# Calculating Key Figures for Radiology Departments Using a Clinical Data Warehouse - A Technical Case Report

Leon LIMAN[a,1], Georg FETTE[b], Jonathan KREBS[a], and Frank PUPPE[a]

[a] Chair of Computer Science VI, Würzburg University, Würzburg, Germany
[b] Service Centre Medical Informatics, University Hospital of Würzburg, Würzburg, Germany

**Abstract.** Optimizing the utilization of radiology departments is one of the primary objectives for many hospitals. To support this, a solution has been developed, which at first transforms the export of different Radiological Information Systems (RIS) into the data format of a clinical data warehouse (CDW). Additional features, like for example the time between the creation of a radiologic request and the finalization of the diagnosis for the created images, can then be defined using a simple interface and are calculated and saved in the CDW as well. Finally, the query language of the CDW can be used to create custom reports with all the RIS data including the calculated features and export them into the standard formats Excel and CSV. The solution has been successfully tested with data from two German hospitals.

**Keywords.** Data Warehouse, Electronic Health Records, Radiology, statistics and numerical data

## 1. Introduction

### 1.1. Background

Receiving and understanding radiologic images of different kinds like e.g., the ones generated by Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) is very important for making a lot of clinical diagnosis. Therefore, optimizing the usage of the limited resources of radiology departments and by doing so improving the treatment of patients is of great interest for many hospitals. A possible area for optimizations is the reduction of the time a patient has to wait for a radiology appointment. This could for example be achieved by reducing the number of unnecessary repeated examinations of the same organ region with different (or the same) modality. Another advantage of such a reduction would be less radiation exposure for patients.

---

[1] Corresponding author: Leon Liman, Chair of Computer Science VI, Würzburg University, Am Hubland, 97074 Würzburg, E-mail: leon.liman@uni-wuerzburg.de.

*1.2. Objectives and Requirements*

The goal of the project is to provide a solution to calculate key figures for the utilization of the radiology department of different hospitals. Therefore, the developed system must be able to read the exported data of a Radiology Information System (RIS). This export should then be imported into an existing clinical data warehouse (CDW). An easy-to-use configuration file should be provided to specify the mapping of the export to the CDW's data structure. Furthermore, it should be possible to calculate additional features on the RIS export during the import. The calculations to be run on the data should then be configurable via a graphical user interface. Finally, the ability should be provided to save the calculation results in the standard formats Excel and CSV.

## 2. State of the art

The interest in providing key figures derived from the radiology data of a hospital [1] as well as various concrete use cases, that are most relevant to experts in the radiology domain [2] have already been described. And while data warehouse solutions in the field of radiology already exist, some of them [3] only offer a fixed set of predefined calculations on the data instead of the ability to define own queries on the data, or they provide fixed interfaces to the information systems of specific hospitals [3, 4] and therefore can't be easily adjusted to new data. More generic tools for importing different kinds of data into a CDW [5, 6] don't offer the option to calculate additional features during the import process and are therefore dependent on a more complex query language to perform advanced types of evaluations.

## 3. Concept

PaDaWaN [7] has been used as a CDW which stores its data in an entity attribute value (EAV) model [8] using either a MSSQL[2] or a MySQL[3] database. For increased performance of executed queries, the data is additionally stored in a Solr[4] index. Another advantage of using a Solr index is the ability to run complex queries on textual data including the option of ad hoc information extraction [9]. PaDaWaN already provides a web-based graphical user interface for configuring the calculations as well as the ability to export these results to Excel and CSV and thus directly meets these two requirements of the project.

PaDaWaN has been developed primarily together with and for the University Hospital of Würzburg, where it is in use as a CDW for the data of nearly all clinical departments and is used on many kinds of request by multiple clinicians.

The data model of PaDaWaN consists of two main tables. One of them is called the catalog and stores all attributes (as well as an optional hierarchy among these attributes) used in the data. The other one contains facts. Each fact has a relation to a single attribute from the catalog and saves a value for this attribute together with the value's time of measurement. Each fact also contains three different identifiers to be mapped to a patient,

---

[2] https://www.microsoft.com/en-us/sql-server/

[3] https://www.mysql.com

[4] https://solr.apache.org

a patient's case (in this project used for an appointment of the patient in the radiology department) and a clinical document created during a case (in this project used for a single radiological examination).

The RIS data was provided as multiple Excel sheets. Each of them contained information from one year and one hospital. Every row in the sheets includes the data from a single radiological examination. Their columns could be mapped to the following attributes: patient identifier (which is already pseudonymized in the RIS export), start and end (each specified by day and time) of the examination, modality (especially radiography, CT, MRI, ultrasound and angiography), device identifier, name of the examination (e.g. thorax in two levels), type of treatment (inpatient or outpatient), identifier and name of the hospital department requesting the examination, day and time of the request, day and time of creation and approval of the examination results, type of the radiology department (neuroradiology or radiology). A sample with some columns of this export is shown in Figure 1. The needed information in this export was specified in detail based on an initial concept intended to enable different kinds of evaluations of radiology departments. This enabled the IT department responsible for the RIS to directly provide the export (except for some information that is not available in the used RIS).

| | A | B | C | D | E | F | |
|---|---|---|---|---|---|---|---|
| 1 | PatID | Durchführungsdatum | Durchführungszeit | Durchführungsendzeit | Leistung / Gew. Unt. | CODE Arbeitsplatz | Durchg. Untersuchu |
| 2 | 1 | 12.01.2019 | 12:00:00 | 13:00:00 | MRT | MRT1 | Prostata nativ + KM |
| 3 | 1 | 14.07.2019 | 09:38:56 | 09:48:56 | Konsil | KON | Einspielen von CT-Fr |
| 4 | 1 | 21.12.2019 | 13:45:19 | 13:55:19 | Sonstiges | ARCH | Erstellen einer CD (R |
| 5 | 1 | 31.12.2019 | 09:08:49 | 09:11:07 | Röntgen | ROE1 | Thorax in 2 Ebenen |
| 6 | 2 | 17.03.2019 | 14:43:56 | 14:47:37 | CT | CT5 | Thorax nativ |
| 7 | 2 | 19.03.2019 | 18:46:05 | 18:48:05 | Röntgen | ROE1 | Thorax in 1 Ebene a |
| 8 | 2 | 20.03.2019 | 09:22:33 | 09:24:33 | Röntgen | ROE3 | Thorax in 1 Ebene a |
| 9 | 2 | 21.03.2019 | 14:54:20 | 14:56:20 | Röntgen | ROE5 | Thorax in 1 Ebene a |
| 10 | 3 | 08.10.2019 | 13:30:00 | 14:45:00 | MRT | MRT2 | Herz (Stress) nativ + |
| 11 | 4 | 04.05.2019 | 14:26:17 | 14:26:17 | Röntgen | ROE2 | Schultergelenk links |
| 12 | 5 | 01.04.2019 | 11:59:29 | 12:14:29 | Nuklearmedizin | CT4 | Diagnostische CT zur |
| 13 | 6 | 03.11.2019 | 13:35:28 | 13:37:28 | Röntgen | POLI1 | Finger V links in 2 Eb |
| 14 | 6 | 03.11.2019 | 13:37:28 | 13:39:28 | Röntgen | POLI1 | Finger IV links in 2 E |
| 15 | 6 | 03.11.2019 | 13:39:28 | 13:42:28 | Röntgen | POLI1 | LWS in 2 Ebenen (R/ |
| 16 | 6 | 03.11.2019 | 14:31:30 | 14:51:30 | US | US5 | gesamtes Abdomen |
| 17 | 7 | 01.08.2019 | 10:38:02 | 11:08:02 | MG | MG | Drittbefundung ohn |
| 18 | 8 | 06.01.2019 | 16:10:07 | 16:20:07 | Röntgen | ACR | Thorax in 2 Ebenen |
| 19 | 8 | 06.01.2019 | 20:21:42 | 20:36:42 | CT | ACT | Schädel nativ (NRAD |
| 20 | 9 | 21.07.2019 | 18:51:42 | 18:56:34 | CT | CT5 | Traumaspirale mehr |
| 21 | 9 | 21.07.2019 | 18:56:44 | 19:11:44 | CT | CT5 | HWS nativ (RAD) |

**Figure 1.** Sample of some of the data in the used RIS Excel export; the columns have German titles and contain (from left to right) the pseudonymized patient identifier, the day of the examination, the start and end time of the examination, the modality, the device identifier, and the name of the examination (the shown data has been altered compared to the real data)

This RIS export is then imported into the PaDaWaN with the following steps (the details of these steps are explained in section 4.1). At first a single RIS exported Excel sheet as well as the configuration file are loaded. Then all the raw data from the export is saved into the database according to the configuration file. Next, the calculation of additional features from the raw data is done, that are used for each examination, as well as additional features per patient and appointment taking into account multiple examinations. These additional features are saved into the database as soon as they have been calculated. Finally, all the database content is saved into a Solr index.

For querying the content of the Solr index the PaDaWaN provides its own "Medical XML Query Language" (MXQL)[5]. The basic building block of this query language is the "Attribute". It represents a single entry from the catalog table and can specify different filters for the found values depending on the type of the catalog entry (e.g., that a number should be lower than a specified bound or that a text should contain some words). Multiple attributes can then be combined using the logical connections "AND" and "OR". These logical connections could then in turn be placed inside of the following XML tags: "DistributionColumn", "DistributionRow" and "DistributionFilter". As the names suggest these tags configure the columns and rows of the result table from the execution of the query. The "DistributionFilter" is used to provide additional constraints that should be applied to all columns and rows. Finally, these three tags are combined under a "Query" tag, that specifies, which count should be returned (number of matching patients, appointments or examinations). An example for such a query is shown in Figure 2. Every possible combination of the specified columns and rows is then created and translated into a Solr query each, which is then executed. To create MXQL queries, the graphical user interface of the PaDaWaN could be used. It also provides the options to directly show the result table as well as to download the result either as an Excel or as a CSV file.

```xml
<Query>
    <DistributionRow>
        <Attribute domain="leistungen_generated"
        extID="modalities_order_per_exectask_abdomen" desiredContent="Nachfolger"/>
    </DistributionRow>
    <DistributionColumn>
        <Attribute domain="leistungen_generated" extID="hospitallocation"
        contentOperator="EQUALS" desiredContent="X"/>
        <Attribute domain="leistungen_generated" extID="hospitallocation"
        contentOperator="EQUALS" desiredContent="Y"/>
    </DistributionColumn>
    <DistributionFilter>
        <Attribute domain="leistungen_generated" extID="datayear"
        contentOperator="EQUALS" desiredContent="2019"/>
    </DistributionFilter>
</Query>
```

**Figure 2.** MXQL query to calculate the number of different types of multiple examinations for abdomen examinations (as rows of the result) per hospital (as columns) filtered on only the examinations of 2019

To prevent the reidentification of patients via PaDaWaN queries (or at least to make it more difficult), PaDaWaN offers the option to only display, and export calculated numbers, if they are larger than a specified threshold. With this option it is not possible if e.g., a rare disease of a patient is known to retrieve more information about this patient.

---

[5] https://www.informatik.uni-wuerzburg.de/is/open-source-tools/padawan-data-query-tool/entwickler-sicht/abfragesprache/ (German)

## 4. Implementation

### 4.1. Solution description

To import the RIS export into the PaDaWaN a configuration file has been provided to easily specify the mapping to the data model of PaDaWaN. An example of such a file is shown in Figure 3. The configuration file uses Excel and contains the following 5 columns:

- **ColumnName** contains the name of the column from the RIS export to be configured. For timestamps the used RIS export has two columns (one for the date and another one for the time). For this kind of attribute, multiple names could be specified in this column of the configuration file.
- **DisplayTitle** specifies the name the corresponding PaDaWaN catalog table entry should get.
- **ExtID** has to be a unique identifier for the current row's attribute, that is used for example to map an MXQL Attribute tag to a catalog table entry.
- **DataType** provides the type of data contained in the corresponding column of the RIS export. Supported types are Text, Number, DateTime, SingleChoice (for attributes with a fixed set of values, e.g., the type of treatment only has the possible values inpatient and outpatient) and Structure (like SingleChoice for attributes with a fixed set of values but for this type for each possible value a separate catalog entry is created; this is used for attributes with a large number of possible values like e.g., the identifier of the device used for the examination).
- **ValueMappings** is an optional column that can be used to map values for attributes of type SingleChoice to a more readable value. Currently this is used for the type of radiology department to map "RA" to "Radiology" and "NR" to "Neuroradiology".

| | A | B | C | D | |
|---|---|---|---|---|---|
| 1 | ColumnName | DisplayTitle | ExtID | DataType | ValueMappings |
| 2 | PatID | Patienten-ID | pid | Text | |
| 3 | Durchführungsdatum;Dur | Durchführung - Start | execStart | DateTime | |
| 4 | Durchführungsdatum;Dur | Durchführung - Ende | execEnd | DateTime | |
| 5 | Leistung / Gew. Unt. | Leistung / Gewünschte | taskReq | SingleChoice | |
| 6 | CODE Arbeitsplatz | CODE Arbeitsplatz | codeStation | Structure | |
| 7 | Durchg. Untersuchung | Durchgeführte Unters | execTask | Structure | |
| 8 | CODE Abteilung | Fachabteilung | reqUnitShort | SingleChoice | RA:Radiologie;NR:Neurora |

**Figure 3.** Example for the configuration file with the mappings for the columns in Figure 1 (from top to bottom: patient identifier, start and end time of the examination, modality, device identifier, name of the examination) as well as the type of the radiology department (in the last row)

All the information necessary to create the mapping (especially the exact definitions of the columns in the RIS export) had been provided in a documentation created by the IT department responsible for the RIS. Therefore, the configuration file could be created without any further involvement of this IT department.

By using this configuration all the raw data from the export is loaded and saved into the database one row (meaning one examination) at a time. As the data contains no

identifier for the appointments, such an identifier is generated by grouping all examinations by the combination of a single stay in the hospital, the day of the year, the shift in which the examination ended and the modality of the examination.

After that, additional derived values are computed from the raw data and stored in the database. For this purpose, two different interfaces (coded in Kotlin) have been provided that compute derived values from a single row related to just one examination or from multiple rows related to a combination of several examinations. For example, the first interface is currently used to get the name of the shift in which the examination ended or to calculate time gaps (in minutes and days) between two DateTime attributes like e.g., the start and end of an examination. Another use case of this interface is the translation of identifiers for e.g., the different hospital departments to a readable name. This was possible based on a mapping between the identifiers and names provided as part of the documentation of the RIS export (compare columns A and B in Figure 3). The second interface spanning multiple examinations is used e.g., to check whether for one patient multiple examinations (with the same or with different modalities) are performed for the same organ region.

When the database has been filled, all its information is transferred to a Solr index, which could then be used to perform the calculations of interest via the graphical user interface of the PaDaWaN. An example for a query to calculate the number of different types of multiple examinations for abdomen examinations is shown below in Figure 4 with the result table presented in Figure 5 (downloaded in Excel format; other options are a preview in the UI or a download in CSV format). This example utilizes the ability of PaDaWaN to hierarchically structure the catalog and automatically generating rows for all the entries below another one by using the query attribute parameter "Successors of" (in German "Nachfolger von"). Therefore, not every possible type (as partially shown in the expanded catalog on the left in Figure 4) must be individually added as a row. The columns are configured explicitly to query two different hospitals (in German "Klinik"). All the possible combinations of columns and rows are filtered to only query data from 2019 (in German "Jahr der Ursprungsdaten").

## 4.2. System in Use

The described system has been developed and used with the RIS exports of two different hospitals from northern Germany containing the data described in section 3 for all radiology examinations from January 2019 to September 2020. In total the export contains data from nearly 490,000 examinations belonging to about 380,000 appointments and more than 125,000 patients. The whole system has been installed on a virtual machine with Ubuntu 20.04 running Microsoft SQL Server 2019 and Solr 8.8.1. The virtual machine got 4 processors and 24 GB of RAM and was installed on a Solid-State-Drive. Details about the process of importing and indexing are shown below (Table 1).

Depending on the complexity of the configured query, the time Solr needs for the calculation could be very different. As an example, the query from Figure 4 resulted in 207 rows and 2 columns (thus requiring 414 individual Solr queries) and took 10 seconds to complete.

Currently the system runs in the research network of Würzburg university. Once the improvements described in the next chapter are finished, the system will also be deployed directly in the network of the hospitals mentioned above. Furthermore, the usage of this system with RIS data in other hospitals is planned.
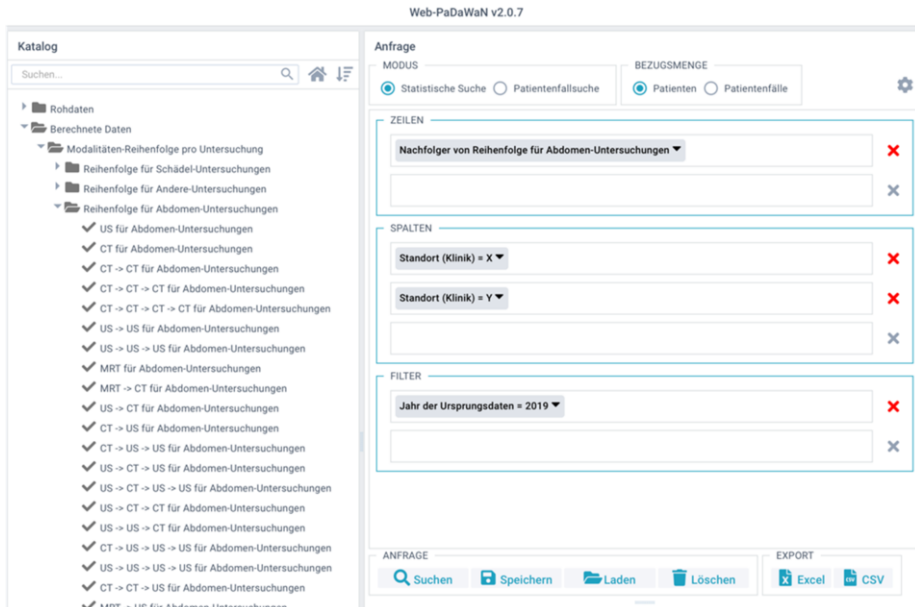
**Figure 4.** PaDaWaN-UI with a query to calculate the number of different types of multiple examinations for abdomen examinations (as rows of the result with some concrete types shown in the catalog on the left) per hospital (as columns) filtered on only the examinations of 2019; this query is identical to the one shown in MXQLformat in Figure 2



**Figure 5.** Example PaDaWaN Excel result with the number of patients for different types of multiple examinations for a single organ region and per hospital (each shown numbers has been randomly altered compared to the real results)

**Table 1.** Measurements for the import and index creation process

| | |
|---|---|
| Duration of database import | 1 h 19 minutes |
| Size of database | 11.35 GB |
| Duration of Solr index creation | 44 minutes |
| Size of Solr index | 9.8 GB |

## 5. Lessons learned

The core of the developed system is the clinical data warehouse PaDaWaN. CDWs in general provide many different advantages by making multiple tasks like e.g., the recruiting of patients for clinical studies or statistical analysis a lot easier. This has been one of the reasons, why the usage of CDWs has been strongly encouraged by the Medical Informatics Initiative Germany[6].

For the use case of computing key figures for radiology departments, the developed system provided many advantages. By using a simple configuration file, the import can be easily adjusted when additional data from the RIS should be added. Even the adoption to a completely different RIS is easily achievable by just adapting the configuration file. Therefore, the whole system could be used in many different hospitals. Furthermore, the PaDaWaN provides the ability to save configured queries and reuse them at any time and even on data sets from different hospitals. This makes it easy to change parts of the configured calculations to e.g., get the data for another year or another modality.

But such an approach also has some disadvantages. If additional features should be calculated during the import, these features have to be planned and implemented before they can be used for PaDaWaN queries. When additional features have been added or the calculation of existing features has been adapted, the import and index creation have to be run again, which (for the amount of data described in section 4.2) takes about 2 hours until the updates are available.

Some improvements for the described system are already planned. For commonly generated features on the raw RIS data (like time gaps) another configuration file will be provided, so that these features could be specified and generated during the import without the need to write any code. Additionally, the import process will be adapted, so that single features could be added to or updated in the database and the Solr index without the need for importing and indexing all the other features again. Furthermore, the PaDaWaN interface will gain the ability to adapt the layout of the generated Excel file (e.g., by specifying names for the generated rows and columns) and to show the numbers of the created tables as graphics.

Currently the clinicians using this system are provided with either the Excel results or with graphics manually created from these results. The operation of this system has only been done by the authors so far. Because the first reactions by clinicians to the generated results has been very positive, it is planned to deploy the system in a way, so that the clinicians can use it on their own. From a technical viewpoint, the system to retrieve the results consists of three parts: (1) PaDaWaN, (2) potentially complex PaDaWaN queries based on the imported RIS data and (3) graphical presentation of the results from (2). The first two components can already be used by clinicians by themselves with a small amount of training. For the third component, a comfortable tool for the creation of graphics from the Excel results is currently developed. And while so far, the results had been only calculated for the whole data of a past year, with this tool (and by changing the used PaDaWaN queries) it will be possible to calculate the graphics based on the data of up to the last day.

---

[6] https://www.medizininformatik-initiative.de/en/

# 6. Conclusion

A solution has been presented to use the Excel export of a RIS and transfer its data into the CDW PaDaWaN. For this purpose, a simple configuration file has been provided to specify the mapping of the columns of the RIS export to PaDaWaN catalog entries. Also, a Kotlin interface was developed to enable the calculation of additional features on the raw RIS data. After the PaDaWaN has been filled with this data its web-based graphical user interface could be used to easily configure multiple kinds of calculations on the data and export their results as either an Excel or a CSV file e.g., for further usage with data analytics tools.

# Declarations

# References

[1] V. Georgiana, D. Kartawiguna, Evaluation of radiology data warehouse implementation on education, research, and quality assurance, 2016 International Conference on Information Management and Technology (ICIMTech) (2016), 277-280.

[2] M. Karami, R. Safdari, From information management to information visualization: development of radiology dashboards, Applied clinical informatics 7(2) (2016), 308.

[3] P. G. Nagy, M. J. Warnock, M. Daly, et al., Informatics in radiology: automated Web-based graphical dashboard for radiology operational business intelligence, Radiographics 29(7) (2009), 1897-1906.

[4] L. M. Prevedello, K. P. Andriole, R. Hanson, et al., Business intelligence tools for radiology: creating a prototype model using open-source tools, Journal of digital imaging 23(2) (2010), 133-141.

[5] C. Bauer, T. Ganslandt, B.Baum, et al. The integrated data repository toolkit (IDRT): accelerating translational research infrastructures, Journal of Clinical Bioinformatics Vol. 5. No. 1 (2015), 1-2.

[6] G. Fette, M. Kaspar, G. Dietrich, et al., A Customizable Importer for the Clinical Data Warehouses PaDaWaN and I2B2, GMDS (2017), 90-94.

[7] G. Fette, M. Ertl, G. Dietrich, et al., An improved data workflow for a medical data warehouse, European Journal of Epidemiology 32 (2016), 54.

[8] V. Dinu, P. Nadkarni, Guidelines for the effective use of entity-attribute-value modeling for biomedical databases, International Journal of medical informatics 76 (11-12) (2007), 769-779.

[9] G. Dietrich, J. Krebs, G. Fette, et al., Ad hoc information extraction for clinical data warehouses, Methods of information in medicine 57 (Suppl 1) (2018), e22.