# Data warehousing in the Danish healthcare sector

**Mathias Abitz Boysen[a], Julian Birkemose Nielsen[a]**

[a]*Department of Health Science and Technology, Medical Informatics, Aalborg University, Denmark*

## Introduction

In the Danish healthcare sector enormous amounts of data are stored as fragments in a large number of unintegrated IT-based information systems (IS). It is a very resource-intensive task, but it has been demonstrated that the existing data can be used to provide an information basis for strategic decisions and research. However, data are rarely used beyond individual patient care because of its contextual nature. Therefore, fragments of data are rarely meaningful. Without data integration it becomes problematic to interpret and use the data for secondary purposes. Consequently retrieval and aggregation are not fully realised which are some of the most prominent advantages of IS. The health care sector has begun to realize the potential of dimensional data warehousing (DDW) to query across several data sources, thus creating data sets for research. DDW's has been developed to extract and load data from many heterogeneous sources into a joint underlying model. The aim of this study was to expand the existing patient administrative data warehouse in the North Region of Denmark (RN) to demonstrate integration of two data sources creating a dataset with type 1 diabetic patients and their laboratory results.

## Materials and Methods

RN's DDW embracing patient administrative data is in production. In this study a copy was expanded with a model embracing laboratory data. The copy is solely in an experimental state.

On the basis of laboratory processes and data extracted from LABKAII, a dimensional model was designed through three steps. 1. Determining the granularity. The highest level of detail in the laboratory data and the business process behind generation of the laboratory data was used to determine the models granularity. 2. Choosing and designing dimensions. Dimensions were designed to be highly denormalized tables containing all descriptive information fitting the granularity of data in the fact table. Dimensions in RN's DDW and LABKAII had overlapping information, why these were used as conformed (shared) dimensions. 3. Designing the fact table. The fact table was designed to only contain aggregable data and foreign keys to dimensions.

The dimensional model was implemented through *extract, transform* and *load* (ETL). A test dataset was extracted from LABKAII through a series of stored SQL procedures from a Microsoft SQL Server 2008. The raw extracted data from LABKAII were transformed to fit the dimensional model: Data types were changed, table columns were sorted, distinct values were extracted and data were filtered to remove redundant historic data. All descriptive information for dimensions was stored in Master Data Services (MDS) entities on a Microsoft SQL Sever 2012. After transformation the dimensional model was loaded into the DDW. The fact table has foreign keys to the dimensions; hence the dimensions were loaded before the fact-table by combining relevant MDS entities.

## Results

The final dimensional model consisted of one fact table, two new dimensions and four conformed dimensions. The fact table contained approx. 95 million rows. The granularity in the fact table corresponds to one laboratory result per row. Descriptive data unique to LABKAII were loaded in two new dimensions. The first contained information related to analysis types, reference limits and instruments. The second contained descriptions of priority, status and archetypes related to each result in the fact table. The dimensional model also relied on four conformed dimensions encompassing information about date, time, citizens, and organization. A query combined laboratory results with patient administrative data belonging to patients with diabetes mellitus type 1. The query yielded 4.949 patients with 2.612.219 laboratory results from 1.539 types of analyses.

## Discussion

The time is right for DDW's, because the Danish healthcare sector does a tremendous job classifying and structuring data in selected sources, making this DDW project manageable. It is important to realize that the DDW does not actually integrate data. It only loads the data into a joint model that can be queried across different sources easily. The secondary purpose decides the integration and it should be considered carefully with each new integrating query.

In conclusion the DDW can be used to retrieve large datasets for research and other secondary purposes. The more data sources it embraces the more secondary purposes it can support. However, concrete examples analysing datasets are needed to show the value and increase the acceptance.

## Acknowledgments

We would like to thank Enversion for support during development. We would also like to thank The North Region of Denmark for access to anonymous data.

## Address for correspondence

mab@enversion.dk