Detlef Hühnlein et al. (Eds.): Open Identity Summit 2015,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015 133

Evaluating Complex Identity Management Systems – The FutureID Approach

Rachelle Sellung[1] Heiko Roßnagel[2]

Abstract: This in-progress paper will discuss the importance of evaluation methods in complex large scale projects, specifically those regarding identity management systems and electronic Identities (eIDs). It will depict the advantages of using a Design Science methodological framework approach and show how the EU project FutureID has utilized this methodology to bring multiple disciplines perspectives together in a harmonized evaluation.

Keywords: Design Science, Large-Scale projects, Evaluation approach, eID's

1 Introduction

A common problem found in many technology-based research projects, specifically in information security, is the sole focus only on the technological aspects. These solutions address mainly issues, such as, security, privacy and reliability [ZR12]. They fail to elicit and consider other requirements such as business and usability requirements. As a result, this approach often veers away from user's needs, markets, and economic contexts. Consequently, there have been multiple security and privacy technologies, which have been designed in a way that often results in market failures; such as, electronic signatures [Ro06] or web anonymity services [FFSS02]. Another strong point mentioned by [ZR12], is that the assumption concluding a technologies market success is solely reliant on their technological sophistication is not satisfactory. When reducing the effort put forth into creating a well-designed business model for the market, it often leads to important factors either not being addressed or not initiated to the best of its capabilities. For example, [GORR04] mentions how technologies often fail to address the user's needs and requirements appropriately with respect to usability and accessibility for both individuals and organizations. In addition, the classical initiation of a technology base project is having the evaluation of the project results being based on a sole evaluation of the pilots. Furthermore, the evaluation results of the pilots are often assumed to be an accurate implication or even forecast on how it would perform in a real market scenario. When including a wider range of disciplines within an evaluation, it becomes quickly apparent that this approach is no longer viable to serve as a well-rounded evaluation for a large complex research project. The FutureID project has taken an alternative approach to address these concerns and challenges. FutureID is a large scale EU project that

[1]Fraunhofer IAO, Identity Management, Nobelstraße 12, 70569 Stuttgart, Germany
firstname.lastname@iao.fraunhofer.de

strives to build a comprehensive, flexible, privacy-aware and ubiquitously usable identity management infrastructure for Europe. Following a viable security approach [ZR12], FutureID considers the interests of all the stakeholders involved in the eID ecosystem to facilitate economic conditions for wide take-up of its results. It combines experts from seven different disciplines that each provided a requirement analysis for three defined artifacts of the project which depict three levels of the project as a whole. These requirements were considered during the design of the artifacts and serve as a basis for the evaluation approach. As a result the project needed to address these needs with a rigorous, flexible, and comprehensive evaluation method. To put into perspective of importance, FutureID's pilots only show a subset of what the results of this project has to offer, specifically in its reference architecture and the implementation. Further, the Design Science Approach is the process decided to address this task. This research in progress paper focuses mainly on the Design Science Evaluation method, specifically how this was addressed in FutureID. The rest of this paper is organized as follows; section two will include the challenges faced, section three goes into detail on the methodology, section four expands on the FutureID approach, section five serves as a discussion and limitations section, and lastly section six is a conclusion.

2 Challenges

Including a variety of disciplines naturally leads to a more complex evaluation. The disciplines included in the FutureID evaluation are Socio-economic, Security, Legal, Privacy, Usability, Accessibility, and Technical. With that, FutureID has faced many challenges in initiating a comprehensive evaluation its artifacts. First, FutureID is a large project that includes 19 different partners from 11 different EU countries. Having such a diverse consortium in many different ways, often leads to challenges regarding harmonizing and compromising all perspectives to create artifacts that are comprehensive and flexible. Second, FutureID aims at having a flexible Reference Architecture, however with that it increases difficulties in initiating an evaluation method that can be just as flexible. For instance, the Reference Architecture evolved throughout the duration and evaluation process of the project due to the increasing needs of requirements from the different disciplines, going beyond what was originally proposed in the project plan. As a result it was not possible to implement all of the new features defined in the Reference Architecture, due to the limited amount of available resources. Further, due to the flexibility of the architecture many different configurations and different forms of deployment are possible, which of course makes an evaluation even more challenging. While FutureID is capable of supporting many different use and business cases, the two pilot applications only focus on two exemplary use cases. One pilot provides Citizen Services in the e-health domain and the other one focuses on e-Learning Services for Enterprises. As a result these pilots are not capable to showcase all of the possibilities. With this conclusive set of challenges, FutureID faced the largest challenge of finding and applying an evaluation approach that would fulfill its comprehensive and flexible needs.

3 Methodology

FutureID uses the Design Science research approach as they are presenting three novel artifacts and a suitable evaluation that address the artifact's appropriateness to contribute to the problems' solution [NCP91] [ZRMS11]. Design Science research is a set of analytical techniques and perspectives that was originally designed for Information Systems. Design Science's achieves knowledge and understanding of a problem domain by building and application of a designed artifact [MMG02] [HMPR04]. The artifact is created to be used as a tool to better understand the problem and to re-evaluate the problem to improve the quality of the design process and to be able to start the process over again [MMG02]. The overall goal of this approach is to create a design process that is a sequence of expert activities that produces an innovative product [WSE09]. Referring to Figure 1, the Design Science research model satisfies two cases; the business needs (relevance) and the knowledge base (rigor). The knowledge base feeds on creating applicable knowledge that will be able to be used to better an artifact that is used in different real world situations. The knowledge base's objective is to be rigorous in a way that the research built upon existing knowledge and then it further contributes as applicable knowledge to an artifact or theory. After it is applied, then it assists in assessment and refinement to further justify and evaluate in a more scientific manner. The knowledge resulting from this process is added to the knowledge base. Simultaneously, the environment side serves more the business-needs assessment of the model. Its goal is to apply the artifact or theory in a relevant way and real world situations. In the Design Science Research model, business needs are assessed and evaluated in consideration of organizational strategies, structures, cultures and already existing business processes [HMPR04]. Furthermore, business needs go through the same process as the knowledge base did, as it is further assessed and refined to justify and evaluate the artifact or theory. The difference with the Environment side is that afterwards it the result is applied in an appropriate environment and then what is learned is returned to the Environment side. This model shows how these two processes work simultaneously and perpetually together to continuously make a method or artifact stronger and more comprehensive. Furthermore, the Design Science Evaluation methods, which are shown in Table 2, are divided into five broad categories; observational, analytical, experimental, testing, and descriptive. These categories cover a wide variety of evaluation methods; such as, case studies, dynamic analysis, simulation, functional testing, or informed argumentation. Each evaluation method shouldn't be considered or weighed at the same consistency as an informed argument is not as credible or reliable as a field study. An advantage to the Design Science evaluation method is that these five categories are flexible enough to be applied in many different disciplines despite the range of different techniques. These methods can be applied to a wide variety of research fields whether it's in law or in a more technical field. The Design Science evaluation methods are flexible, but organized. This provides a strong argument to how one can organize a variety of interdisciplinary evaluation methods. Overall, Design Science has a strong and comprehensive research model, dependable guidelines, and wide spread evaluation methods. In the FutureID project, we have 7 discipline teams of experts, who follow this approach and apply their own evaluation techniques within the realm of the Design Science Evaluation Methods. As a result of this application, each discipline creates a list of requirements that each artifact would have to fulfill. This results in a basis from a Design Science Methodology framework for the interdisciplinary evaluation of the main artifacts of the FutureID project.

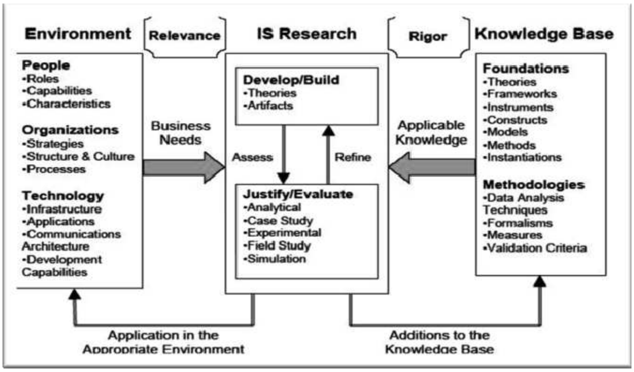

Fig. 1: Design Science Research Model [HMPR04]

Observational	Case Study: Study artifact in depth and business environment
	Field Study: Monitor use of artifact in multiple projects
Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g. complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g. performance)
Experimental	Controlled Experiment: Study artifact in use for dynamic qualities (e.g. usability)
	Simulation: Execute artifact with artificial data
Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Preform coverage testing of some metric (e.g. execution paths) in the artifact implementation
Descriptive	Informed Argument: Use information from the knowledge base(e.g. relevant research) to build a convincing argument for the artifacts utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Tab. 2: Design Science Evaluation Methods [HMPR04]

4 FutureID Approach

FutureID has dedicated a significant effort to evaluate its results in a rigorous manner (e.g. Test beds for the Pilots, Evaluation WP). To provide an overhead of results, FutureID has dedicated part of the Sub-Project Transfer tasks to Consolidation and Evaluation. The work packages dedicated to these tasks present the 'big picture' of the major results for the entire large-scale project of FutureID. The consolidated view forms the basis to give a systematic evaluation. The FutureID Evaluation approach using a Design Science Methodology framework has been a valuable tool in organizing and harmonizing multiple disciplinary evaluation approaches.

To provide a closer look, FutureID has simplified its evaluation process into three easy steps. First, they identify each of the Artifacts, which in their case are two pilots, a reference architecture, and implementation. Second, they clarify where each interdisciplinary team considers the artifacts and develop requirements regarding their disciplinary. This step is ranked regarding importance and is utilized by using the Evaluation Wiki Tool. Lastly, they Re-evaluate, which is when each requirement identified will be reevaluated on whether they should be really implemented or initiated in each artifact. Of course, with the complexity of some of the artifacts a noble evaluation could not be sufficiently executed with just this process, therefore, FutureID has used extra evaluation steps to properly consider specific needs of some of the artifacts. For example, they have used testbeds in grasping a better outlook of the pilots. The Evaluation Wiki tool is a quality control mechanism that has been used for the core evaluation of FutureIDs results. It has a variety of different beneficial functions that lead to a practical and more optimal evaluation method. On the practical side, it presents an easy to read and adjust, while still being a comprehensive solution for documentation of the evaluation requirements needed for each artifact. Each artifact can be sub categorized into viewing each of the importance levels of requirements (must, should, may, all) on the main page of the tool. It classifies each requirement, from which interdisciplinary team it's from, comment section, and its rank of importance. While collaborating with multiple disciplines, harmonizing and consolidating a wide spectrum of requirements proved to have some difficulties and major conflicts. In order to resolve this problem, FutureID included another addition to the Evaluation Wiki tool and to the Evaluation work package. The additions was an added deliverable that focused on the clarification of which requirements are either similar to, relates to, or conflicts with other requirements. This is a necessary task that all large scale interdisciplinary projects should have in harmonizing requirements in evaluations. This task helped provide insight on how all of the requirements can cooperate and be applied all together. In addition to these processes, the testbed has proven to be a great technical method in testing the implementation and pilot applications. It is built of three different levels of testing; unit testing, integration testing, and system testing. The implementation artifact is tested using the unit, integration, and system testing. While, the pilots are tested on only the system level testing, the form of evaluation methods between different artifacts obviously varies. However, the Design Science Evaluation methods are broad enough to cover a wide range of techniques.

5 First Results and Limitations

As FutureID is an ongoing project, this section will elaborate on first results in FutureID and limitations. Until now, the requirements have been formed and harmonized for the evaluation of all of the FutureID artifacts. The advantages could be seen as premature, but as the Design Science Methodology framework has provided mostly positive feedback in research, the outcomes are promising. Overall, this could be seen as one of the main limitations presented in this in-progress paper and application, even though until now there has been promising first results. Continuing, FutureID has already gain first results on the Reference Architecture, which has provided encouraging results. As a way of evaluation, each discipline represented in FutureID established requirements that should be met for each artifact. The Reference Architecture passed all of the requirements in all 'must, may, should' categories regarding the Socio-Economic Requirements. Regarding the Technical Requirements, it also passed 92 % of the 'must, may, should' categories for both the Reference Architecture and the Implementation artifacts. Overall, most of the disciplines displayed similar positive remarks regarding the application of requirements. Even though FutureID is currently in the stage of concluding the evaluation of both the Implementation artifact and the Pilots, it can be foreseen that similar positive results are also to emerge.

6 Conclusion

This research in-progress paper discussed the need for technical projects to focus on multiple disciplines in order to be more inviting to the market. Further, the paper takes a practical focus and goes into detail how the project FutureID has applied a Design Science Evaluation approach to better evaluate, re-evaluate, and harmonize the needs and demands of different disciplines and different perspectives. As the project and this paper are still in progress, only first results were able to be presented. However, FutureID will be concluding its work by fall of 2015, where larger results of this inter-disciplinary evaluation application can be seen and interpreted.

References

[GORR04] Greenwald S, Olthoff K, Raskin V, Ruch W. The user non-acceptance paradigm: INFOSEC's dirty little secret. Proceedings of the 2004 workshop on New security paradigms. ACM, Nova Scotia, Canada 2004. p35-43

[HMPR04] Hevner A, March S, Park J, Ram S. Design science in information systems research, MIS Quarterly 2004. p 75-105.

[FFSS02] Feigenbaum J, Freedman M, Sander T, Shostack A. Economic barriers to the deployment of existing privacy technologies (position paper). Proceedings of the Workshop on Economics of Information Security; 2002.

[Fu14] FutureID- Shaping the Future of Electronic Identity, [Internet]. 2014. Available: http://futureid.eu/.

[NCP91] Nunamaker J, Chen M, Purdin T. Systems development in information systems research. Journal of Management Information, 1991.

[MMG02] Markus ML, Majchrzak A, Gasser L. A design theory for systems that support emergent knowledge processes, Mis Quarterly, 2002. p. 179–212.

[Ro06] Roßnagel, H. On Diffusion and Confusion – Why Electronic Signatures Have Failed. Trust and Privacy in Digital Business; 2006. p. 71-80

[WSE09] Watts S, Shankaranarayanan G, Even A. Data quality assessment in context: A cognitive perspective. Decis Support Syst., 2009. p. 202–211.

[ZR12] Zibuschka J, Roßnagel H. A Structured Approach to the Design of Viable Security Systems. ISSE 2011- S Wiesbaden: Vieweg+ Teubner; 2012. p.246-55.

[ZRMS11] Zibuschka, J., Roßnagel, H., Muntermann, J. und Scherner, T. Mobile Emergency Management Services Targeting Large Public Events. International Journal of Service Science, Management, Engineering, and Technology (IJSSMET). 2011.