User Centred Networked Health Care
A. Moen et al. (Eds.)
IOS Press, 2011
© 2011 European Federation for Medical Informatics. All rights reserved.
doi:10.3233/978-1-60750-806-9-892

IT Infrastructure Components to Support Clinical Care and Translational Research Projects in a Comprehensive Cancer Center

Hans-Ulrich PROKOSCH[a,b1], Markus RIES[b], Alexander BEYER[b], Martin SCHWENK[b], Christof SEGGEWIES[b], Felix KÖPCKE[a], Sebastian MATE[a], Marcus MARTIN[b], Barbara BÄRTHLEIN[a], Matthias W. BECKMANN[c,d], Michael STÜRZL[e], Roland CRONER[f], Bernd WULLICH[g], Thomas GANSLANDT[b], Thomas BÜRKLE[a]

[a] *Chair of Medical Informatics, University Erlangen-Nuremberg, Germany*
[b] *Medical Informatics & Comm. Center,*
[c] *University Cancer Center Erlangen,*
[d] *Dept. of Obstetrics and Gynecology,*
[e] *Division of Molecular and Exp. Surgery,*
[f] *Dept. of Surgery,*
[g] *Dept. of Urology,*
University Hospital Erlangen, Germany

Abstract. This paper presents the concept of an integrated IT infrastructure framework established at the comprehensive cancer center at the University Hospital Erlangen. The framework is based on the single source concept where data from the electronic medical record are reused for clinical and translational research projects. The applicability of the approach is illustrated by two case studies from colon cancer and prostate cancer research projects.

Keywords. Comprehensive cancer, center, cancer documentation, single source concept, translational research, IT infrastructure framework

1. Introduction

Oncology care is provided in complex transsectoral and interdisciplinary networks of service providers. Within cancer research in recent years we have seen a massive growth in data, especially when molecular, genomic and clinical data shall be linked [1]. In Germany comprehensive cancer centers have been established in order to provide centers of excellence for cancer care, medical education as well as clinical and translational cancer research. Traditionally however, many data collections and IT components in hospitals and research institutions have been developed and implemented independently from each other and typically without any crosslinks. In this context, Beckmann and colleagues have complained about the enormous multiplied documentation requirements for physicians [2]; Shortliffe and Sondick have emphasized, "if the submission of data for research and monitoring purposes requires

[1] Corresponding author: E-mail: ulli.prokosch@uk-erlangen.de

an extra step, . . . the process will likely fail" [3]. This has lead to the design and implementation of integrated informatics research platforms on one side [4] and single source solutions [5] on the other side. In the implementation phase of the Erlangen University Cancer Center (UCCE) in 2007/2008 it has been realized, that a comprehensive and integrated information technology framework with a high level of data reuse will be a major pillar of a successful comprehensive cancer center. In this publication we describe the architecture of such a framework supporting both: cancer care and cancer research. We further present two small case studies illustrating the value gained already from this implementation.

2. Methods

At Erlangen University Hospital a comprehensive workflow-based electronic medical record system (EMR Soarian® from Siemens; compare [6]) has been stepwise introduced within the last decade. Furthermore until 2008 clinical cancer registration was still performed as separate data entry based on paper chart review or pathology reports. For this purpose at Erlangen University Hospital an Oracle-based proprietary cancer documentation system (TUREK-2) has been established. Before the UCCE was started clinical trial documentation was based on paper or using individual software solutions. Biobanking was decentralized with specimen tracking and annotation data often stored in excel sheets. When in 2007 the UCCE IT infrastructure concept was defined, it was decided that a single source approach with the Soarian® EMR closely connected with the clinical cancer registry database as core components should be pursued. Further, those core components should be complemented by commercially available standard products wherever possible. Thus the requirement specification consisted of an integrated framework comprising 1) the Soarian® electronic medical record, 2) the clinical cancer registry database, 3) a centralized biobanking management software, 4) a central clinical trials database, 5) a flexible clinical data warehouse and 6) standard services to assure compliance with data protection requirements in this environment.

Keeping the single source documentation approach in mind, it implied that data should be captured only once at its origin and be afterwards available for multiple reuse. Ideally, digital structured data acquisition should be an integrated part of the clinical treatment process. Thus, an analysis of the clinical workflows related with the various steps in cancer care was performed. Based on this the Soarian® electronic medical record system has been extended with numerous workflow-supported assessment forms for the documentation of cancer anamneses, diagnostic data, therapy documentation and follow-up data. Additionally all data entry forms were based on the German wide standardized definition of a minimal basic cancer dataset. The therapeutic decision process for cancer patients, pursued within interdisciplinary cancer conferences has been supported with conference planning and documentation forms.

The proprietary homegrown clinical cancer registry database has been substituted by GTDS® (a cancer registration system once developed with funding support of the German Ministry of Health and the German Cancer Society and today used in more than 60 clinical cancer registries throughout Germany) [7]. Biobank management support is provided by the commercially available Starlims® system. The GCP-certified commercially available clinical trials management system SecuTrial® has been established as a campus-wide platform for clinical trials. In addition to the existing

commercial Cognos® data warehouse, the I2B2 toolbox has been evaluated and established as a user-friendly and flexible clinical data warehouse [8, 9]. Finally, secure data flow between the framework´s components and compliance to the German data protection law is supported by standardized modules provided by the German Technology and Method Platform for Networked Medical Research (TMF) [10].

3. Results

The value of the IT infrastructure framework established at the UCCE shall be illustrated with two early case studies, where the above components have been applied.

3.1. The Polyprobe Project

Polyprobe is a multicentric research project, aiming at validating the major predicitive/prognostic genes for colorectal cancer in a prospective diagnostic study by applying novel automatized nucleic acid extraction procedures from formalin-fixed paraffine-embedded tissues and quantitative RT-PCR procedures for high-throughput gene expression analyses of 61 marker genes. Within a period of 3 years 650 patients shall be included in the study. The IT concept within this project completely follows the single source idea. Nine assessment forms have been implemented within the EMR to capture diagnostic, therapeutic and study-specific data integrated into the colon cancer treatment process. Within the EMR those data are identified by a hospital-wide patient identification as well as a pseudonym generated in advance for all study participants. Patient consent is documented within the EMR as well. After a final quality validation by a research physician, those data are flagged to allow the export of pseudonymized records into a CSV format, which directly matches the import format of the SecuTrial® clinical trials management system. Thus, regular data import of quality assured data into the research database is supported. Biospecimens extracted within surgery or endoscopy are transferred to the pathology department for diagnostic purposes as well as storage within the UCCE biobank for further research analysis. Those specimens stored for research purposes are identified with special probe identification numbers, which are documented as linking information within the patient´s pathology report in the EMR and also imported into probe related records in the SecuTrial® system. Besides its batch import functionality SecuTrial® provides secure web-based data entry forms which support direct eCRF-based data entry for the second study center (Frankfurt University Hospital) which has not yet been able to also implement a single source approach. Until today 141 patients have been enrolled into the study at Erlangen and were documented within the EMR. From those, data of 20 patients have currently been imported into SecuTrial® and released for monitoring purposes. The external project monitors use the SecuTrial® monitoring workflows for their study specific quality management process.

3.2. The German Prostate Cancer Consortium Database

The German Prostate Cancer Consortium comprises a group of more than 70 urologists, pathologists and basic researchers throughout Germany. Founded in 2003 their aim is to improve prostate cancer research with interdisciplinary and cross-institutional cooperation. For this purposes between 2007 and 2009 a web-based joint

research database has been established including data on prostate cancer diagnosis, therapy and follow-up as well as the characterization of biospecimens collected at the participating centers. Data capture for this database was provided through web-based data entry screens. Despite the high research interest of all partners this solution was finally not accepted because it required time-consuming manual data entry of parameters which usually in similar form have already been documented in the local medical record system. Thus, it was decided to move towards a single source/data warehouse approach, reusing data already documented in local electronic medical records. Urology clinics at Erlangen and Münster University hospital were chosen as pilot centers, since both of them had already established a comprehensive prostate cancer documentation within their EMR system. For data protection reasons a two level architecture was established using three I2B2 installations specifically extended and adapted for this scenario. Every participating partner (currently Münster and Erlangen) has a local I2B2 installation. Datasets are regularly exported from the EMR systems, pseudonymized and imported into the local I2B2 data warehouse. Thus, those local I2B2 instances do already provide query and analysis features for the respective Urology Clinics on their "own" data. Regularly researchers can initiate transfer of further anonymized data from the local instances into one common I2B2-DPKK-Research Database. This central I2B2-instance provides a password-protected web-based secure query interface for all DPKK members.

4. Discussion

Due to the complex structure of oncology documentation, which originates over long treatment periods in different clinical disciplines, implementation of an IT-based documentation process is a complex mission. Oncology data are generated by different clinical specialties, clinical care documentation and research databases are traditionally separated, biological and molecular research based on high-throughput systems is often not linked with clinical research. Translational research project in future will need integrated efficient data management platforms which can easily be accessed by various data analysis and data mining tools. The caBIG initiative in the U.S. has aimed at mastering this challenge supported by large funding efforts and a variety of grid-based tools have been developed and applied in various scenarios [11]. Reusing the electronic medical record for clinical research has been identified as one large challenge for medical informatics, therefore tools as the caBIG modules need to be closely integrated with EMR databases [12]. McConnell and colleagues for example have presented a pilot deployment of caTRIP at Duke Comprehensive Cancer Center [4]. Ochs and Casagrande have described their view on "information systems for cancer research" and provided an overview of the systems and interactions needed to handle clinical trials and high-throughput data in cancer research. Their vision was that such systems should ideally interact gracefully with institutional systems for clinical care and would utilize institutional IT infrastructure and expertise.

Large parts of their vision have been implemented at Erlangen University Hospital within the last three years. Workflow-supported EMR-documentation linked with the above described single source concept has enhanced such documentation efforts, making the data available for clinical care (including billing, quality assurance programs and discharge letter creation), clinical cancer registries and research purposes at the same time. Pseudonymization tools developed to meet national data protection

requirements could be integrated seamlessly into the transfer processes between the EMR and research databases.

In the above described case studies the CSV files exported from the EMR database are currently only imported into the SecuTrial® or the I2B2 databases respectively. In a next step those data will also be used as upload-/import-files for the Erlangen Cancer Registry. Additionally, in the future UCC defined core data records of all cancer patients can be exported from the EMR and imported into a joint I2B2-based UCC research platform. Having linked those data also with the identifiers of the biospecimens within the Starlims® biobank management system, those data can also be used as clinical annotations for the biobank. This illustrates the opportunities arising based on the integrated IT infrastructure framework implemented at Erlangen Comprehensive Cancer Center, making EMR data available for multiple secondary use purposes. Nevertheless, even though this paper illustrates the successful implementation of a single source approach, we shall not neglect that on a semantic and process level, implementing those data reuse concepts has been quite complex.

Major challenges which needed to be mastered were related to the definition of common cancer specific minimal data sets and the alignment of process steps for clinical care documentation, register documentation and trial documentation with each other. Describing all those aspects however, would go beyond the limits of this paper and shall be focus of a separate publication.

Acknowledgement: parts of the described projects have been funded by the German Federal Ministry of Education and Research and by the German Cancer Aid.

References

[1] Ochs, M.F. Casagrande. J.T. Information systems for cancer research, Cancer Invest 26 (2008), 1060-1067.
[2] Beckmann, K. Jud, S. Heusinger, K. Schwenk, M. Bayer, C. Häberle, L. et al. Dokumentation in der gynäkologischen Onkologie, Der Gynäkologe 43 (2010), 400–410.
[3] Shortliffe, E.H. Sondick. E.J. The public health informatics infrastructure: anticipating its role in cancer. Cancer Causes Contr 17,7 (2006), 861.
[4] McConnell, P. Dash, R.C. Chilukuri, R. Pietrobon, R. Johnson, K. Annechiarico, R. Cuticchia, A.J. The cancer translational research informatics platform, BMC Med Inform Decis Mak 8, 60 (2008), doi:10.1186/1472-6947-8-60.
[5] Dugas, M. Breil, B. Thiemann, V. Lechtenbörger, J. Vossen, G. Single source information systems to connect patient care and clinical research, Stud Health Technol Inform 150 (2009), 61-65.
[6] Haux, R. Seggewies, C. Baldauf-Sobez, W. Kullmann, P. Reichert, H. Luedecke, L et al. Soarian - workflow management applied for health care, Methods Inf Med 42 (2003), 25-36.
[7] Altmann, U. Katz, F. Dudeck, J. Das Gießener Tumordokumentationssystem GTDS: Software für klinische Krebsregister, Spiegel der Forschung (2002), 4–10.
[8] Murphy, S.N. Weber, G. Mendis, M. Gainer, V. Chueh, H.C. Churchill, S. et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med InformAssoc 17,2 (2010), 124-130.
[9] Ganslandt, T. Mate, S. Helbing K, Sax, U. Prokosch, H.U.. Unlocking Data for Clinical Research - The German i2b2 Experience. Appl Clin Inf 2 (2011), 116–127.
[10] Helbing, K. Demiroglu, S.YRakebrandt, .F. Pommerening, K. Rienhoff, O. Sax. U. A Data Protection Scheme for Medical Research Networks. Review after Five Years of Operation. Methods Inf Med 49, 6 (2010) 601-607.
[11] Fenstermacher, D. Street, C. McSherry, T. Nayak, V. Overby, C. Feldman, M. The Cancer Biomedical Informatics Grid (caBIG). Conf Proc IEEE Eng Med Biol Soc 1 (2005), 743-746.
[12] Prokosch, H.U. Ganslandt, T. Perspectives for medical informatics - Reusing the electronic medical record for clinical research, Methods Inf Med 48, 1 (2009), 38–44.