

Data Warehousing Technology for Surgical Workflow Analysis

Thomas Neumuth
University of Leipzig, ICCAS
Semmelweisstr. 14
04103 Leipzig, Germany
thomas.neumuth@iccas.de

Marc H. Scholl
University of Konstanz
Box D188, FB Informatik
78457 Konstanz, Germany
marc.scholl@uni-konstanz.de

Svetlana Mansmann
University of Konstanz
Box D188, FB Informatik
78457 Konstanz, Germany
svetlana.mansmann@uni-konstanz.de

Oliver Burgert
University of Leipzig, ICCAS
Semmelweisstr. 14
04103 Leipzig, Germany
oliver.burgert@iccas.de

Abstract

Analysis of surgical procedures is an emerging domain of medical engineering aimed at advancing surgical assist systems. The term Surgical Workflows is used for describing the methodical framework for acquiring formal descriptions of surgical interventions. A formal model and a uniform recording scheme of a surgical process are crucial for systematic accumulation of relevant data from running surgical instances in a form, appropriate for analysis. This paper describes the process of designing a data warehouse for surgical workflow analysis. Data warehousing technology, originally developed for quantitative business data analysis for the purpose of decision making, is adaptable to the requirements of surgical data and analysis. We describe a conceptual model of a surgical procedure obtained in accordance with the multidimensional data model. We demonstrate how a subject-oriented multidimensional perspective of a surgery and its components enables powerful analysis and exploration by defining the metrics of interests and aggregating those metrics along various dimensions and levels of details. Apart from its primary function, i.e. quantitative analysis, the data of the surgical data warehouse may serve as input for other data-intensive systems, e.g. visualization and data mining tools.

1. Introduction

An emerging field of surgical workflow analysis fosters intelligent acquisition of process descriptions from running surgical interventions for the purpose of their clinical and

technical analysis [19]. This type of analysis is crucial for the development of surgical assist systems for the operating room of the future [12]. Besides, it provides a framework for evaluating newly developed devices or surgical strategy evolution. A medical engineering term *Surgical Workflows* describes the underlying methodological concept of this acquisition procedure and its post processing. The process data is obtained manually and semi-automatically by monitoring and recording the course of a surgical intervention. Apparently, a well-defined formal recording scheme of the surgical process is required to support such data acquisition [17]. Surgical workflow use cases are manifold, ranging from the preoperative planning support by retrieving similar precedent cases to the postoperative workflow exploration, from discovering the optimization potential for instrument and device usage to verifying medical hypotheses, etc.

1.1 OLAP and Multidimensional Data Model

Data warehousing and OLAP(On-line Analytical Processing) [8, 3] emerged in the 90s as a new technology for advanced user-driven information access. Applicability of data warehousing is by no means restricted to business scenarios. As a generic framework for providing data analysis, data warehouses are increasingly deployed in non-business domains, such as government, science, education, research, medicine, etc. OLAP technology draws its analytical power from the underlying multidimensional data model. The data is modeled as cubes of uniformly structured facts, consisting of analytical values, referred to as measures, uniquely determined by descriptive values drawn from a set of dimensions. Each dimension forms

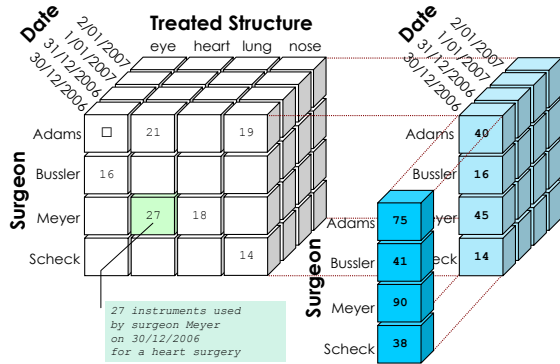


Figure 1. A sample 3-dimensional cube (fragment) storing instrument usage numbers per surgical intervention.

an axis of a cube, with dimension members as coordinates of the cube cells storing the respective measure values. Figure 1 shows a strongly simplified example of a 3-dimensional data cube, storing instrument usage statistics (measure number of instruments) determined by dimensions Surgeon, Treated Structure, and Date. Besides the original cube storing the data at the finest granularity, Figure 1 also displays the results of two “roll-up” operations that summarize the measure across all treated structures, and subsequently, across all dates. In real-world application, data cubes may have arbitrarily many dimensions, and are therefore denoted *hypercubes*.

Member values within a dimension are further organized into *classification hierarchies* to enable additional aggregation levels. Dimension hierarchies are strictly structured, i.e. values at each level of a tree belong to the same *category*. Multiple hierarchies may be defined within a dimension, and can be mutually exclusive (e.g., Date can be aggregated by month or by week, but not both), denoted *alternative*, or non-exclusive, or *parallel* [22].

1.2 Related Work

Due to the multidisciplinary nature of our research, the related work falls into multiple categories, such as medical engineering, data warehousing, and business process management. A pioneering research on designing scientific methods for surgical workflows is carried out at the Innovation Center Computer Assisted Surgery (ICCAS) located in Leipzig, Germany. Major directions of their projects are surgical workflow formalization [17, 19], semantics [2], analysis [20], standardization [1], visualization [18]. Other approaches to surgical process modeling were contributed by several groups ([13], [9], [21], [24]). All approaches

are concentrated on aspects of modeling surgical processes, none of them aims user-oriented data representation. Application of the data warehousing techniques to business process management and analysis is commonly referred to as *Business Process Intelligence* (BPI). Dayal et al. [5] suggest that a business process data warehouse can be built from execution logs by cleaning, transforming, and merging the data. An example of a BPI tool suite based on a data warehouse approach is given in [7]. The process data is modeled according to the star schema, with process, service, and node state changes as facts and the related definitions as well as temporal and behavioral characteristics as dimensions. This approach enables analysis of process execution and state evolution in the environments where processes have a uniform and well-defined scheme. Health care applications are frequently encountered in the data warehousing literature in the role of motivating case studies and usage scenarios. Pedersen et al. proposed an extended multidimensional data model for meeting the needs of non-standard application domains and at the example of accumulating patient diagnosis data [23]. Mansmann et al. describe how surgical process modeling, used as a challenging case study in data warehousing, results in the necessity to extend the conceptual foundations of the multidimensional data model [14] and present the implications of introducing the extensions at the conceptual level to the implementation of the data warehouse and the front end tools for interactive analysis [15]. To the best of our knowledge, the presented contribution is unique in a sense that it adopts established technologies from the domain of business intelligence for providing a solution for a highly-specific sub domain of health care systems.

2 Structuring A Surgical Process

Surgical process modeling is comparable to the domain of business process modeling [19], in conformity with the established definition of a business process as “a set of logically related tasks performed to achieve a defined business outcome” [4]. In this article, the terms “process” and “workflow” will be used as synonyms.

While the concepts *process* and *workflow* are used interchangeably by some authors [27], diverse workflow definitions can be found in the literature. One popular interpretation is that business processes output products while workflows deliver services [25]. Another use of the term workflow is to denote the control flow, i.e., the dependencies among tasks during the execution of a business process [26]. In this work, we adopt the differentiation in the levels of abstraction proposed in [16]: while business processes are mostly modeled in a high-level and informal way, workflow specifications serve as a basis for the largely automated execution and are derived by refining the business process

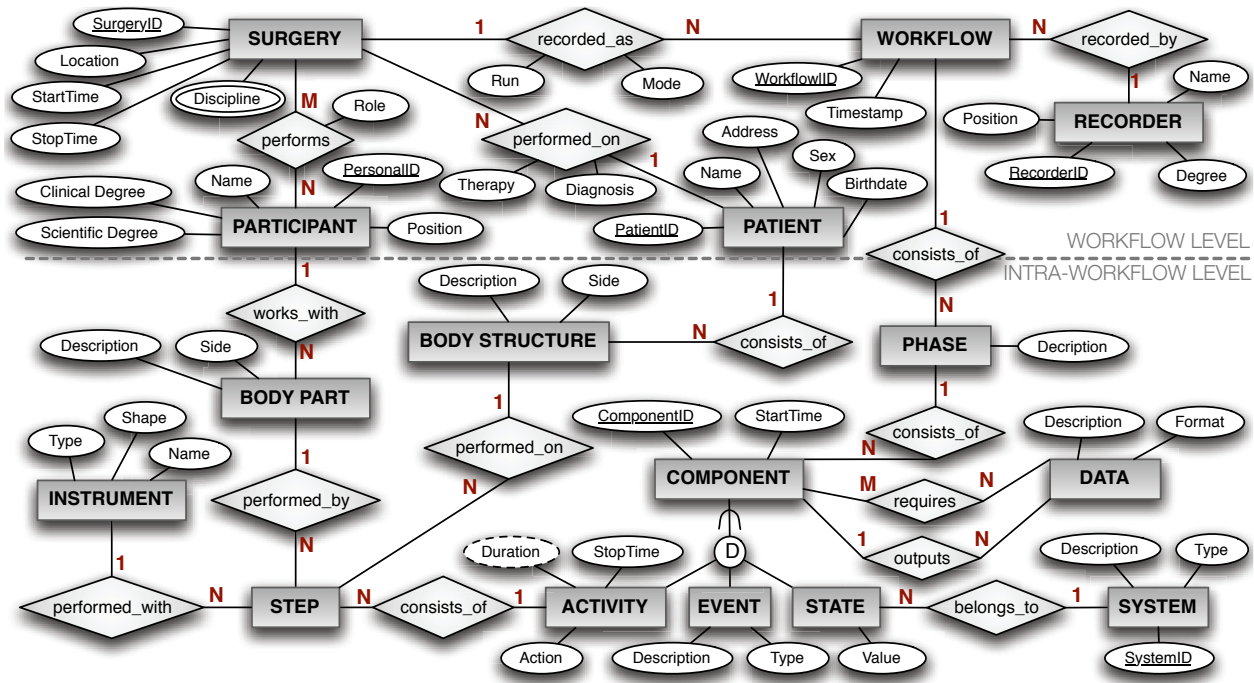


Figure 2. Recording scheme of the surgical workflow as E/R diagram.

specification.

A common approach to structuring a process to decompose it vertically, for instance with regard to the time line, into logical units, such as subprocess, stages, step, etc. From the logical point of view, a surgical process consists of phases, which, in their turn, consists of activities, such as work steps with the goal of performing a certain action. From the surgeon's perspective, an activity is an atomic unit; meaning that it cannot be further decomposed into smaller units. Technically, however, one single action may be executed by multiple agents, using multiple instruments. To account for this observation, we refine the granularity to the instrument usage level. On this level, each usage case refers to a part of the action performed by a single body part of an agent on a single structure of a patient using a single instrument. In the upward direction, surgical instances can be grouped into classes according to their diagnosis, which in turn are associated with surgical disciplines.

The conceptual design of a data warehouse is concerned with establishing relevant relationships of the data to be stored and analyzed in an implementation independent fashion. According to Kimball [10], the design process undergoes the stages of 1) choosing a business process, 2) choosing the grain of the process, 3) identifying the dimensional characteristics, and 4) defining the measured facts. The last step preaggregates the data to the desired level of detail, which, as a consequence, leads to storing predefined measurements of the process performance rather than "raw"

process data. In our application, however, it is necessary to store the original surgical workflow without aggregation. Quantitative queries represent just a fraction of the surgical workflow analysis. Other analytical tasks might be concerned with qualitative issues, such as pattern recognition, relevance assessment, or knowledge discovery.

To support a broad scope of expected queries, we opt for an *event tracking* data warehouse design. Event tracking tables are used for storing a robust set of many-to-many relationships between entities as so called *factless facts* [10]. As preparation phase to the identification of relevant dimension, we modeled the entire surgical workflow recording scheme using extended entity-relationship (E/R) notation, shown in Figure 2. Notice that we distinguish between the surgery itself and its recording in form of a workflow. The original decomposition into activities has been extended to include components of type *Event* and *State*. Both concepts come from the area of workflow design and are used to capture the surgery course's transition points (e.g., a surgeon's directive to his assistant) and the states of various systems (e.g., usage of a device), respectively.

Notice that the entire recording scheme in Figure 2 can be divided into two basis granularity levels:

1. *Workflow level* in the upper part of the diagram contains the characteristics describing the surgery and its workflow as a whole. Examples of such characteristics are *Patient* and *Participant*, and *Recorder*.

2. *Intra-workflow level* contains properties belonging to particular components – activities, event, and states – within a workflow, such as `Instrument` and `System`.

3 Conceptual Design

A multidimensional schema can be derived from the original E/R model by examining relationship cardinalities and functional dependencies between attributes. Guidelines for designing correct data warehouse schemes based on multidimensional normal forms can be found in [11].

Entity X may be mapped to a dimension or a next-level dimension category of entity Y , if there exists a one-to-many relationship between Y and X . In the multidimensional data model, such relationships are referred to as “rolls-up-to”. For example, `Body Structure` and `Patient` qualify to be treated as a dimension hierarchy of `Step`. Entities involved solely into incoming one-to-many relationships should be mapped to facts. Besides, each many-to-many mapping is a fact [10].

As a graphical notation for the conceptual modeling we adopt the Dimensional Fact Model (DFM) [6]. Figure 3 shows the resulting multidimensional schema of the surgical workflow recording. Facts are shown as boxes and dimension hierarchies are directed graphs of ‘circular’ category nodes. Superclass entity `Component` has been resolved into its subclasses (modeling details of `Event` and `State` have been omitted in Figure 3 for space reasons). The bottom-up placement of the elements in the scheme is partially ordered according to the decreasing granularity level.

Entities and attributes identified as dimensions have been arranged to classification hierarchies, whenever applicable. Each dimension is topped by an abstract category $\top_{dimension}$ with the only value *all* as the root for that dimension’s data tree. Property attributes, i.e. those not used for aggregation, are omitted to avoid presentation overload. Shared elements, such as entire dimensions or their parts, are displayed in a non-redundant fashion. The same category node may even be used to define parallel hierarchies. For example, `Time` plays the role of `Start Time` and `Stop Time` of `Surgery`. In such cases, we label the respective edges with the role names (not a part of the DFM). Explicit modeling of shared fragments results in a complex connected schema known as *galaxy*, or *fact constellation*. The advantages of this scheme are semantic richness, visibility of all valid aggregation and join paths. Besides, at the physical design stage, such fragments can also be created and maintained without redundancy. For example, `time` category is mapped to a single table, referenced by five different dimensions.

The original DFM does not distinguish between multi-

ple alternative and parallel hierarchies: both types are represented by multiple outgoing “rolls-up-to” relationships. However, this distinction is crucial for specifying correct aggregation behavior. Therefore, to model alternative, i.e., mutually exclusive roll-ups, we introduce an additional construct: alternative “rolls-up-to” arrows are merged into one line with merging point marked by a black circle, as used to model the roll-up of `Date` by `Month`, `Week`, and `Weekday`.

OLAP queries follow a common scheme: a measure of interest is aggregated along various dimensions. A measure is a formula consisting of a simple (e.g., sum) or a sophisticated (e.g., rank or ratio) function with one or more data attributes as its input. Since the fact tables in the surgical workflow data warehouse are *factless*, i.e., measure attributes are not captured by the model, it is indispensable to use OLAP tools that support dynamic measure definition from any attribute of the data cube. Fortunately, advanced front ends for visual analysis implement this feature.

Consider a sample query “Find the most frequently used instrument of each surgeon”. The grain of `Instrument` usage is `Activity_Step`, therefore, the measure of this query is the number of activity steps, aggregated by surgeon and instrument. The query results should be sorted by surgeon, and subsequently descending by the measure, so that the instrument with the highest number of usages appears at the top of each surgeon’s result set.

4 Demonstration

Data acquisition for the analysis is performed using a graphical workflow editor interface, implemented at ICCAS [17]. A trained and experienced observer generates the workflow data instantaneously by monitoring the running surgical intervention. The originally obtained flow often contains errors, inconsistencies and gaps due to time pressure, limited visibility or audibility, human or software failures. Therefore, prior to transferring the workflow to the data warehouse, the data is placed into temporary storage for verification, correction, and completion.

Once the “cleansed” data has been written to the data warehouse, it can be analyzed and explored using advanced OLAP tools. Standard interface for querying multidimensional aggregates is a *pivot table*, or *cross-tab*. In table 1 an application case of analyzing the use of instruments in the intervention type of discectomy is shown. The surgical goal of this intervention type is the partial removal of the herniated intervertebral disc. For the example we focus on querying the usage of different surgical instruments with the same surgical objective. During discectomies the parts of the vertebra are removed by the help of different surgical instruments: surgical mallet/chisel, surgical punch, and trephine. All instruments are used for ablating bone mate-

medical researchers and engineers, operate on the consistent set, get insight into data and obtain measurements and indicators relevant for solving their tasks.

References

- [1] O. Burgert, T. Neumuth, M. Gessat, S. Jacobs, and H. U. Lemke. Deriving dicom surgical extensions from surgical workflows. In *SPIE MI 2007: Proc. of SPIE Medical Imaging 2007 - PACS and Imaging Informatics*, page CID 61450A, 2007.
- [2] O. Burgert, T. Neumuth, F. Lempp, R. Mudunuri, J. Meixensberger, G. Strauß, A. Dietz, P. Jannin, and H. U. Lemke. Linking top-level ontologies and surgical workflows. *International Journal of Computer Assisted Radiology and Surgery*, 1(1):437–438, 2006.
- [3] E. F. Codd, S. B. Codd, and C. T. Salley. Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate. Technical report, E.F.Codd & Associates, 1993.
- [4] T. H. Davenport and J. E. Short. The new industrial engineering: Information technology and business process redesign. *Sloan Management review*, 31(4):11–27, 1990.
- [5] U. Dayal, M. Hsu, and R. Ladin. Business process coordination: State of the art, trends, and open issues. In *VLDB 2001: Proceedings of the 27th International Conference on Very Large Data Bases*, pages 3–13, 2001.
- [6] M. Golfarelli, D. Maio, and S. Rizzi. The dimensional fact model: A conceptual model for data warehouses. *International Journal of Cooperative Information Systems*, 7(2-3):215–247, 1998.
- [7] D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, and M.-C. Shan. Business process intelligence. *Computers in Industry*, 53(3):321–343, April 2004.
- [8] W. H. Inmon. *Building the Data Warehouse*. John Wiley & Sons, Inc., New York, NY, USA, 1992.
- [9] P. Jannin, M. Raimbault, X. Morandi, L. Riffaud, and B. Gibaud. Model of surgical procedures for multimodal image-guided neurosurgery. *Comput Aided Surg*, 8(2):98–106, 2003.
- [10] R. Kimball. *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. John Wiley & Sons, Inc., New York, NY, USA, 1996.
- [11] J. Lechtenbrger and G. Vossen. Multidimensional normal forms for data warehouse design. *Information Systems*, 28(5):415–434, 2003.
- [12] H. Lemke and M. Vannier. The operating room and the need for an IT infrastructure and standards. *Int J Comput Assist Radiol Surg*, 1(3):117–122, 2006.
- [13] C. L. MacKenzie, I. A.J., C. G. L. Cao, and L. A. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. *Minimally Invasive Therapy Allied Technologies*, 10(3):121–128, 2001.
- [14] S. Mansmann, T. Neumuth, and M. H. Scholl. Multidimensional Data Modeling for Business Process Analysis. In *ER 2007: Proceedings of the 26th International Conference on Conceptual Modeling*, 2007. to appear.
- [15] S. Mansmann, T. Neumuth, and M. H. Scholl. OLAP Technology for Business Process Intelligence: Challenges and Solutions. In *DaWaK'07: Proceedings of the 9th International Conference on Data Warehousing and Knowledge Discovery*, pages 111–122, 2007.
- [16] P. Muth, D. Wodtke, J. W. enfels, G. Weikum, and A. Kotz-Dittrich. Enterprise-wide workflow management based on state and activity charts. In *Proc. NATO Advanced Study Institute on Workflow Management Systems and Interoperability*, pages 281–303, 1997.
- [17] T. Neumuth, N. Durstewitz, M. Fischer, G. Strauss, A. Dietz, J. Meixensberger, P. Jannin, K. Cleary, H. U. Lemke, and O. Burgert. Structured recording of intraoperative surgical workflows. In *SPIE: Proc. of the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 54–65, 2006.
- [18] T. Neumuth, S. Schumann, G. Strauß, P. Jannin, J. Meixensberger, A. Dietz, H. U. Lemke, and O. Burgert. Visualization options for surgical workflows. *International Journal of Computer Assisted Radiology and Surgery*, 1(1):438–440, 2006.
- [19] T. Neumuth, G. Strauß, J. Meixensberger, H. U. Lemke, and O. Burgert. Acquisition of process descriptions from surgical interventions. In *DEXA 2006: Proc. 17th Int. Conf. on Database and Expert Systems Applications*, pages 602–611, 2006.
- [20] T. Neumuth, C. Trantakis, F. Eckhardt, M. Dengl, J. Meixensberger, and O. Burgert. Supporting the analysis of intervention courses with surgical process models on the example of fourteen microsurgical lumbar discectomies. *International Journal of Computer Assisted Radiology and Surgery*, 2(sup. 1):436–438, 2007.
- [21] N. Padoy, M. Horn, H. Feußner, M.-O. Berger, and N. Navab. Recovery of surgical workflow: a model-based approach. In *CARS 2007: Proceedings of the 21st International Congress and Exhibition on Computer Assisted Radiology and Surgery*, 2007.
- [22] T. B. Pedersen and C. S. Jensen. Multidimensional database technology. *IEEE Computer*, 34(12):40–46, 2001.
- [23] T. B. Pedersen, C. S. Jensen, and C. E. Dyreson. A foundation for capturing and querying complex multidimensional data. *Information Systems*, 26(5):383–423, 2001.
- [24] J. Qi, Z. Jiang, G. Zhang, R. Miao, and Q. Su. A surgical management information system driven by workflow. In *SOLI '06: Proc. of IEEE Int. Conf. on Service Operations and Logistics, and Informatics*, pages 1014–1018, June 2006.
- [25] H. A. Reijers. *Design and Control of Workflow Processes: Business Process Management for the Service Industry*. PhD thesis, Technische Universiteit Eindhoven, 2003.
- [26] A. Sharp and P. McDermott. *Workflow Modeling: Tools for Process Improvement and Application Development*. Artech House, Inc., Norwood, MA, USA, 2001.
- [27] W. M. P. van der Aalst and K. van Hee. *Workflow Management: Models, Methods, and Systems (Cooperative Information Systems)*. The MIT Press, Cambridge, 2002.