COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 114 (2014) 70–79

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

A simple versatile solution for collecting multidimensional clinical data based on the CakePHP web application framework☆

Martin Biermann[a,b,*]

[a] *Nuclear Medicine and PET Centre, Department of Radiology, Haukeland University Hospital, N-5021 Bergen, Norway*

[b] *Section for Radiology, Clinical Institute I, University of Bergen, Bergen, Norway*

ARTICLE INFO

Article history:
Received 30 April 2013
Received in revised form
10 December 2013
Accepted 8 January 2014

Keywords:
Web application framework
PHP
CakePHP
Clinical Data Management System
Investigator-initiated clinical research

ABSTRACT

Clinical trials aiming for regulatory approval of a therapeutic agent must be conducted according to Good Clinical Practice (GCP). Clinical Data Management Systems (CDMS) are specialized software solutions geared toward GCP-trials. They are however less suited for data management in small non-GCP research projects. For use in researcher-initiated non-GCP studies, we developed a client–server database application based on the public domain CakePHP framework.

The underlying MySQL database uses a simple data model based on only five data tables. The graphical user interface can be run in any web browser inside the hospital network. Data are validated upon entry. Data contained in external database systems can be imported interactively. Data are automatically anonymized on import, and the key lists identifying the subjects being logged to a restricted part of the database. Data analysis is performed by separate statistics and analysis software connecting to the database via a generic Open Database Connectivity (ODBC) interface. Since its first pilot implementation in 2011, the solution has been applied to seven different clinical research projects covering different clinical problems in different organ systems such as cancer of the thyroid and the prostate glands.

This paper shows how the adoption of a generic web application framework is a feasible, flexible, low-cost, and user-friendly way of managing multidimensional research data in researcher-initiated non-GCP clinical projects.

1. Introduction

Clinical studies in human medicine generate multidimensional data sets with numerous observations that are best administered using dedicated software solutions for data entry and analysis. At our molecular imaging center, we needed a flexible, scalable, and affordable solution for data management in our own researcher-initiated studies.

Clinical Data Management Systems (CDMS) are a family of client–server applications aimed at pharmaceutical trials [1]. Such trials are conducted for regulatory approval of

* *Correspondence to*: Centre for Nuclear Medicine and PET, Department of Radiology, Haukeland University Hospital, Jonas Liesvei, N-5021 Bergen, Norway. Tel.: +47 55977643; fax: +47 55977602.
E-mail addresses: martin.biermann@k1.uib.no, martin.biermann@helse-bergen.no

0169-2607/$ – see front matter © 2014 The Author. Published by Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2014.01.007

a drug or medical appliance by regulatory bodies such as the Federal Drug Agency (FDA) or the European Medicines Agency (EMEA). Design, conduct, and data management in such trials are governed by stringent international conventions such as Good Clinical Practice (GCP) [2] in addition to national legislation [3]. The design of such trials is invariably prospective, usually randomized, and, if possible, double-blinded, and outcome measures (such as total mortality or disease-related mortality) are set in advance [4]. Documentation must be tamper-proof [2] to avoid potential allegations of fraud as billions of dollars are at stake for the pharmaceutical company that developed the drug and sponsors the trial [5]. Independent contract research organizations (CRO) specialize in running trials in a GCP-compliant manner. These days, data entry will most often be conducted via electronic case report forms (eCRF) using CDMS with an internet portal [6].

Non-commercial, researcher-initiated studies will often follow less formal exploratory designs aimed at gaining new insights into a given problem. At our molecular imaging center, we combine hybrid imaging – single photon emission computed tomography (SPECT) and positron emission tomography (PET) both acquired in conjunction with computed tomography (CT) – with other radiological modalities such as ultrasound (US), magnetic resonance imaging (MRI), and US-guided biopsies both in our clinical routine and in our research projects. This yields complex data sets comparing several imaging modalities (such as US, PET, SPECT, and contrast-enhanced CT) with cytological (US-guided biopsy) and histological (after surgical treatment) verification in one or several tumor lesions in a large number of patients. Projects are often interdisciplinary, involving different clinical specialists (e.g. surgeons and oncologists), imaging specialists (nuclear medicine and/or radiology), and laboratory specialists (pathology, cytology, clinical chemistry) in the scope of a single research project such as multimodal imaging for thyroid cancer [7].

For use in our own non-GCP clinical research projects and based on earlier experience with a custom-designed data management system for a clinical trial [8,9], we were looking for a system that met the following specifications: (1) The system should be network-based, allowing for concurrent data entry by several authenticated users. (2) The system should meet all current regulatory requirements in respect to data protection and security. (3) The system should allow for hierarchical data models supporting complex entity relationships and provide built-in mechanisms to enforce relational integrity. (4) Modifications to the data models must be easy to implement even when data acquisition is under way. (5) The system should be cheap so that it can be shared between groups and projects without being limited by software licensing. (6) The software should be vendor-independent and multi-platform (e.g. Linux, Microsoft Windows®) so that it can be expected to be viable for the entire duration of projects spanning several years [9,10].

Finding no suitable software solution that met all our current requirements, we set out to develop a new simpler and more scalable solution for data management in our own clinical research projects.

2. Related work

Requirements for GCP-compliant CDMS have been reviewed in depth by Ohmann et al. [11]. An overview of available systems is provided by a recent European Survey [1]. At the 74 study centers, 39 different systems were in use in 2008/2009: 18 self-developed proprietary, 17 commercial, and 4 open source. The latter include the increasingly popular OpenClinica (https://community.openclinica.com), which is based on 3-tier architecture with an apache tomcat web application server (http://tomcat.apache.org) with a PostgreSQL (http://www.postgresql.org) database backend.

An alternative approach suited for large non-GCP research projects is the establishment of an integrated information technology (IT) framework where structured data from electronic medical patient records are reused for clinical and translational research based on a single source concept of data entry [12–15]. When interfaced against other systems such as laboratory information systems, such frameworks will not only eliminate duplicate documentation requirements for physicians, but can help improve patient safety by providing on-line surveillance of critical events such as adverse drug reactions (ADR) [16,17]. Since these frameworks heavily rely on the exchange of information between different systems, information is most often expressed using standardized dictionaries, such as WHO-ART for coding ADR or LOINC for using laboratory tests [16,18,19]. Due to their complexity, the establishment of such frameworks requires a major commitment from the health care provider such as major comprehensive cancer centers, limiting their availability and accessibility to the individual researcher. In addition, there is a growing number of web-based solutions for outcome surveillance in a clinical or research setting such as CAISIS (http://www.caisis.org/), OIO (http://sourceforge.net/projects/open-outcomes/), Medintux (http://medintux.org) and FreeMED (http://freemedsoftware.org) as well as mobile solutions for data entry [20].

3. Design considerations

We had previously developed our own client–server application based on an Oracle database (Oracle Corp. Inc., Redwood City/CA) with Oracle Forms graphical clients [8] for data management in a prospective randomized multicenter trial. The MSDS trial on external beam radiotherapy (RTx) for locally advanced differentiated thyroid cancer (DTC) was run in close collaboration with the Department of Biometrics/Competence Centre for Clinical Studies (KKS) at the University of Münster. Challenges in managing the trial were its interdisciplinary design involving endocrine surgery, pathology, radiotherapy, and nuclear medicine with separate reference centers for each specialty, and the trial's size (429 patients), duration (10 years), and geographical distribution (50 participating centers in 3 countries). As none of the then available CDMS were found to be suited to the task within the funding constraints of the trial, we decided to proceed with our own development based