Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-554

A Web-based Tool for Patients Cohorts and Clinical Trials management

P. FRACCARO[a] and Mauro GIACOMINI[a,1]
[a] *Dept. of communication, computer and system sciences (DIST) University of Genoa*

Abstract. Clinical Trials (CTs) play an increasingly important role in modern medicine. Often these types of studies, especially in the final phases, require collaboration between several hospitals, laboratories and institutions in different places or countries. The solution proposed is a template which exploits the principles of Networked Clinical Research and the strengths of Clinical Data Management Systems (CDMSs) commonly used (Electronic Data Capture (EDC), Electronic Medical Record (EMR) and Electronic Health Record (EHR). Therefore, the basic structure is a highly normalized and standardized database which is managed by a web platform and, only by recording the required information and developing web pages starting from predefined templates, it is possible to carry out new projects. The result is a hybrid CDMS which preserves the flexibility and user autonomy of EDC systems; and contemporarily, permits the creation of patients cohorts, as in EMR and EHR systems, in which to realize simultaneously different multicentric CTs in several medical fields.

Keywords. Database, Clinical Trials, Electronic Data Capture, Electronic Medical Records, Patients Cohorts.

Introduction

Clinical Trials (CTs) play an increasingly important role in modern medicine and are an essential tool for the Evidence Based Medicine (EBM) approach [1, 2]. Clinical research generates a large amount of data that must be collected, processed and analyzed, often from centers which are in different places or countries. Consequently, Networked Clinical Research (NCR) and Clinical Data Management Systems (CDMSs) have become fundamental instruments for studies especially in the area of biomedicine and biotechnology. In general, trial data is collected at investigator sites with special forms, so called Case Report Forms (CRFs) which are then queried, cleaned, stored and analyzed by CDMSs. The collection of clinical data by means of electronic forms is called Electronic Data Capture (EDC) and the advantages of their use together with the internet for data management, is widely recognized[3, 4].

Recent surveys have shown a substantial heterogeneity in CDMSs in Europe, in fact there are many solutions but none of which are considered optimal [5-7]. In addition, in the academic community, where there is lack of human and financial resources, the open source solutions are widely employed like: (OpenClinica™ [8], OpenCDMS™ [9], PhOSCO™ [10] and REDCap [11]). This type of solution, although low cost and suitable

[1] Mauro.Giacomini@dist.unige.it

for a single trial, neglects two fundamental possibilities in clinical research: the interoperability with medical care and the patients cohorts creation. Unfortunately, nowadays, there are few solutions which fulfil these objectives. Important examples are certainly the NADIS (Fedialis Medica, Marly le Roi, France), which is an electronic medical record (EMR) for HIV-, hepatitis B virus (HBV)- or hepatitis C virus (HCV)-infected adults seeking care in French public hospitals [12], and Julius[13] which is an Electronic Health Record (EHR) adopted in Sweden. An EMR is a legal computerized medical record created in an organization that provides care. While an EHR represents the ability to easily share medical information among stakeholders and allows the following of a patient from different departments or institution[14]. The inserted patients in Nadis constitute a true cohort and entered clinical data remains available in order to carry out different studies contemporarily. On the other hand, Julius, through its template based structure founded on medical domain knowledge, allows the templates reusability and integration with other existing systems[13]. However, Nadis and Julius have some limitations: both are proprietary systems only available in some French hospitals within the French national health system and in Sweden; Nadis maintenance and development requires a large amount of human and financial resources and has an inflexible structure. Additionally, using these systems clinicians encounter restrictions, especially in data extraction. Another important technical aspect is the application structure: Julius exploits the World Wide Web (WWW), while Nadis it is a desktop application which uses a VPN connection and consequently requires more maintenance then a web-application. Undoubtedly their structures and features are highly useful for clinicians and they allow the collection of a vast amount of clinical data for CTs.

The solution proposed is a standard web-based template which permits the exploitation of the strengths of both the EDC, EHR and EMR systems. The result, which can be obtained by simply recording data and developing the web pages, is a hybrid CDMS which consists of a web-based platform that preserves the flexibility and user autonomy of the EDC solutions; while still allowing the creation of patient cohorts through which different CTs are simultaneously managed.

1. Methods

The positive aspects of the previously illustrated solutions were focused on to realize a database structure applicable in different medical fields, safe and correct data sharing among partners and user-friendly applications suitable for operators without an informatics background. After several interviews with experts in this field, the E-R diagram approach[15] was adopted and several conceptual schemas were developed, with different levels of detail, which expressed the information that was collected concerning clinical studies workflows. Subsequently, the E-R schemas were translated into logic diagrams and were then adapted to the specific features of the Data Base Management System (DBMS) of the Microsoft SQL Server 2008 system.

In order to manage any type of patient information a three level structure was established: Clinical Events; Categories; Parameters. Clinical Events were defined as all those happenings of medical interest which patients undergo during a clinical study (ie. a therapy, a blood sample or a hospitalization); and the Clinical Events were then categorized using a meta-description approach. Events attributes, named “Parameters”, are divided into “Categories” and conceptually in each category there is uniform data but in a different format (distributed using just a few tables). Using this approach a

standard structure was developed which is applicable to different types of trials in varying clinical fields, therefore the addition of new events, and the relative categories and parameters, is restricted to an insertion of records in the existing tables and it does not involve changes in the database organization. Furthermore, in order to constitute a cohort, all examinations entered for each patient had to be accessible from all the studies in which the patient participated; but related data had to be entered only once by clinicians therefore the patients' clinical records were separated from the event's temporal features in order to obtain this goal. Another fundamental characteristic of the solution is the data shareable feature amongst partners. In order to avoid legal and privacy problems a Local Identifier Code (LIC) is assigned to patients through a HASH code. In this way, only the authorized operators from the LIC, can access the patients' real identity. Therefore, a choice was made to widely utilize standards for sharing data within the scientific community. Accordingly, a specific structure was developed in order to archive clinical events, categories and parameters by using standards codes (LOINC and ICD). Moreover, a structured output in Health Level 7 (HL7) was created for sharing data with other work groups that can understand this standard interface.

Finally, the third main aspect of the solution is the usability by operators without an informatics background. First of all, the platform, that manages the database, was developed on the World Wide Web in order to limit the maintenance costs, page templates were developed referring to CRFs and the three level structure previously described. Therefore users, that utilize this type of application, have to access and evaluate data. Since records are stored in the database and operators do not have high computing skills, data cannot be autonomously extracted through DBMS by users, consequently an extraction tool was developed for the platform. This tool consists of two parts: the first allows users to insert the requirements and information criteria (such as thresholds or equality/inequality) that have to be extracted; the second, by interpreting the operators' demands, the queries are built dynamically, and then extracts data from the database making them available in Excel format.

2. Results

In order to obtain the database structure described above, several E-R diagrams were produced and then translated into logic schemas.

Referring to Figure 1 it is possible to observe the three level structure of the patients' clinical information. The different types of clinical events ("Eventi") are defined in "Tipi_Evento" and with its foreign key "Id_Tipo_Evento" in "Eventi" it is possible to refer back to the "happening" type. The events information is structured using categories ("Categorie"), which are sets of attributes ("Parametri" with the column "Codice_LOINC" for the HL7 standard interfaces). Moreover, this structure organization is related also to the study ("Categorie_Possibili_Tipo_Evento" with primary key "Id_Tipo_Evento", "Id_Categoria", "Id_Studio") therefore any type of clinical event could consider some categories in one clinical research and others in another. Furthermore the event times, in respect to the patient dates of reference for the studies, are stored in the table "Tn_Evento" whose primary key is indeed "Id_Evento","Id_Tn" and "Id_Studio". Accordingly, the event is seen with different Tn from different on-going studies but its records, organized around the table "Categoria_In_Evento", are entered only once, allowing substantial time-saving and decreasing the possibility of error. Another fundamental table is "Dipendenze_Eventi",

in which "Id_Evento" is the foreign key, which allows the recording of the relationship of dependency between clinical events. For example, in this way, is possible to manage information like a blood sample carried out under a particular therapy or an exam undergone during a hospitalization.

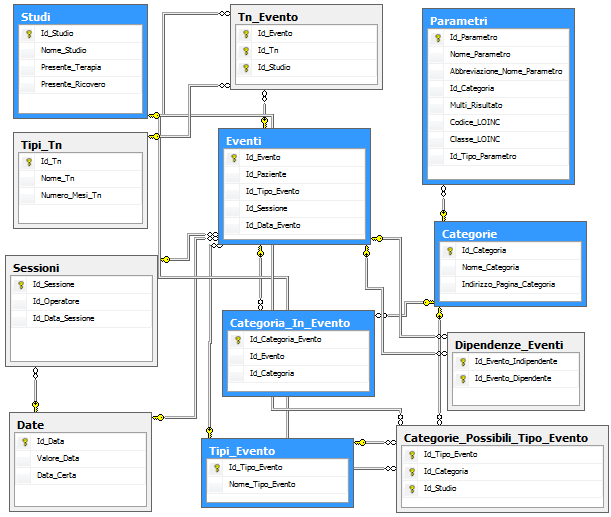

Figure 1: Logic Diagram Level 1.

In addition, a web application template was realized, which with only a few adjustments, permits the database management, referring to the principles above.

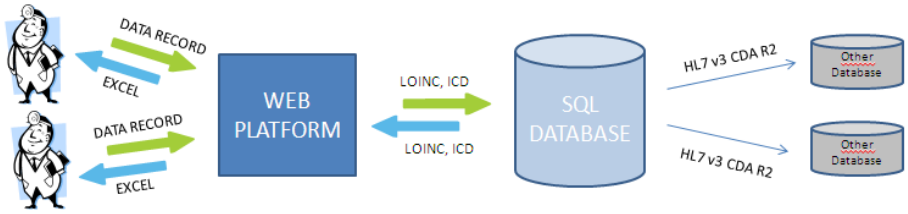

Figure 2: Solution architecture.

In Figure 2 the solution architecture is shown. Users enter data in the web application and information in standard format (LOINC, ICD) through the World Wide Web, which reaches the server and is stored in the SQL database. Data can be also consulted and extracted by the user utilizing the web application and the extraction tool previously described. Moreover, the system exports information in HL7 v3 (version 3) CDA R2 (Clinical Document Architecture, Release 2) in order to share the recorded data with other work groups which have the same standard interface.

At the moment, utilizing the above template and methods, a web-platform concerning Infectious Diseases was created that permits collaboration among eight different Departments of Infectious Diseases and two Laboratories in Liguria and Piedmont. From July 2011 up to March 2012, twenty operators requested the access credentials, four different clinical studies have been developed, and approximately 200

patients were recorded in the database; constituting a true patients cohort available for further research. These results, obtained in only a few months, evidence the effectiveness and flexibility of our method and solution.

3. Discussion

The methodological approach which was realized, allows the exploitation of the strengths of the EDC, EHR and EMR systems. Therefore, the implemented database structure and the architecture are flexible, general and standardized. These features allow, starting from this basis, the development of new projects in a very short time scale. In the future, using a standardized communication interface also for the importation, it will be possible to integrate the sets of data developed with this method with other databases directly on the web-platform. Moreover, with the same interface it will be possible to interact with Hospital Information Systems and Laboratories Information Systems in order to directly import patient information, thus avoiding data entry processes which are both time consuming and can cause errors. Finally, the aim is to develop a statistical area in order to permit assessments and analysis directly on the platform.

References

[1] ClinicalTrials.gov (*2012 January*)
[2] Davidoff F, Haynes B, Sackett D and Smith R. "Evidence based medicine". *BMJ*. 1995-04-29; 310(6987): 1085-1086.
[3] Kubick WR. "The elegant machine: applying technology to optimize clinical trials". *Drug Information Journal*. 1998; 32(4): 861-869.
[4] Paul J, Seib R and Prescott T. "The Internet and Clinical Trials: Background, Online Resources, Examples and Issues". *J Med Internet Res*. 2005; 7(1): e5.
[5] Ohmann et al. "Standard requirements for GCP-compliant data management in multinational clinical trials". *Trials*. 2011; 12(85-94.
[6] Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, Leizorovicz A, Schade-Brittinger C, Wittenberg M, Voko Z, Gaynor S, Cooney M, Doran P, Maggioni A, Lorimer A, Torres F, McPherson G, Charwill J, Hellstrom M and Lejeune S. "Heterogeneity prevails: the state of clinical trial data management in Europe - results of a survey of ECRIN centres". *Trials*. 11(1): 79-89.
[7] Ohmann C and Kuchinke W. "Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration.". *Methods Inf Med* 2009; 48(1): 45-54.
[8] http://www.openclinica.org/ (*2012 January*)
[9] http://www.opencdms.org/features (*2012 January*)
[10] http://www.phosco.com/ (*2012 January*)
[11] Harris P A, Taylor R, Thielke R, Payne J, Gonzalez N and Conde J G. "Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support". *Journal of Biomedical Informatics*. 42(2): 377-381.
[12] Pugliese P, Cuzin L, Cabié A, Poizot-Martin I, Allavena C, Duvivier C, El Guedj M, De La Tribonnière X, Valantin M. A, Dellamonica P and the Nadis Group. "A large French prospective cohort of HIV-infected patients: the Nadis Cohort". *HIV Medicine*. 2009; 10(8): 504-511.
[13] Rong Chen, Gosta Enberg and Gunnar Klein. "Julius - a template based supplementary electronic health record system". *BMC Medical Informatics and Decision Making*. 7(1): 10.
[14] Davis M Garets D. "EMRs and EHRs. Concepts as different as apples and oranges at least deserve separate names.". *Healthcare Informatics* 2005 October; 22(10): 53-56.
[15] Pin-Shan Chen P. "The entity-relationship model: toward a unified view of data". *ACM Trans. Database Syst.* 1976; 1(1): 9-36.