ICT for Health Science Research
A. Shabo (Shvo) et al. (Eds.)

doi:10.3233/978-1-61499-959-1-211

Scenarios for Using OpenClinica in Academic Clinical Trials

Matthias LÖBE[a,1], Frank MEINEKE[a] and Alfred WINTER[a]
[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Universität Leipzig, Germany*

Abstract. Clinical Data Management Systems (CDMS) are used to electronically capture and store data about study participants in clinical trials. CDMS tend to be superior compared to paper-based data capture with respect to data quality, consistency, completeness and traceability. Nevertheless, their application is not default – especially in small-scale, academic clinical studies. While clinical researchers can choose from many different software vendors, the vast requirements of data management and the growing need for integration with other systems make it hard to select the most suitable one. Additionally, the financial and personnel costs for purchasing, deploying and maintaining a commercial solution can easily go beyond the limits of a research project's resources. The aim of this paper is to assess the suitability of the web-based open-source software OpenClinica for academic clinical trials with regards to functionalities required in a large research network.

Keywords. Clinical Data Management Systems, Clinical Research Informatics

1. Introduction

Electronic data capture is supposed to be a state-of-the-art technique for clinical trials. The usage of Clinical Data Management Systems (CDMS) for the electronic collection of data from study participants has increased over the last years [1]. However, implementing a CDMS is a costly and time-consuming task and many single-site-single-investigator studies still rely on simple data acquisition tools like Microsoft Excel. Open-source CMDS have the advantage that they can be tested in practice without having to buy a license first. When we decided for OpenClinica 8 years ago, there were a number of other systems referenced on the internet claiming to be full-fledged open-source CDMS (Phosco, Obtima, Visitrial), but in terms of actual availability, functionality and demonstrable practical application beyond the programming institution, they were not real competitors. The now-popular tool REDCap [2] is available free of charge (if one becomes a member of the REDCap consortium), but the license terms expressly prohibit the provision of paid services to third parties, so it could not be used in our projects.

OpenClinica[2] exists since 2005 and is developed by a company. Since the release of version 3.0 (2009), the popularity of OpenClinica has increased demonstrably in the

[1] Corresponding Author: Matthias Löbe, IMISE Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany; E-Mail: matthias.loebe@imise.uni-leipzig.de.
[2] https://openclinica.com/

academic environment. OpenClinica is licensed under the GNU Lesser General Public License (LGPL), which even allows a commercial exploitation of own developments with OpenClinica as basis. Furthermore, OpenClinica is regulatory compliant to the rules of Good Clinical Practice (GCP), the data protection regulations of HIPAA as well as the regulations of the FDA 21 CFR Part 11. The company behind OpenClinica generates revenue through user support, training, study hosting and system validation support. Another advantage of OpenClinica is its modern web-based architecture, without the use of bleeding edge libraries that are prone to changes and potentially prone to errors, which makes it technically very stable. It is written in Java, uses the Apache Tomcat application server, PostgreSQL as database and proven libraries like JSP, Spring or Hibernate (all components are also open-source). The internal data model is closely based on CDISC's Operational Data Model (ODM), so data can be exchanged with other CDMSs without major changes. The internal conceptual model of the database resembles an EAV (entity - attribute - value) scheme.

2. Methods

Our work took place in the context of the Center for Sepsis Control & Care (CSCC) Jena. CSCC covers all aspects of sepsis, from risk prediction to long-term sequelae and health economy and provides resources for several clinical trials. Because of a very tight timeframe, an explorative evaluation method was applied in which the suitability of OpenClinica was tested on concrete practical implementations of various clinical trial designs (see Table 1).

Table 1. Overview of the most important requirements

Requirement	Type	Met	Study
Support for multiple study sites	Functional	Yes	MEDUSA
API for automated data import	Functional	Yes	MEDUSA
Support for different versions of a CRF	Functional	Yes	MEDUSA
Data quality – format checks	Functional	Yes	REGISTRY
Data quality – mathematical validations	Functional	Yes	REGISTRY
Data quality – logical checks	Functional	Yes	REGISTRY
Patient calendar with scheduling and reminders	Functional	Partly	REGISTRY
Repeating of CRFs or groups of data elements	Functional	Yes	ALERTS
Support for Subject Screening	Functional	No	ALERTS
Generating pseudonyms out of identity attributes	Functional	No	ALERTS
Support for trial arms	Functional	Yes	NeuroPAIN
Reusing existing CRFs in another study	Functional	Yes	OSARST
Randomization of subjects	Functional	No	SMOOTH
Database locking/freezing	Functional	Yes	EIDECS
Restrict visibility of data only to the subject of a CRF	Functional	No	EIDECS
Support for paper-based CRFs	Functional	Yes	ACTION
Double Data Entry	Functional	Yes	ACTION
Dunning for incomplete CRFs	Functional	No	ACTION
Easy to use graphical user interface	Non-functional	Yes	MEDUSA
Stability during heavy use	Non-functional	Yes	INSEP
Sufficient performance during heavy use	Non-functional	Yes	INSEP

The first study, MEDUSA, was a multicenter trial with 45 sites and 250 users. We installed three instances of OpenClinica to reflect the traditional approach of development - test - production. OpenClinica supports multiple centers, which can assign their own user roles and the patient data is only visible within this center. A

special feature was that some sites did not want to enter the data via the graphical user interface, as they already enter the data in another documentation system. Here the data was imported via a SOAP web service in ODM format. Due to lack of time, no training could be carried out, but the users were able to get along without any problems with the help of an instruction document. Because of the fast start some Case Report Forms (CRF) had to be corrected later. OpenClinica supports the concept of different versions of a CRF in one study.

The second project was called REGISTRY. The sepsis registry acquires baseline data from Jena sepsis patients at different time points. It served as a core data set which means that wherever adequate and possible, data elements from future clinical trials should refer to data elements in the registry instead of collecting redundant copies. The sepsis registry made heavy use of data quality rules, with format checks (room numbers, time stamps) and validations (laboratory value ranges) as well as many cross-field rules with a need for calculations („Not more than 2 days after admission to ICU"). OpenClinica was able to represent all the required check logics. Although visits can be scheduled, there is no active reminder of upcoming appointments or reminder of overdue visits.

ALERTS was a complex trial with a large number of subjects, because every patient of the University Hospital, which met certain signs of infection, should be recorded as part of a screening process. ALERTS made strong use of OpenClincia's ability to repeat individual CRFs or groups of data elements without having to specify them redundantly. However, integration with the Hospital Information System (HIS) was not easy; a separate screening list application had to be provided. Support for single-source scenarios (importing data from or into the HIS) was made possible in a later project with the help of the MirthConnect software [3]. Since no pateint-identifying data should be contained in a study database, a service for generating and linking pseudonyms from patient clear names had to be connected.

The next project, NeuroPAIN, was a complex trial with mixed cohorts. It investigated subjects from four different groups: ICU patients with sepsis, ICU patients without sepsis, patient suffering from hepatic encephalopathy and healthy volunteers. While some data was collected using the same CRFs for all four groups, other CRFs are very specific and must not be collected from the wrong group. This could be realized by designing the individual study arms as sites. Individual sites can have different events and CRFs in OpenClinica.

The OSARST trial was the first one to allow for reusing complete CRFs originally specified in another project made possible by a universal Data Dictionary.

SMOOTH was the first study to require a randomization function. OpenClinica does not offer such a function, but external solutions with advanced algorithms and stratifications can be integrated [4].

The EIDECS trial tried to develop improved structures of communication in end-of-life decision making and also aimed at reducing symptoms of burnout in caregivers. In this study, four different cohorts are subject of investigation: patients, doctors and caregivers at the ICU, relatives, and the family physician. Since the study also contained questions on burnout and the agreement of nursing staff and assistant physicians with decisions made by senior and chief physicians (and since data in OpenClinica are always visible for every investigator at a site), special levels of data privacy were demanded. Furthermore, EIDECS was the first trial to consist of several phases and idle periods between where recruiting /data entry must not be possible. This could be realized by the database locking/freezing feature.

ACTION was a study implemented in OpenClinica, where it was known that the later data collection would be done first on paper CRFs. This meant that features such as the page of the CRF, numbering of questions and more complex structuring with (sub-) headings, explanations and text markups were particularly in demand. This is supported by OpenClinica, whereby more complex layouts are also possible through the possibility of specifying one's own CSS declarations. However, the print layout of the CRFs was not exactly perfect. Double Data Entry is supported by OpenClinica, but a dunning procedure for overdue CRFs is missing.

Finally, the INSEP trial by the Sepsis Competence Network (SepNet) did not pose any particular challenges to the functional richness of the software, but showed the performance of the system even in low resource settings. In order to reliably estimate the incidence of severe sepsis, 446 investigators in 154 intensive care units recorded and documented 12.305 suspected cases of sepsis during a three-month period (see Figure 1). Application server and database were located on the same server, an obsolete 2005 Xeon with 2 CPUs each with one core, 4 GB RAM and 30 GB hard disk, without any complaints about performance bottlenecks.

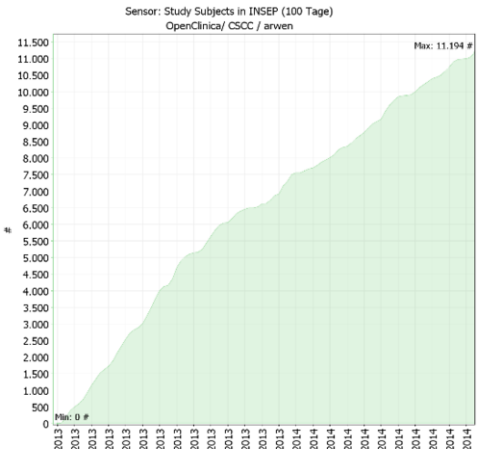

Figure 1: New study subject enrolments for the INSEP trial for a three-month period

3. Results

Although the report presented here was not preceded by a formal evaluation, the experience gained and the multitude and heterogeneity of the clinical trials implemented demonstrate the suitability of the chosen approach. OpenClinica showed to be a quite simple and robust tool for clinical data management. Our overall basic requirements for the EDC system were: platform independent web-interface, simple e-CRF modeling, support for monitoring, query management and reporting, support for exports to statistical software like SPSS or SAS, full GCP conformance and low license costs. Further requirements, which became apparent during the course of the project, could be covered by the majority (see Table 1).

Since none of the studies was aimed at the approval of new drugs, various aspects of pharmacovigilance, such as management and reporting of serious adverse events, coding with medical terminologies such as MedDRA or annotation of study data with CDISC SDTM could be dispensed with.

Nevertheless, OpenClinica is not a sole solution for all problems related to clinical trial management. We have developed a number of our own auxiliary tools, e.g. for creating XML rules, generating Annotated CRFs, storing data elements in a central Metadata Repository, highlighting differences between CRF versions or outputting CRFs in a desired PDF layout. But while some functions may be missing, the availability of source code, the presence of programming interfaces (SOAP and REST) and the available technical documentation allow OpenClinica to be embedded into

complex research infrastructures (see e.g. [5]). The support of widespread interoperability standards of clinical research such as CDISC ODM also easily allows further utilization in clinical research data warehouses such as i2b2 or graphical analysis frameworks such as tranSMART [3, 6].

4. Conclusion

OpenClinica has shown that free and open-source software tools exist that cover a wide range of clinical data management functional requirements and so satisfy academic user needs. The paper is therefore aimed first and foremost at researchers who do not have the means to procure a commercial system or to outsource tasks to external service providers. Even if the learning and maintenance of such a system requires not negligible resources, electronic data capture using a professional software is highly recommended with regards to data quality, data privacy and traceability of changes (audit trail) [7]. It is therefore preferable to error-prone solutions for data acquisition such as direct input, e.g. in Microsoft Excel, or self-programmed and difficult to maintain web forms.

Acknowledgments

This work has been supported by BMBF research grants No's. 01ZZ1609A 031L0026 and DFG grant WI 1605/10-1.

References

[1] W. Kuchinke *et al.,* "Heterogeneity prevails: the state of clinical trial data management in Europe - results of a survey of ECRIN centres," (eng), *Trials*, vol. 11, p. 79, 2010.

[2] P. A. Harris *et al.,* "Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support," (eng), *Journal of biomedical informatics*, vol. 42, no. 2, pp. 377–381, 2009.

[3] J. C. Camacho Rodriguez, S. Stäubert, and M. Löbe, "Automated Import of Clinical Data from HL7 Messages into OpenClinica and tranSMART Using Mirth Connect," (eng), *Studies in health technology and informatics*, vol. 228, pp. 317–321, 2016.

[4] D. Schrimpf, M. Haag, and L. R. Pilz, "Possible combinations of electronic data capture and randomization systems. principles and the realization with RANDI2 and OpenClinica," (eng), *Methods of information in medicine*, vol. 53, no. 3, pp. 202–207, 2014.

[5] T. M. Deserno, V. Deserno, D. Haak, and K. Kabino, "Digital Imaging and Electronic Data Capture in Multi-Center Clinical Trials," (eng), *Studies in health technology and informatics*, vol. 216, p. 930, 2015.

[6] F. A. Meineke, S. Stäubert, M. Löbe, and A. Winter, "A comprehensive clinical research database based on CDISC ODM and i2b2," (eng), *Studies in health technology and informatics*, vol. 205, pp. 1115–1119, 2014.

[7] C. Daniel and D. Kalra, "Clinical Research Informatics: Contributions from 2017," (eng), *Yearbook of medical informatics*, vol. 27, no. 1, pp. 177–183, 2018.