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

doi:10.3233/978-1-61499-101-4-1105

Conducting Requirements Analyses for Research using Routinely Collected Health Data: a Model Driven Approach

Simon de LUSIGNAN[a,1] Josephine CASHMAN[a] Norman POH[a] Georgios MICHALAKIDIS[b] Aaron MASON[b] Terry DESOMBRE[a] Paul KRAUSE[b]
[a] *Department of Health Care Management and Policy, University of Surrey, UK*
[b] *Department of Computing, University of Surrey, GUILDFORD, UK*

Abstract. Background: Medical research increasingly requires the linkage of data from different sources. Conducting a requirements analysis for a new application is an established part of software engineering, but rarely reported in the biomedical literature; and no generic approaches have been published as to how to link heterogeneous health data. Methods: Literature review, followed by a consensus process to define how requirements for research, using, multiple data sources might be modeled. Results: We have developed a requirements analysis: *i*-ScheDULEs –The first components of the modeling process are indexing and create a rich picture of the research study. Secondly, we developed a series of reference models of progressive complexity: Data flow diagrams (DFD) to define data requirements; unified modeling language (UML) use case diagrams to capture study specific and governance requirements; and finally, business process models, using business process modeling notation (BPMN). Discussion: These requirements and their associated models should become part of research study protocols.

Keywords. Medical Informatics, research methods, computing methodologies

Introduction

Requirements analyses are part of software engineering, with the purpose of reducing the chances of failure or the need for frequent expensive changes further along the development pathway [1,2]. Conducing requirements analysis as a routine part of protocol development may be appropriate because medical research is a complex process increasingly linking heterogeneous data sources held in different computerised data repositories.

There is little written about appropriate methods for conducting a requirements analysis in the biomedical literature. The following have been proposed: mind maps and scenarios (use cases) [3]; Rational Unified Process (RUP) – an agile process which lists business process modeling and agile methods of assessing requirements as the first two components of its application development process [4], with and without the use of the Generic Component Model [5]; Domain analysis modeling, which compares

[1] Corresponding Author. Simon de Lusignan, Professor of Primary Care and Clinical Informatics, Department of Health Care Policy and Management, University of Surrey, GUILDFORD, GU2 7XH, UK s.lusignan@surrey.ac.uk

applications in the same field [6]; and Action Design, another interactive agile method of application development and theoretically a fusion of action and design research methods [7].

Our aim was to develop a step-by-step approach that standardized the way that requirements analyses were constructed for single studies using heterogeneous datasets; where possible, creating reference models which could be customized to meet the needs of different studies.

1. Methods

Our method combined a literature review; experiential learning from over 15 years processing of routine data, and the opportunities presented through national and international medical informatics groups.

2. Results

The experiential learning from the clinical informatics research group (www.clinif.eu) focused on data, study specific, and governance requirements: making sure that we could identify and extract the correct variables [8], and ensured that data were traceable with processes and metadata in place to do this [9]. We increasingly link datasets using secure and private methods [10] and have to ensure appropriate information governance [11].

Table 1. The i-ScheDULEs approach to developing a requirements analysis for combining heterogeneous data for research and how these requirements are met

Requirement	Detailed requirement & Output	Source of information
Indexing	Consistent naming of all actors, stakeholders and entities. To be used in all study documentation	Study protocol & documents these may not be consistent
Data	Map and link data to achieve some degree of interoperability between data sources. Defined in data flow diagrams DFD	Study protocol, from data providers, and stakeholder interviews / analysis
Study specific	Understand the particular study requirements; described in the schema/rich picture and defined within the UML use case diagrams	Study protocol, participants, principal and other study investigators, stakeholders
Standards	List all relevant standards to enable any interoperability of data	Information from study protocol, and data providers
Governance	Ensure permissions & governance process are in place for data extraction, use, storage & curation	Study protocol, registration, legislation and policy
Business	Taking the study as a whole there are sufficient incentives for participation, and that barriers to participation will be overcome. Defined in BPMN diagram for the study	Study protocol and written data sources may provide some insight but largely form stakeholder analysis
Sensitivity analysis	Track record maybe predictive	Bibliographic databases

The international medical informatics working groups reinforced using standards, and added a socio-technical perspective [12]; suggested how informatics requirements might be incorporated into study documentation [13]; and suggested that we should

start to formally model participants' business requirements using business process modeling notation (BPMN).

We use the acronym i-ScheDULEs (indexing, Schema, DFD, Use case UML diagram, businEss process model, and sensitivity Analysis) to describe the process. (1) Indexing: Actors, stakeholders, and entities (using the Unified Modeling Language (UML) definitions of each); (2) Creating a schema or rich picture of the process; (3) Defining the data requirements using a data flow diagram (DFD); (4) Capturing the process as an UML use case diagram to capture study specific, stakeholder and governance requirements; (5) Capturing the business requirements and summarizing that the other requirements have been met; and (6) Conducting a sensitivity analysis based on the track record.

3. Conclusion

The i-ScheDULEs offer a practical approach to conducting a requirements analysis for studies involving the use of heterogeneous data sources. These requirements and their associated models should be a standard part of research study protocols.

References

[1] Cadle J, Paul D, Turner P. Business analysis techniques, 72 Essential Tools for Success. Swindon; BCS, 2010
[2] Reddy M, Pratt W, Dourish P, Shabot M. Sociotechnical requirements analysis for clinical systems. Methods Inf Med. 2003;42(4):437-444
[3] Hanss S, Schaaf T, Wetzel T, Hahn C, Schrader T, Tolxdorff T. Integration of decentralized clinical data in a data warehouse: a service-oriented designand realization. Methods Inf Med. 2009;48(5):414-418
[4] Guo JW, Poynton M, Matney S, Ellington L, Crouch BI, Caravati EM. Requirements analysis for HL7 message development in poison control center. Stud Health Technol Inform. 2009;146:758-759
[5] Oemig F, Blobel B. Harmonizing the semantics of technical terms by the generic component model. Stud Health Technol Inform. 2010;155:115-121
[6] Barton C, Kallem C, Van Dyke P, Mon D, Richesson R. Demonstrating "Collect once, Use Many" - Assimilating Public Health Secondary Data Use Requirements into an Existing Domain Analysis Model. AMIA Annu Symp Proc. 2011;2011:98-107
[7] Timpka T, Johansson M. The need for requirements engineering in the development of clinical decision-support systems: a qualitative study. Methods Inf Med. 1994 May;33(2):227-233
[8] Michalakidis G, Kumarapeli P, Ring A, van Vlymen J, Krause P, de Lusignan S. A system for solution-orientated reporting of errors associated with the extraction of routinely collected clinical data for research and quality improvement. Stud Health Technol Inform. 2010;160(Pt 1):724-728
[9] van Vlymen J, de Lusignan S, Hague N, Chan T, Dzregah B. Ensuring the Quality of Aggregated General Practice Data: Lessons from the Primary Care Data Quality Programme (PCDQ). Stud Health Technol Inform. 2005;116:1010-1015
[10] de Lusignan S, Chan T, Theadom A, Dhoul N. The roles of policy and professionalism in the protection of processed clinical data: a literature review. Int J Med Inform. 2007 Apr;76(4):261-268
[11] de Lusignan S, Navarro R, Chan T, Parry G, Dent-Brown K, Kendrick T. Detecting referral and selection bias by the anonymous linkage of practice, hospital and clinic data using Secure and Private Record Linkage (SAPREL): case study from the evaluation of the Improved Access to Psychological Therapy (IAPT) service. BMC Med Inform Decis Mak. 2011 Oct 13;11:61
[12] de Lusignan S, Pearce C, Shaw NT, Liaw ST, Michalakidis G, Vicente MT, Bainbridge M, What are the barriers to conducting international research using routinely collected primary care data? Stud Health Technol Inform. 2011;165:135-140
[13] de Lusignan S, Liaw ST, Krause P, Curcin V, Vicente MT, Michalakidis G, Agreus L, Leysen P, Shaw N, Mendis K. Key Concepts to Assess the Readiness of Data for International Research. Contribution of the IMIA Primary Health Care Informatics Working Group. Yearb Med Inform. 2011;6(1):112-120