CENTERIS 2013 - Conference on ENTERprise Information Systems / ProjMAN 2013 - International Conference on Project MANagement / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies

Is Common Test Data the Solution to Poor Quality?
Solving the Right Problem
– An Analysis of a Public Health Information System

Sigrid Eldh[a,b], Martin Blom[a]

ªDept. of Mathematics and Computer Science, Karlstad University, Karlstad, Sweden
ᵇEricsson AB, Stockholm, Sweden

Abstract

This paper reports our initial findings regarding the state of testing of software in the Swedish public health information system. At present, the system is only available through a black-box interface, i.e. through the GUI. This and other issues related to politics, management and organization indicate that much work is needed in order for the software to have the quality level expected by a safety-critical system. The proposed solution by the public health organization for raising the quality is to use an independent test database. Based on our initial understanding of the problem, we argue that there might be other solutions that would perhaps be more cost-effective and have a stronger impact on the quality of the system. Our main contribution lies in the data analysis, where we have collected the problems and suggested alternative cost-saving solutions.

Keywords: Software testing; Public health care; Acceptance testing; Test automation; Interoperability; Information System;

1. Introduction

This paper describes a public health information system provided by Cambio Cosmic HealthCare Systems [1] that manages administrative work such as e.g. patient medical records, registration, billing and journal management. Our focus in this case study is to describe the context of this case as we have understood the

current political, managerial and technical state oich the regional council (Swe. Landstinget) Värmland, which is one of the users of this system.

The research questions we aim to answer are the following:
- RQ1: What are the reasons for the perceived poor quality of the system?
- RQ2: What are the proposed remedies to these problems?

The problems with computerization of a public health information system show similarities to other computerizations of information systems. A recent study (Oct 2012) shows that Sweden is taking the world lead in having fully implemented an electronic patient health record (EPR) information system and eHealth [12]. Cosmic holds 25.8 % of the market in Sweden, making it the second largest provider behind Meilor, which holds 27.6%.

There have been at least two previous case studies performed on this system. The first study [2] from 2007 was conducted with interviews and one survey. Their main findings were:
- Advantages in data accuracy when transforming patient medical notes into accessible and readable electronic information
- Some resistance to computerization among users
- None or little training preceding the introduction
- The slow performance of the system and
- The limited access to hardware

This study does not address testing or quality aspects of the system, other than in comments like "could improve quality and customer security" and "some tasks take longer". Both statements might be based on the users' unfamiliarity with using computers.

The second study [3] from 2011, talks about the system "As a savior or a curse?" It contains direct comments on poor quality, including examples of software failures and problems with system upgrades as well as problems with interoperability with other systems. New risks have been introduced as the system is now accessible from "anywhere", but old risks like losing physical papers have been removed. In this case study, the opinions are more balanced in terms of improvements and problems and paint a more nuanced picture of how software is changing the public sector.

We can add to the above studies the observation that the expectancy of how a "great" software system should behave, is changing due to maturity of computer usage. Users are now demanding fast and user-friendly information access available at an instance in conjunction with high reliability of quality data. There is a wish to improve the process and flows in the system to better accommodate the users. It seems that the views of the users has changed in the computerization process and with growing maturity and that the users demand more of the system. With the more flexible and available system some safety-critical risks becomes more in focus as well as security concerns. A better testing process would not only remove some of the risks, but also make the quality visible, something both provider and customer would benefit from. It confirms the Weber-Janke et. al. [6] insight that the socio-technical fit in conjunction with engineering a robust system creates a challenge to engineering.

Our paper is organized as follows: In chapter 2 we present the background and context of this case study. In chapter 3 we describe the problems and our suggested solutions. In chapter 4 we describe the NordicMedTest suggested test data solution and in chapter 5 we discuss our threats to validity. Finally, in chapter 6, we conclude our paper and describe some of our further work.

2. Background and Context of the Case Study

In this section we describe the methodology used as well as the background and context [5] of the case study.

2.1. Methodology

A series of data has been collected using the following sources;
- A system overview and capture of screens
- Short informal interviews with the acceptance testers and their manager
- An assessment by independent consultants
- A workshop with representatives from a series of stakeholders of the project NordicMedTest
- Selected documentation and related work

The system overview was presented in a two sessions with the users and management of the acceptance testing team present at the regional council Landstinget Värmland. Several short informal interviews have been held with stakeholders of the system with both authors present, e.g. managers, users, test-consultants. The assessment by independent consultants was conducted in collaboration with one of the authors and a short informal survey was made with acceptance testing team at Landstinget Värmland. A two day workshop was held with a series of stakeholders of the project NordicMedTest, including users, administrators of the system, local architects and several levels of management from Landstinget Värmland, and also Inera AB (owned by government i.e. the county councils). This workshop was well documented and gave a more in depth reflection and justification of our informal interviews, as well as solution proposals. In addition we have reviewed selected documentation and related work. We can conclude that many details were hidden from us researchers since no non-disclosure agreement were signed giving access to non-public information. Our information is based on informal meetings and interviews as well the public workshop. We researchers want to be clear that we have not had contact with neither Cambio nor Inera, and they have not had a chance to comment on our findings. Equally, NordicMedTest or Landstinget has not yet taken part or being able to comment on the result of our analysis. Issues related to the quality and validity of these sources will be discussed in the section on validity threats.

2.2. System Description

The focal system [1] in our study is a commercial software product consisting mainly of a public administrative relational database. The system is currently in its 7th version, and could be considered mature. It is tailored to each seven regional governments that use it, all with individual adaptations, and all with between 100-300 care centers geographically spread in their region.

The quality of the system is generally considered "poor" according to the limited number of users we have been in contact with – which is a qualitative assessment in lack of defect and size data on the system. We cannot know if the problems that have caused health consequences in the past still persist. Apart from these indications, not much of the product development, such as architecture, performed tests, quality assessments, or even records of defects are available to users, since it is a proprietary commercial system.

2.3. Government Procurement

All Public Health Software Systems are specified by a government body, Inera AB, whose purpose is to collect requirements and handle the procurement in a bidding situation. It is required by regulation that the lowest bid fulfilling the requirements specification wins the contract. The requirements specification must encompass all aspects of what is expected of the software, and the focus is on the (formerly) manual processes that are transferred into software. As Sweden have laws of co-decision, making e.g. users of systems have a great impact on the result, but there is no prototyping or co-development going on as described in [13]. Buying software systems on specification up front does not result in the best systems. Inera, who held the initial bidding process, has revealed that they only required 3% of the total development cost to be spent on testing, as a general rule. Since the software provider was allowed to spend this little effort on testing in combination with all the local adaptions, it has been necessary to make sure that testing is done by the customers themselves in order to meet the quality requirements of the users. A consequence of the local adaptation is that all regions have built up groups for acceptance testing for their version of the system.

Inera, as the council representatives, has with the increase cost of testing and support understood that 3 % testing is a grave underestimation. Especially in the light of some serious incidents, Inera has made sure that in the procurement of new systems, the software development cost should include at least 20% testing. In our opinion this estimate might still be a grave underestimation, since the system should be classed safety-critical, and would probably require a quality and testing effort about 50-70%. Secondly, increasing time spend is not sufficient. Any future requirement specifications must improve the quality requirements and include specific requirements on test procedures and measurements. It is also not sufficient to only focus on the process and procedures, but new quality requirements must also be added and improved on the system, including non-functional requirements, such as robustness. Furthermore, the procurement and bidding should also include intermediate trial periods with users, regular assessments of quality and internal access to some of the assets, e.g. source and test code metrics, to be able to perform independent quality assurance controls on the software.

2.4. Other System Concerns

These above facts, and the history of the system, are used as our hypothesis - that we consider the overall system as poorly tested, which – in lack of other factual evidence will be our basic position.

The size of the system could be considered small by looking at the GUI, which provides the only available access to the system for researchers as well as the regional and local governments. No exact information model or access to the code is available. The data model variation could be considered large, probably due to regional usage.

For all systems in the health sector, security must be enforced. In systems like this, security is mostly confined to access privileges, and extensive logging to be able to spot illegal accesses, which in itself is a large dataset to manage. Additionally, the system is subjected to large amounts of legislation, rules and regulations. For us, and it seems also the acceptance testers, it is unknown how much of these rules and regulations are implemented in the system and how much is left to the user to know and thus to "not make mistakes". A simple example is prescription of drugs. Does the dosage have a maximal limit, or can people by mistake add several zeroes at the end of a number, thus multiplying the dosage by orders of magnitude, leading to disaster? That some of these rules are only acceptance tested by users, might not detect some types of faults. On the other hand data quality and usability of the system, which also includes rules and procedures, cannot be judged easily by a tester. A tester should be able to capture an entire set of user scenarios – and

different user profiles, including logging of access to specific data entries, and a security model of different user access profiles, which also creates a large data volume to handle.

2.5. The Test Process

The steps in the test process are largely unknown since it is internal to Cambio and we have no access to their internal processes. From our interviews with the regional council, Landstinget Värmland, who acceptance test each release, we have understood that these releases come relatively irregularly, and local adaption requires extensive acceptance testing before releasing the new versions. We have observed through our interviews that no test database has been created, and because of limitations in the test scope, software often goes untested in a release. Since all testing is done manually, regression testing is very limited. A set of specific test cases are used, and new test cases are created by bringing new users into the study. Acceptance test is limited, since the only artifacts available apart from the end-user GUI are release notes, updated documentation of the system and specific system change information. These mean that test cases are based more on the user (acceptance testers) specific experiences rather than on test design techniques adapted systematically. System access is only possible in a truly black-box manner since no code is available, probably due to contractual issues with Cambio made by Inera in the bidding process. This prevents many quality improvement techniques and thus the insight in the current testing and quality assurance practices are almost non-existent in the acceptance testing group.

An overview of the stakeholders in relation to the development process, as described in a V-model can be found in Figure 1. In the initial acceptance test the Landstinget Värmland was very active. Now, each of the regions using the system is independently collecting requirements, suggesting changes and reporting faults directly to the commercial company. This also means that each region is getting a new release to acceptance test individually that has the consequence that each region has to acceptance test each other's changes. The result of this process is twofold: It both aligns the usage and processes, and it also causes unwanted local testing. There are also other external users which impact the system releases such as private health providers.
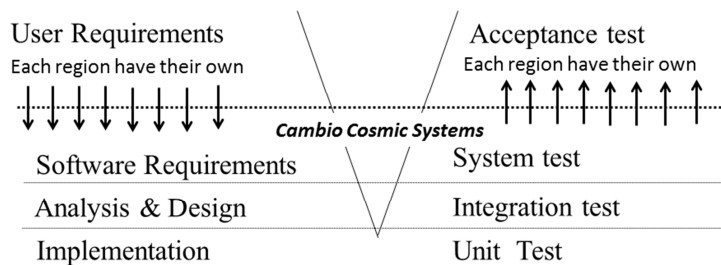

Fig. 1. Overview of the V-model including stakeholders' responsibility of the system

Improving quality in the actual Cambio Cosmic System would be easier, if it was possible to have access to and to influence their test process, something we as researcher hopefully could support. Otherwise the councils have to rely on the current approach to improve the software quality indirectly: by improving all mentioned aspects of requirements and through the defect reports. This is from the regional councils a serious investment and from users and testers standpoint, a much more challenging task to achieve quality assurance.

2.6. Development and Defects

According to the users, and confirmed by the Cambio Healthcare systems website [1] (2013-05-13), the software development is outsourced to Sri Lanka. This fact is used as a claim why "…they are not familiar with the Swedish system and way of working in our Public Health Systems" which in turn causes the software to be less usable. The common opinion of the users and the regional council, is that the system is "correct to the letter" of the requirement, but not in the spirit of usage. This is the most debated problem and almost all users we researchers have met in this process, has lashed out negatively about the system in terms of usage – also corroborated from other regions, as well as the public opinion reflected in newspaper articles. This has caused several doctors, and local care centers to have their own system on the side (one causing a cancer patient not getting treatment on time). In fact, a claim is made that about 20% of Lex Maria complaints are directly or indirectly caused by poor software systems [8]. Lex Maria is a law that requires all incidents to be reported and investigated if they have or could have led to severe harm to an individual [11].

To our understanding, based on questions to the testers of the system, no defect management system exists. This seems to be a serious flaw, as failures and problems are only documented on "loose papers" and not classified and managed in appropriate detail.

2.7. The Organization

Health care processes are highly localized [6] indicating there is a lack of sufficient collaboration. One problem with the autonomy is that the processes are different in different regions and even worse, different within the same region, resulting in different adaptions of the software to fit the local processes. If the processes were unified into one process based on best-practice the software would be enough existing in one version and testing could be centralized and made more efficient. There is no common user group where these sorts of problems could be discussed and solved, most likely due to inter-organizational difficulties. The test know-how in the organization is in general rather low and the main reason for this is that most regional and local public health authorities have no software education. This means that also that there is no coordination regarding requirements and acceptance testing, making this an issue to deal with for the company that also has to try and please all. This lack of coordination seems both costly and time consuming for all stakeholders in the system and creates unnecessary tension. The origin comes from political and regional independence in decision making.

2.8. NordicMedTest Project

The local public health organization joined with Compare in a specific project called "NordicMedTest" [8] with the focus of testing health related software. Compare is a foundation organizing about 100 regional IT-companies, a majority of them being consultant businesses. A major income for many consultant companies is software development in general and software testing in particular. Compare's main focus is called Compare TestLab and most of the foundation's work has been hosting test systems for industry and hosting a vocational education on software testing. Being owned by the local software industry, the test lab is a natural place for locating an independent test center for public health. Compare has together with the regional public health government received funding from the government funding agency, Vinnova [7] to create "new innovative system and services" and also to create an "independent test data" center.

3. Problem Description and Our Suggested Solutions

This section describes the current problems, based on our observations, followed by suggestions on how these problems might be remedied.

- *Perceived lack of usability (causing faults, misuse and possible malpractice)* – A well tested system should not allow the entering of incorrect information. A common problem with system requirements is that they only include what software should do, and omit what it should not do [9]. We suggest identifying critical fields and making sure that safety rules are implemented and certified. A simple investigation and assessment of robustness of the system should also be conducted. Some of the non-critical usability problems could stem from the lack of user computer experience and will hence be remedied as the users mature. Additionally one can make sure users are certified to use the system, requiring mandatory education.

- *No defect management system exists* on the user side – Creating a defect management system is a simple and probably a very effective approach, especially if the defect system is not made local, but open to all regions.

- *No access to the system other than through the end-user GUI* – There is a discrepancy between the interests of a private software developer and a governmental body and apparently the deal was constructed more to the advantage of the software developer. To meet both the need to protect the intellectual property of the company and the need for transparency from a customer standpoint, the code and other artifacts should be accessible under non-disclosure agreement for all testers, acceptance testers of the system as well as researchers.

- *No cooperation (or organized joint work) between user groups in different regions* – In the current context, it would be a good starting point to form small user groups that could be made virtual by modern techniques to save costs.

- *Each region has its own adaption of the system performing acceptance testing of its own branch* – If the system actually was just one system instead of existing in many versions, the problems arising from local adaptations would be removed and all testing could be done at a common test center. This is politically very difficult to go about and would require large organizational changes at the various local sites, but if done money could then instead be spent on improving the system, rather than paying for several adaptions.

- *Nurses (users) are doing the acceptance testing* – Nurses should participate with some tailored test training, but their main role should be as user experts. Testing should be done mainly by software testers with adequate qualifications. Thus, acceptance testing should be automated to create adequate robust legacy regression testing. Software testers would be able to add more negative testing as well as technical testing that probes the system more thoroughly. This work should be thoroughly assessed with high coverage achieved to assure the full penetration in addition to thorough robustness testing.

- *The management of the acceptance test has no experience of testing* – This is an unfortunate situation, common in many organizations, but it has easy remedies. Courses are readily available, as are consultants with this experience, both of which would solve this problem.

- *Procurement of external test support could be challenged* and it fuels the discussion of how to capture expertise. There is no guarantee that hired help will solve the problem and produce a good result, without proper assessments and measurable outcomes in the system. This could be considered a high risk, since it potentially could waste money. A more long term solution is a better focus on establishing higher education (university) courses in software and system testing across the globe as any formal education. In addition, education on software health system and data quality should also be a natural and mandatory part

in both university and vocational education for nurses, doctors, pharmaceutical personnel and other user groups.

- *The testing seems ineffective and inefficient* – This is purely a cost and time matter, as well as the know-how of asking for and assessing the right quality of the company. External quality assessments should be performed on a regular basis in a collaborative manner.
- *Insufficient funds are admittedly invested in testing, quality and efficiency of the adaptation and IT-solutions* – The easy decision is to invest more; the difficult decision is to convince the politicians and leadership in the management and government to do so. Money could be spent on training the local testers, investing in testing tools or paying consultants to help with testing.
- *The system is dependent (in different ways) on 40 other systems* – More research must be made into interoperability requirement management. As so far, we have not focused our effort in this direction, but can easily see this is not solely a testing issue as the current opinions seems to be. In addition, the government procurement body could simply define an agile team to work specifically with both development and testing on interoperability, to make sure these problems does not occur in the future.

4. The Suggested "Test Data" Solution

The idea of a test data center was presented at a workshop describing the NordMedTest project. The goal is that an independent model of test is created at Compare Test Lab. The idea is to create an independent location, where users from all regional councils as well as providers of health-related software can test "new or updated services and solutions in a uniform environment with appropriate integrations to related systems". If the test data should be originating from real data, the question is how one could modify this data without losing its testable value. Real data is both unethical and by law prohibited to use, so some sort of modification is needed.

The problem of having a test data set that is real enough but still preserves the personal integrity is common, especially in medical systems where most data is sensitive. An interesting report on how to construct a dataset using various methods can be found in [10] where the main conclusion is that there is a trade-off between realness of the data, and the preservation of personal integrity, making it in effect a legal and ethical problem. Secondly, this would indeed challenge any local adaptations of the system and imply conformity. Usage process best practices studies will be needed to achieve a best fit.

Complicating the creation of an independent test data solution are the 40 systems that connect to the main health administration system. The interfaces are specified to communicate – but do not do this satisfactory, in part due to lack of proper interface descriptions or possibly lack of understanding of the interfaces. This is interpreted as a test problem, instead of a requirement, re-design and test problem, which it actually is. The test data solution would have to address the connection to the other systems in one way or another. Either the other systems can be seen as input/output ports and entirely faked, or the other systems would have to be emulated in more detail, depending on what aspect of the system is being tested.

5. Threats To Validity

Internal Validity – Much of the present information stems from interviews and informal talks, so there is a risk that we measure personal opinion or preference rather than the quality of the system. The information obtained is rather unison, thus collaborated evidence, as is the assessment by a third party, which allows for some mitigation of this threat. In addition there is the obvious researcher bias on the test and quality matters that probably influence our problem description and solutions proposed.

External Validity – The results are most likely transferrable to other health systems and to any other governmental system that handles personal data that is sensitive.

Construct validity – We have not constructed any artifacts for this study and the construct validity is in principle a non-question.

Conclusion validity – We believe that we have understood reasonably well how the system works, how the users perceive it and what problems exist today regarding the test process. Since the data is informal in its nature and since no formal statistical methods have been used to gauge the results, our conclusions are weak in statistical terms. We believe that our conclusions are probably slightly biased, but are per se correct and well-grounded in the data and hence valid.

6. Conclusions and Further Work

We have attempted to shed light on the current state of affairs regarding testing of Swedish health systems e.g. Cambio Cosmic Healthcare system. The few conclusions that can be drawn from the material presented in this paper are that current test practices are neither efficient nor effective, that testing is done manually and by end-users rather than by professional testers and also that the suggested test data solution only partly will solve the quality problems. In any case, careful handling and of the sensitive health data (EMR) to create unidentifiable test data for use in testing is needed and the trade-off between realness and integrity must be addressed. It seems like the connection to the surrounding 40 systems must be addressed and handled appropriately which is not solely a test problem. It is a combination of proper integration requirements, design and testing problem to solve. Our plan is to support the acceptance testing by aiding in several more complex tasks, e.g. evaluating test automation options, defect analysis and classification, robustness and testability investigations as well as quality assessments. With our work we hope to support Landstinget Värmland and the NordicMedTest Project as an objective and scientific observer without competition with local businesses.

Acknowledgements

References

[1] Cambio Health Care Systems http://www.cambio.se/
[2] Andersson Nazzal L, Ryberg A. Ett vårdinformationssystem i vårdens frontlinje - En fallstudie om Cambio Cosmic på en vårdcentral i Landstinget Kronoberg, Student Thesis, Linneus University, Växjö, Sweden, 2007
[3] Andersson E, Melin U. COSMIC - syndabock eller frälsare - En fallstudie av införandet av och arbetet med ett IT-system för vård- och patientadministration inom Landstinget i Östergötland, Linköping University Electronic Press, LIU-IEI-R 143, 2011
[4] Petersen K, Wohlin C. Context in industrial software engineering research. In: 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM), Orlando, Florida, USA, Oct. 2009
[5] NordicMedTest homepage, www.nordicmedtest.se, retrieved 2013-05-03
[6] Weber-Jahnke, J.H., Price, M., Williams, J.:Software Engineering in Health Care, Is It Really Different? And How to Gain Impact, IEEE WS Health in , 2013
[7] Vinnova, Swedish Governmental Agency for Innovation Systems, www.vinnova.se/en/, retrieved 2013-05-03
[8] NordicMedTest. Vinnova application 2012-0206, http://www.vinnova.se/sv/Resultat/Projekt/Effekta/Testbadd-for-IT-inom-halso--och-sjukvard/

[9]   Eldh S, Sundmark D. Robustness Testing of Mobile Telecommunication Systems, IEEE WS TAIC-PART, Montreal, Canada, Apr.2012

[10] Raza A, Clyde S. Testing Health-care Integrated Systems with anonymized test-data extracted from Production Systems, International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, 2012

[11] Nordlund, Y. G., & Edgren, L. (1999). Patient complaint systems in health care a comparative study between the Netherlands and Sweden. European Journal of Health Law, 6(2), 133-154.

[12] Jerlvall, L. and Pehrsson, T.; eHealth in Swedish County Councils, Inventory comissioned by the SLIT group, CEHIS (swe: Center för eHälsa i Sverige /Eng: Center for eHealth in Sweden), Oct 2012

[13] Hartswood, M., Procter, R., Rouncefield, M., & Sharpe, M.: Being there and doing IT in the workplace: A case study of a co-development approach in healthcare. Proceedings of the CPSR/IFIP WG 9.1 Participatory Design Conference, New York, November 28th-December 1st. 2000.

[14] Detmer, Don E., and Elaine B. Steen. The computer-based record: patient moving from concept toward reality. International journal of bio-medical computing 42.1 (1996): 9-19.