rating the HITV claims, the remuneration has to be distributed in a way that each speciality group receives their stipulated part of the overall budgetary remuneration. (Speciality groups are groups of specialist physicians such as gynaecologists, general practitioners.) The KVB defines the point values for the fee schedule items and finally transfers the medical fee to the practices each quarter.
f) The KVB sends information about the payments of the out-patient medical services to the insurance organisations. Based on this information the insurance remuneration of the services is calculated for the next quarter.

g) Auditing of the accounting process is carried out in two steps: There is a standard audit which is part of the accounting process. The second audit is done after the completion of accounting in spot tests. While it may seem strange that auditing is partly done after completion of accounting, the task of accounting 200 million services as fast as possible and on the other hand the very time consuming process of auditing must be considered. In the audit process the KVB audits the HITV-Data in terms of the compliance of physicians´ accounting with the conditions as stipulated. This is done mainly by standard accounting procedures. Furthermore the KVB audits
   - the appropriate cost of the service provision,
   - budget compliance of the individual practice and,
   - the plausibility of their statements.

   The KVB corrects statements that are not in accordance with the stipulated conditions. Should suspicions of implausibility and inefficiencies arise, the KVB investigates by interviewing physicians and patients.
h) KVB provides public health information to SHI - patients. When physicians are off duty their patients can avail themselves of medical treatments organized by KVB on-call service.
i) The patient is insured with a statutory health insurance organisation of his own choice and pays his contribution
j) The insurance organisation guarantees the coverage of SHI – services for its patients.

The accounting of medical service is a complex process with many steps and intermediate results. After app. 10 days the services are processed and a first analysis can be done. The ensuing time-consuming process of service rating and distribution of remuneration takes nearly three months. The value of each service can change at any step of the accounting process, since different rules are applied at each step to guarantee the stipulated conditions. For example, in the first step a standard value is used which is replaced in the second step by a higher value, if a physician meets certain conditions, such as graduation from a certain training program or a minimum number of services supplied. To quote another example: If a practice exceeded its budgets for a given service, the payment for that service is very low. Not only can individual budgets have an impact on the service value but also the budget of the speciality group as a whole. In that way, the value of a service can change several times in the accounting process. All different value states are interesting and saved as intermediate result, as they help explain the accounting process. The data generated during that process form the basis for the KVB DWH.

4 The KVB DWH System

In order to understand how the components involved in the KVB DWH system interact, it is essential to understand its architecture and its data model which are described in the following chapters.

4.1 DWH Architecture

Figure 2 depicts the KVB DWH system.

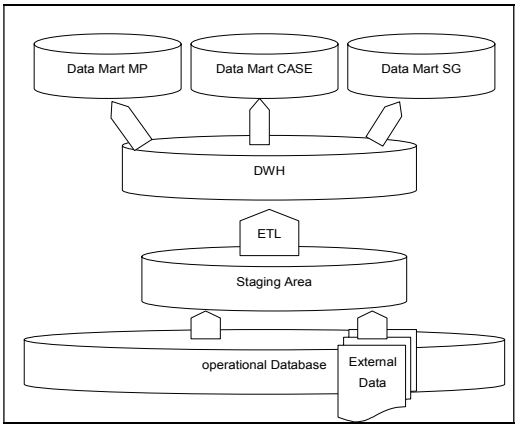

Figure 2: KVB DWH System

The **operational database** is used for the accounting process. The **DWH** is loaded from an export of the operational database called the **staging area** so as not to disturb the accounting process. The loading takes place quarterly due to the fact that the accounting is processed once in a quarter. During the DWH loading process a data-volume of app. 150 GB is extracted, cleaned, transformed and loaded (ETL). The loading-process takes about 3 days. All analytical data since 3/1997 are stored in the DWH.

For performance reasons the DWH is aggregated in form of **Data Marts** [Ka99]. Levels of aggregation can be implemented by each set of dimension attributes. Since there are 12 dimensions with an average of 13 attributes there are $\sum_{a=1}^{12*13}(\frac{(12*13)!}{a!((12*13)-a)!})-1$ that is about 3.418.931.628.970.130.000 possible combinations of aggregates. Obviously, not all of them make sense, given that combinations are only sensible if aggregated on different hierarchical levels [Kr95]. But there are still a lot of reasonable combinations, which could be saved as aggregate tables causing a large number of views. Therefore the aggregates are built on user demand. There are three different types of aggregates at the moment. KVB as a service provider for physicians focuses on practice level and speciality group level aggregates: the **Data Mart MP (Medical Practice)** aggregates the performance values of each medical practice by means of 5 fact-views and the **Data Mart SG (Speciality Groups)** covers all aggregates for speciality groups in a multidimensional database (Oracle Express) for high performance queries. Finally there is the **Data Mart CASE** with 3 aggregates on cases. There are plans to offer more aggregates to keep response time down.

4.2 DWH Data-Model

In this chapter the base views of the DWH Data-Model will be described. All the KVB accounting data are stored in two base star schemas. They contain atomic business data items: each single service or diagnosis for every single day. All other fact-views are aggregates of the base fact-views.

Since services and diagnoses are not related to each other on a HITV they need to be modelled in two different star-schemas: the fact-views F_Service and F_ICD as well as the most important dimensions will be described in the following chapters in order to give a general idea about KVB DWH. The description is a simplification of our warehouse schema in which no mention will be made of attributes and dimension hierarchies.

4.2.1 F_Service

The star-schema of **F_Service** (Fact-view of Services) supplies information about the process of medical service. All services paid by the SHI and therefore accounted by KVB are classified within the doctor's fee scale. F_Service is the central fact view, and the surrounding views are the descriptive dimensions. The fact contains the quantifiable attributes of medical services such as

- frequency of service provision
- value of services at various stages of the KVB accounting process. These facts reflect the impact of the use of rules and regulations.

F_Service has app. 200 million records each quarter.

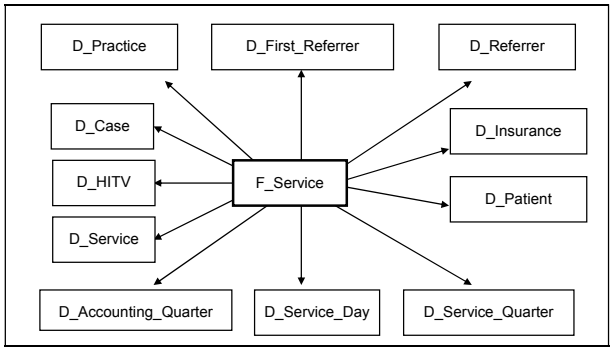

Figure 3: Star-Schema F_Service

4.2.2 F_ICD

The second base view is called **F_ICD** (Fact-view of International Classification of Diseases). It contains information about the diagnoses physicians have documented on the HITV. Since diagnoses and services are both documented with the same information on the HITV, the dimensions are basically the same. There is only one very obvious difference: Diagnoses cannot be described by the dimension D_Service but by the dimension D_ICD, which is highlighted in the diagram. Therefore the two base-views F_ICD and F_Service cannot be related to each other at the most granular level. The F_ICD contains the frequency of diagnoses and has app. 80 Million records per quarter.

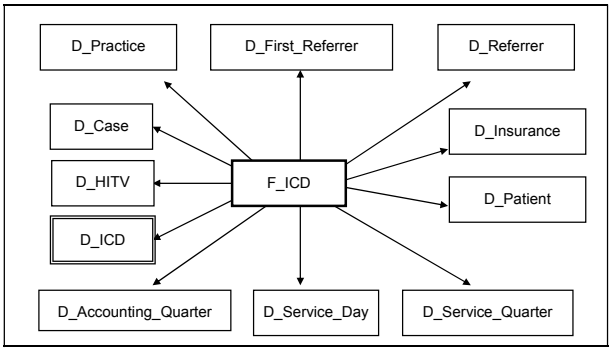

Figure 4: Star-Schema "F_ICD"

4.2.3 Dimensions

The fact-views relate to a set of dimensions. There are several dimensions expressing different roles of the medical practice, depicted at the top of the diagram: **D_Practice**, **D_First_Referrer**, **D_Referrer**. In the present article the model is described in terms of user views. There are some differences in table and view structure, e.g. there is only one table for a practice, but several views to depict the roles of a practice such as referrer or practice. The personal data of the practice and the patient is encrypted for reasons of protection of data privacy. The policies of data privacy are implemented by means of role administration on views.

Since the billing is done by each practice rather than by each physician, data are also expressed with the grain of practice. Medical practices refer patients to other practices for treatment. When more than one laboratory is involved in one case, it is important to know, which medical practice first referred the patient to the laboratory, since that practice has to cover any laboratory costs out of its budget.

Another set of dimensions, depicted at the bottom of the diagram, is needed for time series analyses: **D_Accounting_Quarter**, **D_Service_Quarter**, **D_Service_Day**. The redundancy of the dimension D_Service_Quarter is notable, since all the data contained in D_Service_Quarter is also covered by D_Treatment_Day. However users want the D_Service_Quarter as a separate dimension. Since nearly all analysis is done on a quarterly basis, only specific questions are to be answered on a finer time granularity. For this reason the redundancy of dimensions was allowed to enhance usability.

On the left side of the diagram there are dimensions describing the facts, **D_HITV**, **D_Service** and **D_ICD** respectively, as well as **D_CASE**. The right hand side dimensions, **D_Patient** and **D_Insurance**, describe external instances.

At the users` request all dimensions are modelled as a full set of records per quarter, so each dimension contains the attribute QUARTER. This is acceptable, since the descriptions frequently change and users often analyse dimensional data only in their historical context. In a sense this implies a fourth type of slowly changing dimensions (SCD) [Kr95]: Each quarter a record is generated for each new description. That way the user can query all services valid in a given quarter or compare patients of different quarters with regard to gender and insurance status without joining with a fact-view. That is important, considering that fact-views can consist of 200.000 million records per quarter. The size of the dimensions is negligible with the exception of the dimension patient, which contains about 9 million records per quarter.

It is also easier to load and administer the warehouse as a complete set of quarterly data. If data changes after a backup has been done, it becomes difficult to keep the backup in order, unless data is saved as a full snapshot of each quarter.

5 Scope of the DWH-System

I'll furnish what you wish and more. It's true,
It is a light task, yet the light's a burden too.
The gold lies there and yet to win it,
That is the art - who knows how to begin it? [Gj95]

The DWH itself is not of any primary use for the organisation. Value is added to an organisation if the information queries are the basis for better decisions or an improved perception of the organisation.

Our original intention in implementing the KVB DWH was to support negotiations with the insurance organisations and evaluate the contract conditions. It was assumed that comprehensive information would improve the negotiating position of the KVB. This assumption was proven true on many occasions, e.g. KVB was able to verify an increase

of morbidity and enforce a claim of higher per capita fee against the insurance organisations. In the meantime the DWH has turned out to be useful for most strategic and operative information queries

5.1 Strategic analysis

Supporting the strategic positioning of the KVB with information is the focus of the DWH strategic analysis. The original motivation of building the DWH – supporting the contract negotiations - is still one of the key tasks in this context. Model calculations are carried out to see the impact of decisions, changes in law and context, e.g. the impact of the Local Health Care Fund contract for general practitioners within the scope of integrated care. The validity of publications by other organisations can be checked against our data and commented on. KVB is in the best possible position to comprehend the gains and losses resulting from changes, which enables it to influence these changes.

For information support there are **strategic standard reports** providing information needed at all times, such as the performance measure "cases per physician" or "number of patients per physician". KVB runs multidimensional (implemented with Oracle Express) as well as relational standard reports. Multidimensional reports are the "Fee and Service Reports" with drill down and slicing for all relevant dimensions. The data are aggregated on specialty group and district office. The user interface operates as an aggregate navigator greatly simplifying the task of choosing the right aggregate each time the user changes his information query. The preaggregation also has an enormous impact on response time. Some of the queries returning within seconds would run for hours if not days on the DWH.

Most of the information requests cannot be answered by standardised reports, but by **ad hoc information queries**. Therefore KVB has a group of strategically skilled users (around 10 DWH analysts), creating management reports with SQL-Query and statistical tools. Apart from management information they provide information for public support and pensions research.

Figure 5 is an example of a report generated in response to an ad hoc query to support the KVB board with information geared to guaranteeing provision of services. This illustrates the increase in average age between 1999 and 2003.

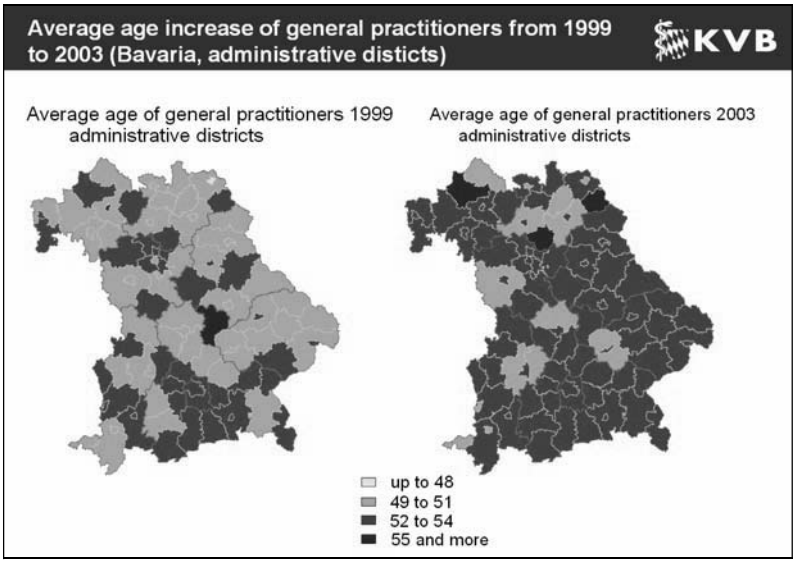

Figure 5: Strategically Standard Report

The quarterly report for the KVB Board and medical associations "Fee trends per individual and group practice" (Figure 6) is a good example for a standard report. It compares the most significant service measurements of a quarter with the corresponding quarter from the previous year, to analyse the development of the services of each specialty group.

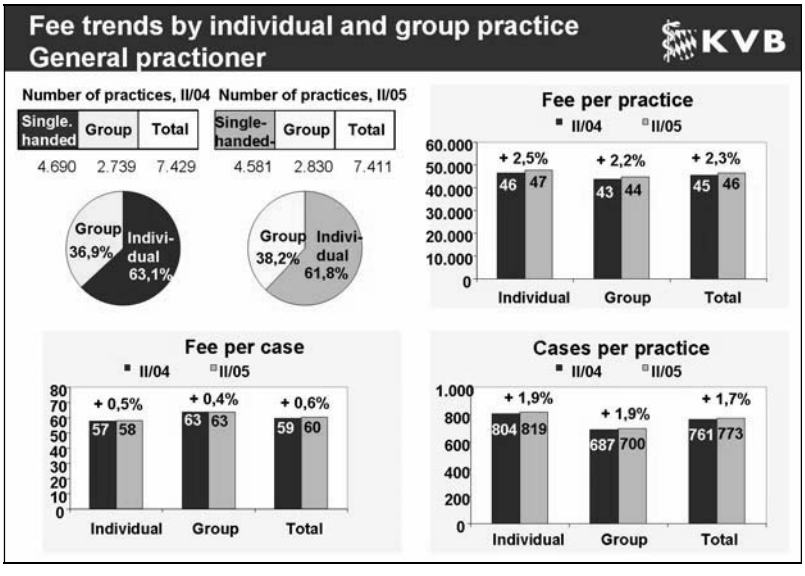

Figure 6: Strategic Report

5.2 Operational Reports

Apart from its strategic use for decision support, the value of KVB Data Warehousing was also discovered recently by the operative business. Since last year app. 100 operational users have been employing standard reports for quality control of the accounting process. KVB did not implement a separate ODS [Iw99], but instead the DWH stores detailed information for auditing the physicians service provision as well as accounting and reporting of service quality. For instance, practices receive reports showing their personal prescribing patterns and their diagnosis documentation in relation to their speciality group each quarter. Figure 7 depicts an example of the kind of report supplied to every practice showing the documentation of diagnoses in relation to each practice's speciality group.

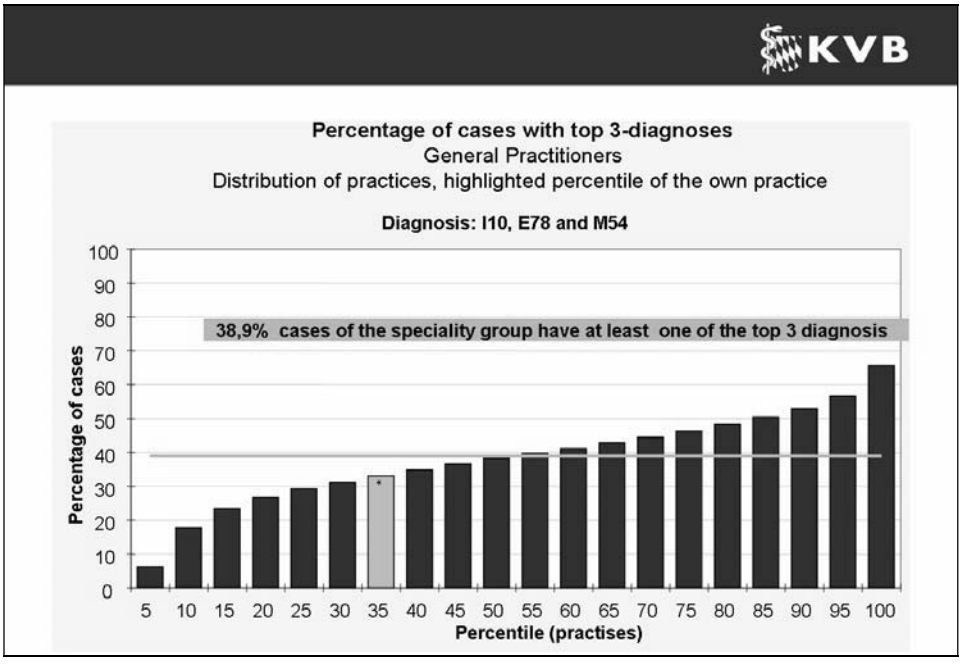

Figure 7: Operational Report

6 Concluding Remarks

State of the art technologies are used in the above described DWH-system. The specific value of the KVB DWH-System lies not in technology but in its informational content. Public Health is a big topic in Germany, costing about 239 billion Euro [NL05] each year and employing app. 11% of all workers according to the Federal Statistical Office [Wj02]. The demographic trend towards an aging society and the medical progress which enables the treatment of more and more diseases, leads to rising public health costs. Therefore there is a need to consistently try to find better and more efficient approaches to reallocating resources. To evaluate the efforts of the approaches a reliable information source is needed. The KVB DWH system could be the first choice.

There is no comparable information system in the field of out-patient public health care in Germany, which since 1997 has stored health care data and costs in an integrated, quality secured and easy to access manner. The KVB DWH-System can be regarded as a very valuable and unique pool of data for health economy. Unlike field research patients, physicians and their relationships (in terms of services and diagnosis) can be tracked over many years. Since app. 90% of the Bavarians are members of the CHI the KVB DWH-Systems also delivers a basis for population based studies. The primary organisational use of the DWH-System was described in chapter 5, but the societal use in terms of research has only recently been discovered.

Furthermore, the developed DWH-system is reusable for the other 16 regional associations. Since the accounting processes of the regional Association of Statutory Health Insurance Physicians within Germany are very similar, the described DWH-system could be used as an example to implement DWH-systems for other regional associations. There are already attempts to implement the DWH-system for the association of Südwürtemberg.

Bibliography

[Iw99] Inmon, W.H.: Building the Operational Data Store. John Wiley & Sons, 1999.

[Gj95] Von Goethe, J.W.: Faust – Erster und zweiter Teil. Deutscher Taschenbuch Verlag GmbH & Co.KG, München 1996, p. 147.

[Ka99] Kennel, A.: Wie viele Dimensionen hat ein Würfel? Der Einsatz von Datamarts als Analysehilfen in einem Data Warehouse. In: INFORMATIK/INFORMATIQUE 1/1999, SVI/FSI, pp. 3-7, Zürich 1999; p. 3.

[Kr95] Kimball, R.: Data Warehouse Toolkit. John Wiley & Sons, New York u.a. 1995; pp. 100-106

[NL05] Neubauer, G.; Lewis, P.: Gesundheit als Wirtschaftsfaktor im Untersuchungsraum. München, final report, november 2005.

[Wj02] Weinmann, J.: Innovationsreport. http://www.innovations-report.de/html/berichte/statistiken/bericht-10211.html , 2002-05-28, referenced on 2006-03-17.

[Wk02] Wirtz, K.: Der Data-Warehouse-Rahmenplan. Gabler Edition Wissenschaft, Wiesbaden 2002; pp. 64–68; pp.187-209.