journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Development of a clinical data warehouse from an intensive care clinical information system

Marleen de Mul[a,*], *Peter Alons*[b], *Peter van der Velde*[c], *Ilse Konings*[c], *Jan Bakker*[d], *Jan Hazelzet*[e]

[a] *Erasmus University, Institute of Health Policy and Management, PO Box 1738, 3000 DR Rotterdam, The Netherlands*
[b] *Atos Origin Nederland, Enterprise Solutions/BI-CRM, Papendorpseweg 93, 3528 BJ Utrecht, The Netherlands*
[c] *Erasmus University Medical Center, Department of Information Technology, PO Box 2040, 3000 CA Rotterdam, The Netherlands*
[d] *Erasmus University Medical Center, Department of Intensive Care, PO Box 2040, 3000 CA Rotterdam, The Netherlands*
[e] *Erasmus University Medical Center, Department of Pediatric Intensive Care/Department of Information Technology, PO Box 2040, 2000 CA Rotterdam, The Netherlands*

ARTICLE INFO

Article history:
Received 31 July 2009
Received in revised form
18 January 2010
Accepted 14 July 2010

Keywords:
Data warehouse
Intensive care
Medical records systems, computerized
Clinical information system

ABSTRACT

There are relatively few institutions that have developed clinical data warehouses, containing patient data from the point of care. Because of the various care practices, data types and definitions, and the perceived incompleteness of clinical information systems, the development of a clinical data warehouse is a challenge.

In order to deal with managerial and clinical information needs, as well as educational and research aims that are important in the setting of a university hospital, Erasmus Medical Center Rotterdam, The Netherlands, developed a data warehouse incrementally. In this paper we report on the in-house development of an integral part of the data warehouse specifically for the intensive care units (ICU-DWH). It was modeled using Atos Origin Metadata Frame method. The paper describes the methodology, the development process and the content of the ICU-DWH, and discusses the need for (clinical) data warehouses in intensive care.

1. Introduction

Information technology in health care is still a topical subject, and reports like IOM's *Crossing the Quality Chasm* have stimulated developments in physician order entry, decision support systems and shared patient records [1]. Despite all the efforts, many health care organizations still have stand-alone information systems that do not communicate with each other. Therefore it remains an enormous task to use data collected at the point of care or in the supporting administrative processes for other purposes than they were collected for (e.g. management information, quality assessment and research). More importantly, clinical information systems (CIS) such as electronic patient records, are often designed to support hands-on care for individual patients, but are not well suited for analyses on an aggregated level, for example on groups of patients with the same disease.

Data warehousing is one of the techniques that seems promising for healthcare information systems. In its simplest definition, a data warehouse is a copy of transaction data specifically structured and optimized for query and analysis [2]. Use of data warehouses in healthcare is not new. The first articles on data warehousing date from two decades ago. In

* *Corresponding author*. Tel.: +31 10 40 81 86 8; fax: +31 10 40 89 09 2.
E-mail addresses: demul@bmg.eur.nl (M. de Mul), peter.alons@atosorigin.com (P. Alons), J.vandervelde@erasmusmc.nl (P. van der Velde), i.l.m.konings@erasmusmc.nl (I. Konings), Jan.bakker@erasmusmc.nl (J. Bakker), J.a.hazelzet@erasmusmc.nl (J. Hazelzet).
0169-2607/$ – see front matter
doi:10.1016/j.cmpb.2010.07.002

the period 1995–2000, it was a topic in healthcare management journals, but many articles provided only viewpoints, and lacked technical and empirical data [e.g. 3,4]. Certainly back then, the focus was on financial data warehouses, and data warehouses are still used for research into in-hospital costs [e.g. 5–7]. Since then, the focus has shifted somewhat to data warehouses for the bioinformatics domain [e.g. 8]. To date, there are but a few published examples of *clinical* data warehouses, using data from CIS, that are implemented and in use [9–13]. Because of the various care practices, data types and definitions as well as perceived incompleteness of clinical information systems, the development of a clinical data warehouse is a challenge [14].

Intensive care units (ICUs) might benefit from clinical data warehouses, as they can be characterized as information-rich environments with a high degree of automation and information technology. The last decades many ICUs especially in the larger university and teaching hospitals have implemented clinical information systems (CIS), sometimes called patient data management systems (PDMS) in critical care. These systems sample and store data from the monitors and other bedside devices, as well as manually entered observations. CIS are used for charting, fluid balance, medication lists and care planning. Some CIS also have physician order entry functionalities, or connections to the hospital information system (e.g. the laboratory system). Since these systems are designed for the point of care, the aggregation of data for management and research is limited. Querying the database of a CIS requires technical skills and knowledge of the database structure, competences which most managers, doctors, and nurses lack. Moreover, querying can be a burden for the operational database because it slows down performance [14].

There is still little evidence in the literature of the presence and use of data warehouses containing ICU data. Brammen et al. [15] provided the only example of a data warehouse that uses ICU data as a source for research on the interface of intensive care and genetics. The paper does not delve into the technical details of the development, but it is clear that this data warehouse has a specific and narrow focus that fits current research interests but is not prepared for future needs. Other studies on data warehousing in intensive care use a combination of a data warehouse (containing only administrative or financial data) and a separate query on the ICU's CIS [e.g. 16–19]. From this we conclude that data warehousing in intensive care is emerging, but not yet documented very well. Moreover, there is little experience with data warehouses containing clinical data about ICU patients, to be used for quality monitoring and research.

In this paper we report on the in-house development of an ICU-part of the data warehouse at Erasmus Medical Center, Rotterdam, The Netherlands (ICU-DWH). This part of the data warehouse collects data from the CIS installed at the ICUs, as well as data from the hospital information system. It was modeled using Atos Origin Metadata Frame method. The paper starts with a short analysis of the characteristics of intensive care, to understand the information needs of ICU staff and management. After that we describe the Metadata Frame method and its application in this project. We also present some examples of the way the ICUs currently use the data warehouse. In Section 5.1 we discuss the advantages of the Metadata Frame method, and present our lessons learned.

2. Background

Every clinical department is, or at least should be, interested in quality information and other issues requiring the analysis of patient data. However, intensive care units have some characteristics that make clinical data warehouses even more interesting and relevant. A first characteristic is the high availability of and dependency on technology and data. The devices on the ICU support vital functions of patients, while producing a lot of information about the patient: heart frequency, blood pressure, lab results, fluid intake, medication, etc. Producing a 'snapshot' of the status of an individual patient requires expert knowledge, but generating valuable information about groups of patients is even more complex. Many data elements have to be combined, which is difficult in a CIS.

Secondly, the cost of and access to intensive care are important (management) issues. The ICU is the most expensive ward in a hospital, ICU beds are scarce, and should be allocated to those patients who benefit from an ICU admission. Both to make a good admission and discharge policy and to monitor its application, data about the patient *population* of the ICU is crucial, yet not available in a CIS.

Related to cost and scarcity, a third characteristic of intensive care is the obligation to be accountable for the quality of care. In many countries, for example, standardized mortality rates (SMR) and other quality measures are benchmarked between ICUs and reported to health inspectorates. Many quality measures and prediction models have been developed for ICUs, and because of the drive to measure and improve quality ICUs are continuously in need of performance data. Quality measures such as "number of ventilator associated pneumonia infections per 1000 ventilator days" can be derived from the data in a CIS, but this requires a complex query. Likewise, severity of illness models (e.g. SOFA[1] [20]) or mortality predication models (e.g. APACHE[2] IV [21]) require complex queries if they are not available in a CIS.

A fourth and final characteristic that is relevant for data warehouse developments is the emergence of intensive care medicine as an independent discipline. Although the first ICUs date from the 1940s, it was not until the 21st century that intensive care medicine was regarded as a medical specialism. Knowledge on the best treatment of critically ill patients is still growing, and CIS can be a powerful data source for scientific research, especially for university hospitals.

Because of all these characteristics of intensive care, the need to analyze ICU data from different perspectives and on different levels (individual patient, patient group, department) is self-evident. Therefore a data warehouse combined with easy-to-use querying and reporting tools would be a big step forward. However, the CIS used in intensive care units usually have very complex data models. There is a lot of technical knowledge needed for making the right queries, while the

1 Sepsis-related Organ Failure Assessment.
2 Acute Physiology And Chronic Health Evaluation.

interpretation of the data requires clinical knowledge and an understanding of the context in which the data was recorded. Re-using patient data from CIS without data warehouses is, thus, a challenge in many ways.

2.1. *Data warehouse developments at Erasmus MC*

Erasmus Medical Center, Rotterdam, is a university hospital with 1200 beds and 37,000 admissions per year. Divided over three locations there are three ICUs with 104 beds in total (34 adult, 36 pediatric, and 36 neonatal). The ICUs have implemented their clinical information system, Critical Care Manager/CareSuite (Picis, Wakefield, MA, USA) in the period 2000–2006 [22]. The CIS database contains 75 GB of data and grows 15 GB each year.

The hospital has experience with data warehousing since 2000, and in several increments the scope of the overall data warehouse has broadened. However, the parts of the data warehouse were developed independently and separately from each other, using different methodologies. First, a financial part of the data warehouse was developed, which contained data on costs, production, personnel and absence through sick leave. From 2002, a separate data warehouse for the operating rooms was developed externally. Since 2004 a DRG-based part was added to the financial data warehouse, and a data warehouse for patient logistics was developed by staff of the IT department, supported by external parties. In this process some care was taken to avoid “stove pipe” solutions – separate data warehouse parts that cannot communicate with each other. All these data warehouse parts used (administrative) data from modules of the Hospital Information System (HIS) (iSoft, Leiden, The Netherlands). Erasmus MC uses SAP® BusinessObjects software (SAP AG, Walldorf, Germany) for generating reports from the data warehouse.

In 2005 the focus of Erasmus MC was expanded as they started the in-house development of a data warehouse part that was to contain patient information from clinical information systems and parts of the hospital information system (e.g. the laboratory module of the HIS). The intensive care data warehouse (ICU-DWH) was the first clinical increment of the data warehouse; the project started in September 2005, and was supported by Atos Origin.

3. Design considerations

For modeling ICU-DWH the Atos Origin Metadata Frame was chosen as method. This method fully supports the design principles of multi-dimensional modeling, as advocated by Kimball and Ross [2]. The Metadata Frame method is based on *fact* and *communication* oriented thinking and uses FCO-IM (Fully Communication Oriented Information Modeling) – the most modern form of complete communication oriented information modeling. The method is relatively new; it was developed in The Netherlands in the 1990s [23]. The Erasmus MC data warehouse is the first large scale application of the Metadata Frame method in health care. This method was chosen for several reasons. The first reason was the focus of this method, which places end-users like doctors, nurses, and researchers, as experts of the domain, in a central position. The second reason is the efficiency of the method, because the creation process of the required dimensional models is fully automated. The third, and most important reason was that the method also supports the maintenance of metadata. Because of the large amount of tables and data elements in a (clinical) data warehouse, technical and functional maintenance is a precarious matter. Extensions and changes in the models have to be implemented integrally throughout the data warehouse structure in order to maintain internal consistency and so to avoid stove pipes. Because of the several increments of the data warehouse, this was an important issue for Erasmus MC.

4. Description of method

For an extensive description of the Metadata Frame method and FCO-IM, we refer to [2,23,24]. In this section show the application of this method for the development of ICU-DWH. The Metadata Frame method distinguishes three phases: the preparatory phase, the data modeling phase, and the building phase. Its aim is to provide all information necessary for the builders who have to fill the database. However, the method should be integrated in a regular data warehouse development process, which does not end with an information model, but also includes the filling of the database, testing, and implementation. These last phases will be described here as well.

4.1. *Phase 1. Preparing*

The aim of the preparatory phase is to make a picture of the total user domain of the ICU-DWH. At Erasmus MC, a multidisciplinary project team was installed, consisting of two metadata experts, one data warehouse developer, two domain experts (one ICU doctor and one researcher), and three experts on the CIS (both technical and functional), of which two were also ICU nurses. They met regularly during all the phases, to provide the DWH developers with the necessary input for their work.

At the start, the DWH project leader and the Atos Origin consultant interviewed key users of the CIS: doctors, nurses, and (clinical) researchers, and the head of one of the ICU departments ($n=7$). These interviewees had experience with analysis of the CIS database, and therefore were potential users of the data warehouse. The interviews focused on current use of the CIS database, and wishes for future use of the data warehouse. Analysis of the interviews resulted in an Excel spreadsheet with five lists:

- a list of the organizational units involved – such as the pediatric and adult ICU;
- a list of ICU processes – such as medication, ventilation, scoring, diagnostics (in total, 20 processes were defined);
- a list of required and desired reports – such as a report for the mandatory national registries for ICU patients, and ad hoc reports for quality monitoring, research, or financial topics;
- a list of all performance indicators and key-performance indicators – such as length of stay, re-admissions within 48 h, transfers from and to other departments, number of central venous line infections, central venous line days,

positive blood samples, number of ventilator associated pneumonias, ventilator days and hours, mortality, blood transfusions, medications, and scores like APACHE IV (in total, 146 (K)PIs were defined);
- a list of candidate dimensions – such as date, time, patient, department, specialism, specialist, treating physician, diagnosis, device type, medical staff, medication (in total, 43 dimensions were defined).

In the same Excel spreadsheet a 'bus matrix' was created to visualize the relation between these five lists. Again, the project team members were involved in this process. For example, the (K)PIs were further defined, grouped and classified by assigning them to the various processes, the reports that Erasmus MC needed, and the candidate dimensions.

4.2. Phase 2. Modeling

In the second phase of the ICU-DWH development, examples were made for each of the items in the bus matrix and for the types of data available in the CIS (starting from the charts, screens and flow sheets used by ICU staff). These examples, which were formally verbalized with the domain experts are called 'fact expressions' in FCO-IM. Fact expressions are complete sentences in common language. Again, a spreadsheet was used to present the examples and share them in the project team. The fact expressions were presented together with the charts of the CIS containing the original data. This enabled the domain experts in the project team to understand the examples, correct verbalizations (if necessary) and validate the facts expressed in the sentences. Special care was given to ensure that the examples and their verbalizations contained all pieces of information at their 'natural lowest grain' (e.g. single events, nursing activities, and applied medications) and not just arbitrary aggregations of these facts. By these activities, it became clear which of the items defined in the first phase could indeed be made available in the ICU-DWH and which ones not, because they required (additional) recording in the CIS first. The team decided to model all items, whether available in the CIS database or not, in order to meet future needs.

When all the sentences were validated by the domain experts, CaseTalk™ software (BCP Software, Utrecht, The Netherlands) was used to derive a complete information model from these facts, with a minimum number of tables. The software uses an exact algorithm, dictated by 'predicate theory' (the mathematical theory behind FCO-IM) and this process is fully automated. An example of the fact expressions and their transition in CaseTalk™ is presented below.

4.3. Example

In the interviews, ICU staff expressed the need for real-time patient data (e.g. blood pressure, heart frequency, temperature) in the data warehouse, to be able to compare patient groups, to study trends through time, etcetera. This data, which is generated by the bedside monitors, can be found on the patient chart in the CIS (Fig. 1).

Fact expressions on the real-time (RT)-measures *Heart Frequency* and *Temperature* were defined as examples for the project team to validate:

- For patient 1234567 on 14-03-2008, 5:00:00 h for heart frequency a value of 98 beats per minute was recorded.
- For patient 1234567 on 14-03-2008, 6:00:00 h for temperature a value of 37.5 °C was recorded.
- ...

In addition, the domain experts in the project team agreed that for each real-time measurement only one result can be recorded in the CIS database. In CaseTalk™, the first example (fact expression) is 'grammatically' analyzed as is shown in Fig. 2.

As a result of the analysis of all fact expressions about real-time measurements, the table real-time measurement is generated (Fig. 3). This table has six attribute fields (the original label types in Fig. 2), of which 'Date', 'Patient_number', 'RT_variablename', and 'Time' are primary-key fields.

4.4. Phase 3. Building

When CaseTalk™ applied its algorithm, an optimal normal form "logical data model" was produced for the Corporate Fact Base. This data model was then transported to an Entity-Relationship (ER)-environment with the ER-Bridge, a tool originally developed by HAN University. This tool assures that all information that is relevant to the model (particularly the semantics underlying the model) is available in the ER repository as comments on the attributes. The meaning of every field in the model can therefore still be understood in terms of the original fact expressions.

At this point, all phases of the Metadata Frame method were completed. The ER model was implemented in the database using Oracle Designer. For the ETL (extract, transform and load) process, the Extelligence® Critical Care Export Tool was used, which was an extraction tool developed by the CIS vendor, Picis. This had two advantages. First, it saved a lot of developing time, because the vendor had the knowledge of the CIS database necessary for making complicated extractions and calculations relatively easy. Second, the extraction tool will be updated with every new release of the CIS, which guarantees future data quality and usability. Still, the ETL was very complex because the CIS data model and the DWH model were completely different. Data elements that were not in the extraction tool, as well as the data elements from the HIS, were extracted by Erasmus MC. Oracle Warehouse Builder was used to fill the tables of the data warehouse with the extracted data.

In this phase, some new attributes had to be defined, because they had not been mentioned by the interviewees, and thus they had not been modeled.

4.5. Phase 4. Testing

In this phase the data warehouse was tested. The testing phase involved two activities: the technical tests, where it is tested that the ETL-processes that transport data from the sources to the DWH does extract, transform and

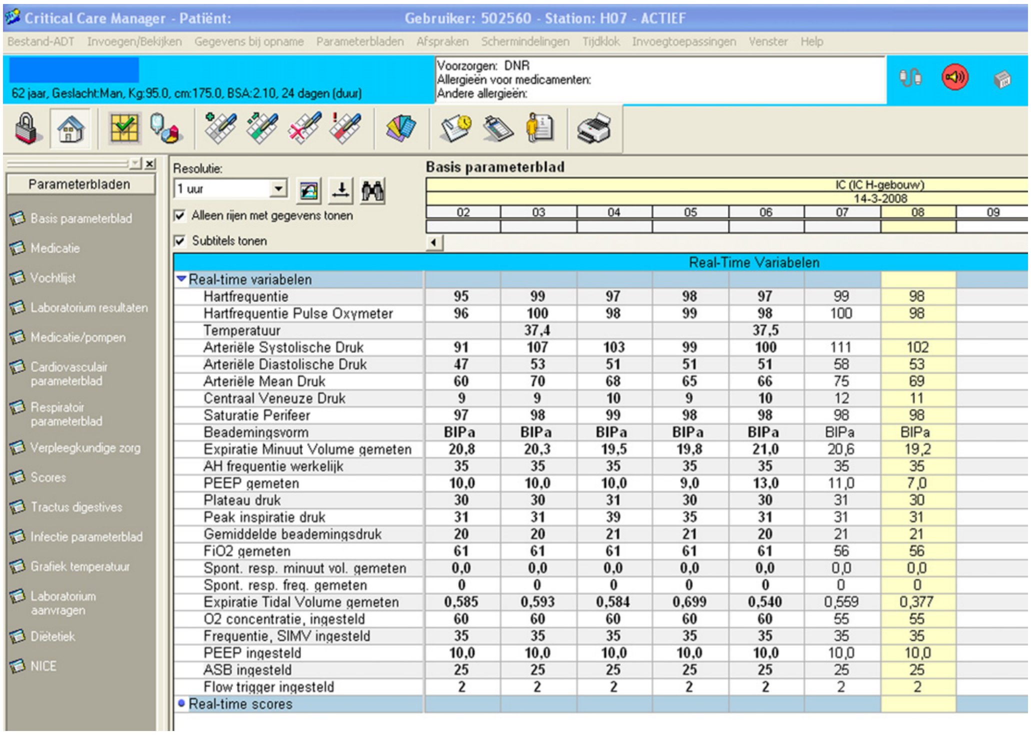

Fig. 1 – Patient chart in Critical Care Manager/CareSuite.

load these data in a correct fashion; and the user tests, which tests whether the anticipated reporting does what it is expected to do. Of course, the latter tests can never be considered to be complete to any full extent, as in a DWH-environment new reporting needs emerge continuously. The test group consisted of staff from the Department of Information Technology, the Division of Medical Information, and 12 end-users (managers, researchers, doctors and nurses).

The tests revealed only a few modeling errors. For example, in one table, the primary key was not defined correctly. Because the ER model was generated automatically, as a perfect translation of the fact expressions, it was to be expected that only these types of errors would be detected in the test phase. Still, the test phase lasted almost 2 years. This was partly due to the limited availability of the end-users, but more importantly to problems with the source system (the CIS and the HIS), that had to be solved first. Moreover, some errors in the ETL were detected.

Another problem was related to the metadata and the hospital information system. In all parts of the Erasmus MC data warehouse, the location of the patients (and thus an admission in a specific department) was derived from the HIS. The times of admission and discharge, however, did not match

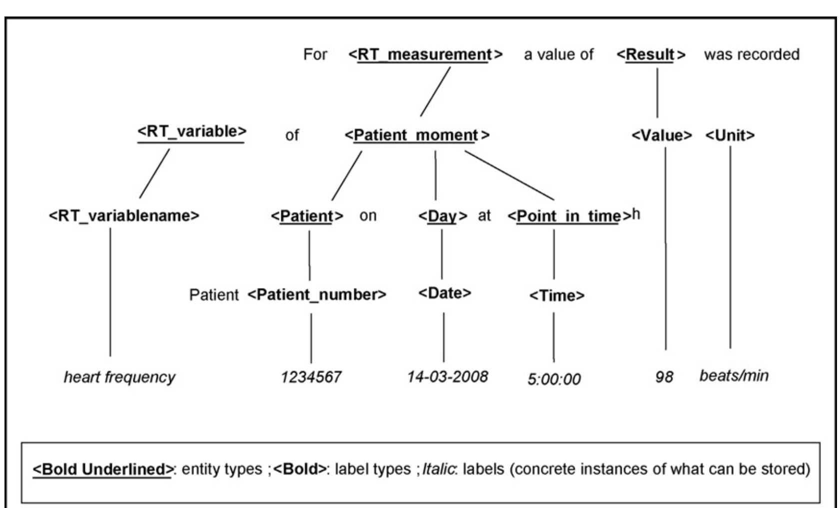

Fig. 2 – Expression tree in CaseTalk™ for the fact type real-time measurement with result.

Table Real_time_measurement

RT_variablename (RT_variablename) NN *Primary Key*	Patient_number (Patient_number) NN *Primary Key*	Date (Date) NN *Primary Key*	Time (Time) NN *Primary Key*	Value (Value) NN	Units (Units) NN
Heart frequency	*1234567*	*03-14-2008*	*5:00:00*	*98*	*beats / min*
Heart frequency	*1234567*	*03-14-2008*	*10:00:00*	*113*	*beats / min.*
Temperature	*1234567*	*03-14-2008*	*6:00:00*	*37.5*	*ºC*
Temperature	*1234567*	*03-14-2008*	*7:00:00*	*38.2*	*ºC*

FACT TYPE EXPRESSIONS:
F1: "For the <RT_variablename> of Patient <Patient_number> on <Date> at <Time> hours a value of <Value> <Units> was recorded."

FOREIGN KEYS:
Real_time_measurement (RT_variabelenaam) → RT_variabele (RT_variablename)
Real_time_measurement (Patient_number) → Patient (Patient_number)
Real_time_measurement (Date) → Day (Date)
Real_time_measurement (Time) → Point_in_time (Time)

Fig. 3 – ICU-DWH Table Real_time_measurement as emerging from the model creating algorithm of CaseTalk™.

the time recorded in the CIS of the ICU. This was problematic because many patients were admitted to the ICU, while they were not administratively discharged from a ward. The ICU-DWH only extracted data from the CIS within the boundaries of HIS-admission and HIS-discharge. While the first few hours of an ICU admission are important for data collection, a lot of these data were missing in the ICU-DWH.

Other errors were found by the users, when they tested the BusinessObjects reports. For example, when building a report on X-rays of ventilated patients, the users discovered errors in the calculation of length of ventilation. These errors were related to both modeling errors and other limitations in the CIS, and also to improper quality of manually entered data in the CIS.

4.6. *Phase 6. Implementation*

In this last phase, the ICU-DWH was officially put in use. From each ICU a few staff members were trained in using the SAP® BusinessObjects (BO) software. In addition, maintenance of the data warehouse was assigned to the Division of Medical Information. This division resides under the Department of Information Technology, and one of their tasks is to support hospital management, researchers, students and doctors with data. They have much experience with the SAP® BusinessObjects (BO) software.

5. Status report

The ICU-DWH consists of 24 dimensional star models, containing 49 different tables and 578 attributes. The data warehouse consists of 15 data marts (groups of one or more tables around a single ICU process): Allergies, Blood products, Catheters, Diagnoses, Events, Fluid balance, ICU-logistics, ICU-medication, Nursing activities, Observations, Precautions, Procedures, real-time measurements, Scores, and Ventilation. Central Fact tables and the approximate number of facts (data entries) are presented in Table 1.

Since the ICU-DWH was modeled on the lowest grain of data available in the clinical information system, the data can be used for (research) questions on various levels of detail; from the patient group or department level down to the individual patient level. Some examples of current use of the data warehouse are listed below:

1. A simple report is made for the secretaries of the pediatric ICU listing all children that were admitted and discharged in the previous week.
2. The ICU-DWH is used to calculate ventilation time and other quality measures Erasmus MC has to report to the Dutch healthcare Inspectorate.
3. The ICU-DWH is used for clinical research by identifying patients eligible for a study, by extracting the relevant patient data for this group, and by presenting this data in the desired format.

Table 1 – Fact tables of ICU-DWH and the number of facts they contain.

Fact table name	Number of facts (approximately)
Blood product administration	77,000
Catheter day fragment	197,000
Fluids per hour	10.0 million
ICU admission	24,900
Medication task	3.3 million
Medication administration per hour	7.9 million
Nursing activities	618,000
Observation item recording	6.5 million
Patient event	531,000
Real-time measurement	81.9 million
Scores item recording	1.6 million
Ventilation hour fragment	1.2 million

To date, three increments of the clinical part of the data warehouse have been completed: the intensive care increment (ICU-DWH), the radiology increment (RAD-DWH) and the laboratory increment (LAB-DWH). These three parts of the data warehouse were developed with the same methodology, which assured cross-referentiality. Alongside the clinical data warehouse development, Erasmus MC improved the non-clinical data warehouse parts in a way that allows linking to the clinical parts. This process is not finished yet, but in the future it should be possible to connect clinical patient data to, for example, DRG data.

5.1. *Discussion of the Metadata Frame method*

In the development of the ICU-DWH, the Metadata Frame method was used, which is based on Fully Communication Oriented Information Modeling (FCO-IM) and an automated conversion to Dimensional models as advocated by advocated by Kimball and Ross [2]. In the literature we found one hospital that used the same multi-dimensional modeling methodology to build a data warehouse for the whole organization, containing clinical, administrative and financial data [13].

Although the ICU-DWH is complex, its development process was transparent for the users (managers, doctors, researchers), because all elements of the ICU-DWH were defined using their own language. This was experienced as the main advantage of the methodology used (Metadata Frame, based on Fully Communication Oriented Information Modeling). Because an automated algorithm is used, The ER model is a perfect translation of all examples that were validated by the clinical experts.

However, a critical note to the methodology is the focus on the end products of the ICU-DWH: reports with fixed data elements, produced for a defined user group in the organization. This focus can be explained by the origin of the data warehouse as a managerial tool. However, the ICUs had mainly ad hoc research questions with a clinical focus. These questions were not known in advance, in phase 1 of the development process. Combined with the large amount of data and data types in the CIS, this introduced a risk for the data warehouse developers: how do we know if the information model is complete? Put differently, the automated algorithm can only guarantee a perfect data warehouse to a certain extent; it depends on the completeness and accuracy of phase 1 of the Metadata Frame method. During the testing phase of ICU-DWH we discovered that not all data elements had been modeled, because not everything was verbalized in a fact expression. Although only a few data elements were missing, it caused some delay in the project. On the other hand, since the method is fully automated where possible and fully aimed at delta-wise improvement of the implemented models, the necessary corrections were made very easily and with great precision.

The Metadata Frame adds tools for storage and maintenance of metadata. Because Erasmus MC (and many other hospitals) has a patchwork of information systems, it is crucial that definitions of (clinical) terms are stored unambiguously. For example, in the context of an ICU admission, length of stay might be defined with a precision of hours, starting with the physical admission of the patient and ending with his or her physical discharge. In the administrative part of the data warehouse, length of stay is defined on a day level, using the administrative admission and discharge dates. To avoid mistakes, these two "lengths of stay" should have distinct attribute names, and in the description of the attribute (the metadata) this distinction should be clear as well.

Another characteristic of the Metadata Frame method is its sensitivity to data quality issues in the source systems. ICU staff suspected errors in the data model of their CIS, which were confirmed by the Metadata Frame method. For example the medication table in the CIS, in which medication is classified by chemical group. However, one of these groups is "pumps", which is a way of administering medication. In this table, semantic rules are violated, leading to data base entries for medication that are not unique. Of course, the Metadata Frame method only reveals such inconsistencies, but it cannot solve the problem automatically. The only way to solve such issues in an automated fashion is by coupling data quality tools directly to the tools of the Metadata Frame method, where all business rules can be stored and made available for use by other tools.

6. Lessons learned

1. The preparation phase is crucial, but takes a lot of time. This was partly caused by the complexity of the clinical information system and the amount of wishes expressed by the domain experts, but also by the intensive user participation in the project team. Active user participation has proven its value in the past, when the clinical information system was implemented in the ICU [22], but there is also a risk. Compared to IT-driven, top-down projects in health care, projects like these, which place end-users in a central position, are highly dependent on input from doctors. Their work in the ICU, and the continuous availability to the clinic sometimes conflicted with the 'linear' structure of the DWH project management and planning of team meetings. For the benefit of continuity, this issue was solved by using a spreadsheet to exchange updates of the bus matrix in the project team. All members could add their input at the time that was most convenient, allowing the project to proceed without the regular attendance of clinical domain experts.
2. We learned that the scope of DWH projects should be broad from the start: not only focusing on the technical aspects, on the development of the data warehouse, but also on the services needed to optimize its use. For further organizational embeddings of the ICU-DWH (and related IT projects), Erasmus MC is preparing a Business Intelligence Center. This center aims to support the ICU-DWH users with – mainly – their knowledge on the On-Line Analytical Processing tools, while key users (e.g. research nurses) support their colleagues with their knowledge on ICU processes and CIS data. Thus, part of the functional maintenance is decentralized, with the key users being 'linking pins' to the Business Intelligence Center. In other hospitals, similar organizational solutions have been implemented [e.g. 26–28].
3. In addition to the previous lesson, it is crucial to train the users of the data warehouse in using the analytical and

reporting tools, and to provide them with the tools and support they need. Most doctors and researchers are familiar with a statistical software package, but not with On-Line Analytical Processing tools, such as SAP® BusinessObjects (BO). Moreover, in the complex ICU-DWH a query cannot be built through simple trial and error, but it has to stem from a clear and unambiguous (research) question. Often these research questions are presumptions from experienced ICU staff. For example "what is the mortality rate for postmenopausal female patients admitted from the general ward with fever?" While this question can be answered using data from the ICU-DWH, building the right BO report is very difficult for these types of questions. The BusinessObjects environment is not appropriate for answering all queries pertinent to given research issues. A report that appears valid, might still present misleading or wrong data. Moreover, the users of (performance) reports need to learn how to interpret the data because its format is different from the medical data presentations they are familiar with [25]. We propose that a course on ICU-DWH and BusinessObjects and extended reporting methods is offered a few times a year, for those nurses and (junior) doctors who want to do research in the ICU.

4. Developing clinical data warehouses places data quality high on the agenda. Streamlining data is challenging because definitions for individual items must be clear and unambiguous throughout the organization, while in practice shared data elements have alternative definitions, owing to a range of different (clinical and administrative) users with a variety of different information needs [14]. Thus, the data warehouse development raises new questions about system integration, definitions and data quality. Especially data warehouse parts that use manually entered patient data face quality problems regarding completeness, accuracy, timeliness, and so forth. The ICU-DWH is fully dependent on the data in the source systems. Health care staff is usually not aware that the data they enter in an electronic record is used for other purposes. However, if professionals learn what they can and cannot do with the data, they will probably be more motivated to improve their recording practices. That way, the DWH can be a catalyst for data quality improvement, and so for *information* quality improvement. This takes, however, continuous effort.

7. Future plans

Future plans include data warehouse parts to monitor clinical pathways, logistics of the outpatient clinic, pain in cancer patients, and the automatic generation of the Dutch quality measures for hospitals.

Epilogue

This project has won the Dutch Computable Award 2007 for best IT-innovation in healthcare.

Conflict of interest

Dr. P.W.F. Alons is a Business and Management Consultant of Atos Origin, and has been involved in the project as Information Architect.

REFERENCES

[1] Committee on Quality of Health Care in America, Crossing the Quality Chasm: A New Health System for the 21st Century, National Academy Press, Washington, 2001.
[2] R.R. Kimball, M. Ross, The Data Warehouse Lifecycle Toolkit, 2nd edition, Wiley, 2002.
[3] M.A. Geisler, D. Will, Implementing enterprise wide databases: a challenge that can be overcome, Topics in Health Information Management 19 (1998) 11–18.
[4] R. Scheese, Data warehousing as a healthcare business solution, Healthcare Financial Management 52 (1998) 56–59.
[5] A. Sacchetti, R.H. Harris, T. Warden, S. Roth, Contribution of ED admissions to inpatient hospital revenue, American Journal of Emergency Medicine 20 (2002) 30–31.
[6] R.J. Weber, S.L. Kane, V.A. Oriolo, M. Saul, S.J. Skledar, J.F. Dasta, Impact of intensive care unit (ICU) drug use on hospital costs: a descriptive analysis, with recommendations for optimizing ICU pharmacotherapy, Critical Care Medicine 31 (2003) S17–S24.
[7] F. Muranaga, I. Kumamoto, Y. Uto, Development of hospital data warehouse for cost analysis of DPC based on medical costs, Methods of Information in Medicine 46 (2007) 679–685.
[8] S.P. Shah, Y. Huang, T. Xu, M.M. Yuen, J. Ling, B.F. Ouellette, Atlas – a data warehouse for integrative bioinformatics, BMC Bioinformatics 6 (2005) 34.
[9] A. Grant, A. Moshyk, H. Diab, P. Caron, F. de Lorenzi, G. Bisson, L. Menard, R. Lefebvre, P. Gauthier, R. Grondin, M. Desautels, Integrating feedback from a clinical data warehouse into practice organization, International Journal of Medical Informatics 75 (2006) 232–239.
[10] B. Collins, M. Wagner, Early experiences in using computerized patient record data for monitoring charting compliance, supporting quality initiatives and assisting with accurate charging at Allina Hospitals & Clinics, International Journal of Medical Informatics 74 (2005) 917–925.
[11] S.J. Welch, S.S. Jones, T.T. Allen, Mapping the 24-hour emergency department cycle to improve patient flow, Joint Commission Journal on Quality and Patient Safety 33 (2007) 247–255.
[12] M.F. Wisniewski, P. Kieszkowski, B.M. Zagorski, W.E. Trick, M. Sommers, R.A. Weinstein, Development of a clinical data warehouse for hospital infection control, Journal of the American Medical Informatics Association: JAMIA 10 (2003) 454–462.
[13] C. Bréant, F. Borst, R. Nkoulou, O. Irion, A. Geissbuhler, Closing the loop: bringing decision support clinical data at the clinician desktop, Medinfo 12 (2007) 890–894.
[14] G.W. Gray, Challenges of building clinical data analysis solutions, Journal of Critical Care 19 (2004) 264–270.
[15] D. Brammen, C. Katzer, R. Röhrig, K. Weismüller, M. Maier, H. Hossain, T. Menges, G. Hempelmann, T. Chakraborty, An integrated data-warehouse-concept for clinical and biological information, Studies in Health Technology and Informatics 116 (2005) 9–14.
[16] G.K. Nishi, R.H. Suh, M.T. Wilson, S.A. Cunneen, D.R. Margulies, M.M. Shabot, Analysis of causes and prevention of early readmission to surgical intensive care, The American Surgeon 69 (2003) 913–917.

[17] R.F. Alban, A.A. Nisim, J. Ho, G.K. Nishi, M.M. Shabot, Readmission to surgical intensive care increases severity-adjusted patient mortality, The Journal of Trauma 60 (2006) 1027–1031.
[18] C.L. Byington, K. Korgenski, J. Daly, K. Ampofo, A. Pavia, E.O. Mason, Impact of the pneumococcal conjugate vaccine on pneumococcal parapneumonic empyema, The Pediatric Infectious Disease Journal 25 (2006) 250–254.
[19] J.F. Dasta, S.L. Kane-Gill, A.J. Durtschi, D.S. Pathak, J.A. Kellum, Costs and outcomes of acute kidney injury (AKI) following cardiac surgery, Nephrology, Dialysis, Transplantation 23 (2008) 1970–1974.
[20] J.L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C.K. Reinhart, P.M. Suter, L.G. Thijs, On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine, The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure, Intensive Care Medicine 22 (1996) 707–710.
[21] J.E. Zimmerman, A.A. Kramer, D.S. McNair, F.M. Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients, Critical Care Medicine 34 (2006) 1297–1310.
[22] M. De Mul, M. Berg, J.A. Hazelzet, Clinical information systems: CareSuite from Picis, Journal of Critical Care 19 (2004) 208–214.
[23] G. Bakema, J.P. Zwart, H. van der Lek, Fully Communication Oriented Information Modeling (FCO-IM Consultancy, 2002) [cited January 15, 2010]. Available from: http://www.fco-im.nl/pdfFiles/FCO-IM%20book.pdf.
[24] P.W.F. Alons, The Metadata Frame in Action: Unconfusing the Babylonians, DB/M 19, issue 6–7, 2008.
[25] J. Schulman, D.J. Spiegelhalter, G. Parry, How to interpret your dot: decoding the message of clinical performance indicators, Journal of Perinatology 28 (2008) 588–596.
[26] J.G. Dewitt, P.M. Hampton, Development of a data warehouse at an academic health system: knowing a place for the first time, Academic Medicine 80 (2005) 1019–1025.
[27] C.R. Weir, B.L. Hicken, H.S. Rappaport, J.R. Nebeker, Crossing the quality chasm: the role of information technology departments, American Journal of Medical Quality 21 (2006) 382–393.
[28] C.M. Munsch, Managing the measurement: a model of data support in an integrated delivery system, Journal of Nursing Care Quality 16 (2002) 1–9.