Clair-DB: A Standardized Departmental Research Database

Enno T van der Velde, Thekla Jansen, Victoria Delgado, Douwe Atsma

Leiden University Medical Center, Leiden, the Netherlands

Abstract

The objective of the CLAIR-DB (Cross Linked Advanced Integrated Research Database) project is to create one single, controlled research database for our department (Cardiology) from which all clinical research will be developed. The ProMISe system is used as the underlying database system. The approach: definition of a limited set of database tables; definition of common variables; all variables unambiguously defined (based on international standards and guidelines). All clinical data are automatically imported from our Cardiology Information System (EPD-Vision) into CLAIR-DB. Therefore, all clinical data are directly available from the source, there is no need for retyping data, data are always up-to-date. Research-specific data are entered directly in the CLAIR-DB web interface.

To perform statistical analysis, an export from CLAIR-DB (ProMISe) can be made in SPSS format.

Results. Data harmonization from 16 existing databases has resulted in one single well defined data set (CLAIR-DB). Existing research databases are now being transferred to Clair-DB.

Conclusion. The CLAIR-DB research database will help to facilitate and improve retrospective analysis of standardized clinically acquired data common to several research lines developed within a single department.

1. Introduction

The efficacy of implementation of a common standardized central database to be used by an investigational group for several lines of research remains unknown. The objective of the CLAIR-DB (Cross Linked Advanced Integrated Research Database) project is to create one single, controlled research database for our department (Cardiology) from which all clinical research will be carried out. Existing research databases are migrated into the new research database.

1.1. Background

Current research projects rely on accurate management of the databases, ensuring accurate extraction of data from Electronic Health Record Systems or collection of data from Clinical Research Forms. Post-processing and analysis of data from databases may result challenging, mainly if different types of software are used (e.g. Microsoft Excel or Access, SPSS, …). Therefore, having a common and unique database for all research projects would be ideal to ensure the integrity of the data.

Thus, the objective of the CLAIR-DB (Cross Linked Advanced Integrated Research DataBase) project is to migrate all existing stand-alone research databases in the department of Cardiology of the Leiden University Medical Center (LUMC) into one single and controlled environment. All research should then be carried out only on data extracted from the CLAIR-DB database. We expect that new and existing research projects will greatly benefit from cross-link between all available datasets.

2. Methods

The ProMISe system from the Leiden University Medical Center (LUMC) department Medical Statistics and Bioinformatics is used as the underlying database system [1]. The approach: definition of a limited set of database tables; definition of common variables; all variables unambiguously defined (based on international standards and guidelines).

2.1. The underlying database: Promise

The ProMISe (‘‘Project Manager Internet Server’’) system has been developed and is maintained at the department of Medical Statistics of the Leiden University Medical Center. ProMISe is a generic Internet based application for the design, maintenance and use of (clinical) data management projects, and is presently used by a number of other registries, such as the European Blood and Marrow Transplantation (EBMT) Group. Both the design and data handling for a project is integrated into the web application, allowing the designer to change any aspect of the project while in production. No additional software is required.

A large number of security measures have been taken to ensure the integrity of the data. The ProMISe server is located in one of the dedicated server rooms of the

LUMC, configured in a dedicated, demilitarized zone, and protected using 2 separate firewalls. The only access to the server that is allowed is through the https and ftp protocol (the latter only for static web pages). Users (or the designer or administrator of the project) can log on to their own project with a standard web browser (Microsoft Internet Explorer) over a secure connection from any location on the Internet, using individual username-password combinations. Extra authentication can be required using a TAN code SMS.The passwords that can be used must adhere to the specific Microsoft recommendations and are stored encrypted. All activities and modifications are logged on IP and user level.

In addition, external XML-formatted research data can be acquired in ProMISE employing secure techniques..

ProMISe currently runs a variety of national and international clinical studies and registrations in haematology, cardiology, orthopaedics, oncology, neurosurgery, paediatrics and neurology.

2.2. The uniform database model and data definitions

The basis for the definition of the CLAIR-DB database model is:

- definition of a limited set of database tables
- definition of a common set of variables

16 stand-alone research databases have been identified and have been selected to merge into CLAIR-DB. The 16 databases together resulted in a total of 2364 different variables, with a large number of variables shared by more than one research database. All variables were then allocated to one of the defined database tables (see below).

In order to build an uniformly structured database, the meaning of all variables should be unambiguously defined. For instance, we have decided on standard definitions of variables such as ‘hypertension’, of ‘diabetes’ and other. The definitions were based as much as possible on existing standards.

The CLAIR-DB database consists of the following (small) set of tables:

- patient table
- study table (information on which studies this patient is enrolled in)
- procedure table (info on procedures performed)
- imaging table (results from echo, X-ray, CT, MRI, …)
- assessment table (history, risk factors, ..)
- medication table
- complications table

All tables, except for the patient and study table, have as index-field the date of the procedure/observation.

2.3. Sources for the data

The data from the 16 stand-alone research databases are, one by one, imported into CLAIR-DB, after conversion (if necessary) to the definition of the uniformly defined variables. During the import, the data is checked with the corresponding information in our Cardiology Information System (EPD-Vision). If the information in EPD-Vision was more coherent, that data was used instead of the data from the research database.

After importing the data from each stand-alone research database, newer information from EPD-Vision was added if available to CLAIR-DB. Research-specific data not available in EPD-Vision is entered directly in CLAIR-DB using the ProMISe web interface. If it concerns variables that can be of importance for the clinical work, these are added to the EHR (EPD-Vision) and are transferred automatically from there to CLAIR-DB.

As soon as all data are imported and available in CLAIR-DB, the database is automatically updated and complemented from EPD-Vision. This means that new data are directly available from the source; no retyping of data; data are always up-to-date.

2.4. Patient privacy protection

Privacy legislation places great weight on personal autonomy, and requires informed consent for accessing personal health data for research. A legitimate public concern related to the use of personal health data is the risk of privacy breaches.

The right to data privacy is heavily regulated and actively enforced in Europe, including the collection of medical data in a system of personal identification. Many European countries have implemented national privacy laws; however more and more countries have now implemented the "EU Directive on the protection of personal data” [2;3].

De-identification of research data according to the specifications in the European Directive is advocated. In CLAIR-DB attributes with identifying information such as ‘Name’, ‘Phone number’, ‘Social Security Number’ will be omitted.

2.5. Database export for research

To allow statistical analysis with the data, an export from CLAIR-DB (ProMISe) is made in SPSS format. The export file can be limited to exactly those variables as defined in the original research project setup, or can

contain additional data to allow correlation with novel parameters. Data can only be exported and analysed from patients that have been included in the specific study.

3. Implementation status

The following stand-alone research databases are now being imported into CLAIR-DB:

- ICD research database
- CRT research database
- Mission! Myocardial Infarct research database
- AF (Atrial Fibrillation) research database

The other research database will follow thereafter.

4. Discussion and conclusion

The current practice of Healthcare generates, exchanges and stores huge amounts of patient-specific information in various Electronic Health Record (EHR) systems. This wealth of information can be exploited for care, statistics and clinical research, by connecting the EHR system with a clinical research database such as CLAIR-DB.

The CLAIR-DB system architecture certifies that the source of the data used for specific clinical research projects is well defined and can traced back at any later stage.

The use of a common set of variables ensures that variables such as ‘hypertension’ and ‘diabetes’ have the same meaning across studies, and also that the same measurement results are used across studies.

In conclusion, the CLAIR-DB research database may help to facilitate and improve retrospective analysis of standardized clinically acquired data common to several research lines developed within a single department.

References

[1] van der Velde ET, Vriend JW, Mannens MM, Uiterwaal CS, Brand R, Mulder BJ. CONCOR, an initiative towards a national registry and DNA-bank of patients with congenital heart disease in the Netherlands: rationale, design, and first results. Eur J Epidemiol 2005;20(6):549-57.

[2] Carinci F, Di Iorio CT, Ricciardi W, Klazinga N, Verschuuren M. Revision of the European Data Protection Directive: opportunity or threat for public health monitoring? Eur J Public Health 2011 Dec;21(6):684-5.

[3] European Commission. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Communities 2011 May 11;L 281/31.

Address for correspondence.

Enno T. van der Velde, PhD, FESC
Medical Physicist
Department of Cardiology, C5-P-32
Leiden University Medical Center (LUMC)
PO Box 9600, 2300 RC Leiden, The Netherlands
ETvanderVelde@lumc.nl