Suitability of Design Science Research - Demonstrated on a Data Quality Project in Ubiquitous Computing

Plamen Petkov[1], Markus Helfert[1]

[1] Dublin City University, Glassnevin, Dublin 9, Ireland
{petkovp, markus.helfert}@computing.dcu.ie

Accepted to the 3rd International workshop on IT Artefact Design & Workpractice Improvement, 2 June, 2014, Friedrichshafen

Abstract.
Design Science research, at its present, provides relatively coherent phases for conducting information systems' research, however its application to Data Quality research is still limited. In this paper, we apply Design Science to a Data Quality Project and introduce an extension to Design Science research in terms of techniques to create artifacts. The artifact in this research is a process for assessing Data Quality in Ubiquitous Computing. In this domain context, Design Science provides relevant knowledge based on domain-specific concepts and relations. Particularly, this is achieved by literature review, collaboration with practitioners, information and process synthesis. The contribution of this paper is an application of Design Science to the domain of Data Quality and a further clarification of its process. We conclude that techniques are appropriate for designing artifacts in the Data Quality field.

Keywords: *Design Science Research, Data Quality, Ubiquitous Computing, Process, Information synthesis, Service Oriented Architectures*

1 Introduction

Over the past decade, the service oriented approach has become a preferable way of building information systems. By assembling functional components into a network of loosely coupled services, flexible and dynamic environments are composed. Following the service oriented paradigm [1], a type of such environments are Ubiquitous spaces. In a high level sense, ubiquitous spaces can be seen as abstract logical environments where distributed devices provide complex functionality that users claim to consume [2]. Far away from the traditional interaction between users and devices, Ubiquitous Computing (UC) reaches to deploy complex proactive behavior by means context awareness [3]. However, in more complex compositions, loosely coupled data and dynamic behavior of the environment can become a source of serious problems. These problems can spread throughout whole composition and in this way make them difficult to handle. Moreover, in some cases they can cause serious

damage which can result even in casualties. So, the question here is how we ensure the quality of the data within these environments? On the highest level, the solution of this problem may require incorporating a correct research methodology in order to deliver reliable results. But, then, which methodology is suitable in this particular case and why?

A suitable research methodology, namely Design Science, has been identified among other candidates to support our research. In the past 15 years, DS research methodology has received an increased attention in computing and Information Systems (IS) research [4]. It has become an established approach for conducting research in the IS field, with remarkable growth in related literature [5]. However, at its current stage, it does not offer consistent and comprehending phases, which will guide researchers in their choice of techniques [6]. This paper presents a suitable research methodology, based on the well-known design science principles [7], which covers techniques to shape and develop artifacts in the area of Data Quality (DQ). The artifact in this paper is a process for assessing DQ in Ubiquitous Computing (UC). The objective of this paper is **not** to show **what** the artifact does **but** rather to demonstrate **how** it was constructed using and extending the DS principles.

The remainder of the paper is organized as follows: the next section we present a discussion of the suitability of the Design Science in this research. In section, we also present an overview of the accommodated DS model for this research topic. Then, in section 3, we explain in details how we composed DQ process within Ubiquitous Computing field and we present the research techniques we used to build/evaluate the solution. Then the paper is completed with discussions, concluding remarks and outlines the opportunities for future work. This paper helps to define future directions and phases of design science methodology within the spectrum of Data Quality and Ubiquitous computing.

2 Selecting a Suitable Research Methodology to conduct a Data Quality research

The Objective of this section is to describe a suitable research methodology in order to provide a prescriptive knowledge in terms of creating a DQ process that later is accommodated within a ubiquitous computing case.

In this study, we turn to the literature for advice at almost every step of our research. An important stage was to identify suitable research methodology. Our observation in literature provided us a list of available research methodologies (e.g Action Research [11], Ontology Engineering [12], Method Engineering [13], etc.) that might beneficial for this type of project. However, none of them was likely to match this research case. Action Research, for example, focuses on solving a socio-technical problem by developing new solution and evaluating it in an organizational context. This type of research is often biased or influenced by the organization that it has been conducted within. Another example – Ontology Engineering – it provides a common vocabulary of an area and defines, with different levels of formality, the meaning of

the terms and the relationships between them. Nevertheless, the outcome of that type of research can hardly deliver a process, which eliminates it as core of our case. Despite this fact, some of its techniques would be still used in our research. Method engineering approach might be seen as the most suitable one as it offers a solution as a conjunction of available methods. However, as a research process, it does not take into account the impact from engaging various practitioners that may provide valuable insights to the required method.

A suitable candidate that combines features of previous the commodities is the Design Science research methodology. Design science is a methodology for generating knowledge – descriptive and prescriptive to solve problems that stem from both literature and practice and are evaluated trough collaborating of academia and industries.

In the past 15 years, DS research methodology has received an increased attention in computing and Information Systems (IS) research [4]. Views and recommendations on the DS methodology vary among papers, e.g. ([14] [15]). One of the main issues with DS is that guidelines provided from the precursors [9] are seldom 'applied', suggesting that existing methodology is insufficiently clear, or inadequately operationalized - still too high level of abstraction [15]. Descriptions of activities (procedures, tools, techniques) that are needed to follow the methodology are only briefly indicated. Thus we are going to extend the native DS methodology by specifying these tools and techniques in the DQ context.

In this research case, the artifact is a process for assessing and analysing the output data quality within the ubiquitous computing environment. As it was mentioned, we are going to clarify the well-known DS approach by specifying techniques used to find, solve and evaluate the particular problem. The clarification described below is based on the process-oriented reference model described by [17] It was introduced as a part of the DS methodology for researchers aiming at process construction. Fig. 1 provides an overview of research methodology in the Design Science settings.

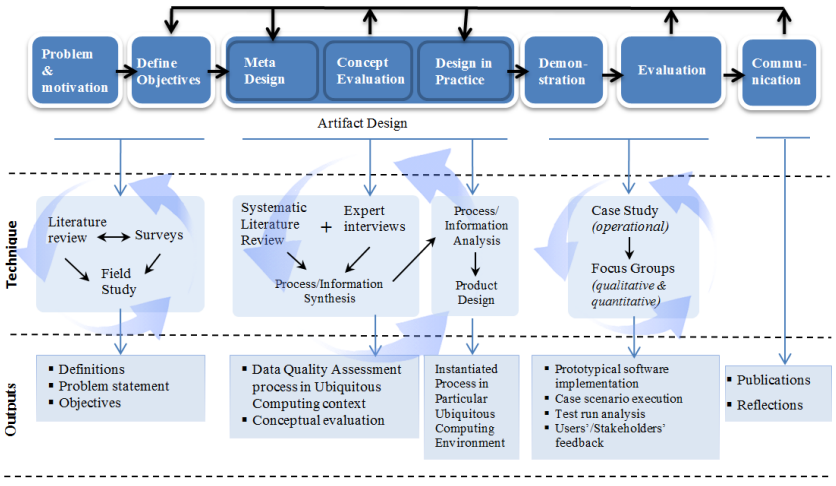

Fig. 1. Design Science Process in DQ research

Furthermore, it demonstrates the relevant techniques and methods selected to accompany the DS process. Also, an overview of the expected outputs and research questions are presented along with the process.

The following illustration (fig. 1) describes the Design Science Reference model steps applied to this research project. The first step includes problem identification and motivation (1) followed by defined objectives (2). Next, is the phase of designing and developing the artefact (3). This step consists of three sub-steps: Meta Design, Concept Evaluation and Design in Practice. To evaluate the artifact demonstration (4) and evaluation stage (5) are followed. The last step is the communication step, which links all previous stages through a set of publications (6). The following section will provide further explanation about each step applied to this research case.

3 DSR in Data Quality: Analyzing and Measuring Data Quality in Ubiquitous Computing

The output of this section is to portray the successful usage of the adapted DSR methodology outlined in Section 2. The following research is conducted by a research group that included representatives from the academic field mainly from Dublin City University, Ireland, in collaboration with University of Granada, and was focused on the incorporating the Data Quality assessment method within Ubiquitous Computing Context [18]. Before we continue any further with explaining the research methodology, we will outline the research context.

The objective in this research case is to investigate how to analyse and measure data quality within Ubiquitous Computing context. In order to reach given objective, assessment process is designed and evaluated. More specifically, this process is composed by few simpler processes that take into consideration ubiquitous computing principles. The final output is guide for assessing DQ in UC and it is, ultimately, implemented as a software tool in a particular case. An overall picture of the research context is presented on fig 2. Further aspects on the content of the artefact were discussed in [18] [10] [19].

(1) Research objective and (2) motivation. The problem statement and motivation were stated in this stage. This may be seen as preparation, gathering knowledge, or building foundation on which the artifact is being constructed [7]. The latter was defined by obtaining information mainly from the literature. Systematic Literature Review (SLR) [8] was chosen in order to get accurate results.

The ultimate goal of the literature review is to identify activities from the found materials. Each activity should be accompanied with a meaningful description and rational of selection. It is a good routine to keep the search for materials transparent and replicable as much as possible. In this research case the main goal here is to extract all the relevant concepts that defines Data Quality, and which are believed to be its enabler factors. Moreover, it aims to identify possible approaches to detect and analyze DQ.

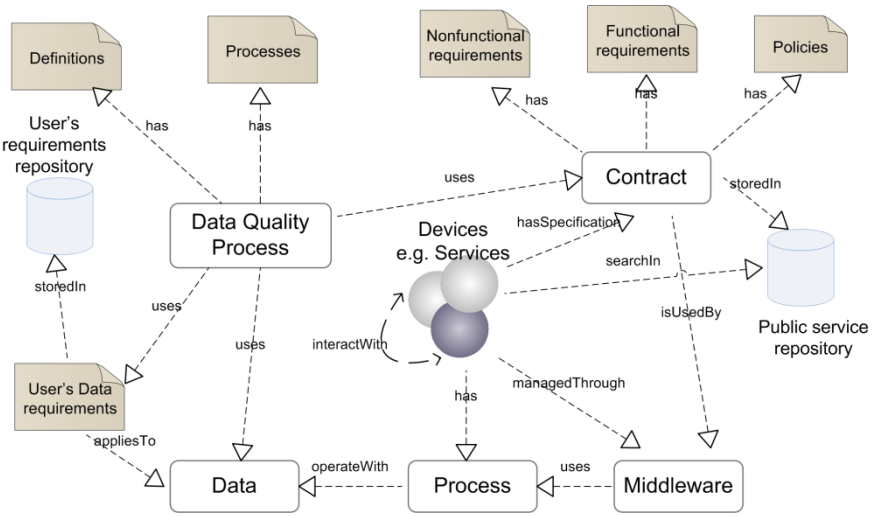

Fig. 2. DQ Process within UC context

Basically literature review aims to answer the research question defined during the scoping phase. Next paragraph aims to explain how the SLR was conducted. Figure 3 illustrates the process that was follow to conduct the literature review.

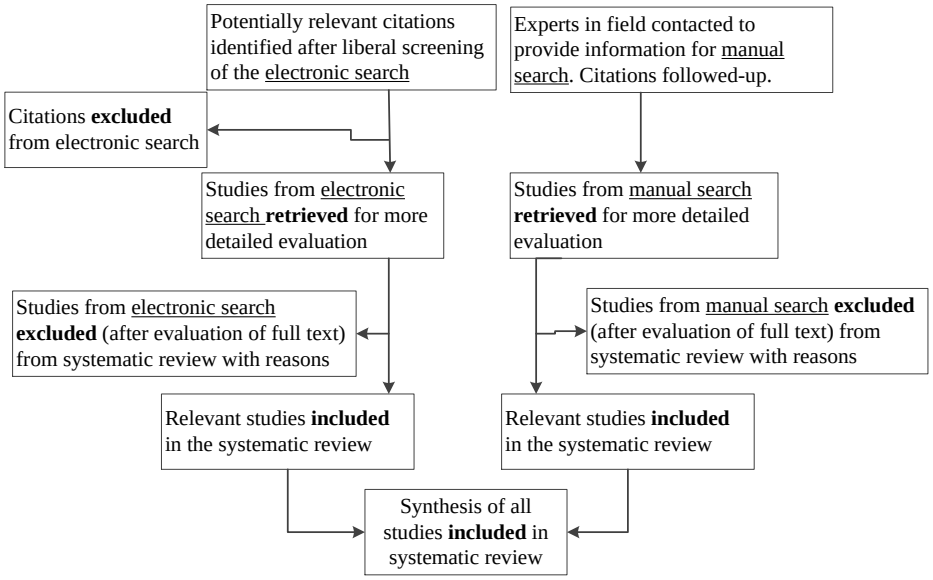

Fig. 3. Flow Diagram of study selection procedure [9]

In this paper the SLR is into two activities - broad and advanced literature search.

The broad search concentrates on finding research relevant materials. The main focus is put on reading abstracts, conclusions, prefaces, and references of the found materials in order to collect as much potential leads to relevant information as possible. The rigor of the systematic search process is one factor that distinguishes this approach from others. It is iterative and benefits from identification of existing systematic reviews, assessing the volume of potentially relevant materials, and using various combinations of search terms derived from the initial scope.

In this research context the target of the broad search was to gather as much possible information about problems related with DQ. Then we referred these findings to the domain of Ubiquitous Computing. The types of literature that were inspected were consisting mainly of periodic issues, reports, journals as well as professional blogs and forums over the Internet. Some main keywords that were used to perform searching were “ubiquitous computing”, “data quality”, “data issues in ubiquitous computing”, etc. At that point 241 were articles were selected as contacting potential knowledge to support the solution.

Synthesis of Studies(figure 3, bottom block) is an important block of the process of conducting systematic literature review. In this stage we used revised literature, structured by topics, to derive with conclusions about the artifact or use it to derive new content that will positively contribute to the solution. . The result of this investigation was the development of a matrix table. This table has been previously publish in [10] and contains different criteria in rows and associated solutions in columns. Its development allowed us to have a complete overview of current research development in the area of DQ. In conclusion, resulting from the analysis made about current DQ approaches, and the ubiquitous computing profile build using the advance literature search, allowed us to ultimately synthesize a list with requirements for applicable DQ assessment process within Ubiquitous Spaces.

The advanced search focuses on analysing and assessing the actual relevance of the found materials. Main and secondary objectives along with the exclusion criteria from the initial scope are intended to identify those materials that provide direct evidence about solution for the domain. In order to reduce the likelihood of bias, these criteria may be refined during the search process. Moreover, the quality of those materials is assessed. The aims are to weight the importance of individual information when retrieving particular activities for the desired process; to lead the interpretation of findings and determine the strengths of results; to provide recommendations for further research.

In addition to the findings in the literature, an independent survey was been conducted on the data quality issues in ubiquitous computing. Targets of this survey were practitioners developing solutions in the Ubiquitous Computing domain. 34 people were surveyed of which 6 projects coordinators, 10 system architects and 18 software developers. Questions in the survey were determined based on findings in the conducted literature review. Results support the problem identified from the literature and reported that the 75% of surveyed companies underestimate the problem of the poor semantic data.

The output of this step was to have a complete list of the literature that is considered for the review. In our case we went through it twice. First we narrowed down the total amount of articles to 39 by reading the abstracts of all the papers we had previously collected. Then, we carefully read all the papers. In particular, we verified their consistency with the research questions. In this study, the research questions are RQ1: "What are the requirements needed for applicable Data Quality assessment process in the ubiquitous context?" and RQ2: "How to incorporate Data Quality assessment process within Ubiquitous Spaces?".

Based on the results from the literature and conducted survey, it was also observed that in complex ubiquitous environment data quality problems are persistent and information flow can be difficult to handle without having awareness of the data. Hence, it was set as the research objective discovering a process that will assess data quality in ubiquitous context.

(3) Design phase. In the design and development phase the artefact (RQ2) was presented in the form of a complex process (set of processes) for defining, detecting and analysing poor semantic data within UC context. The process of designing the artefact was explained in [10]. To build the process in the Meta-phase, literature was used as main reference to extracted knowledge. Extracted knowledge was in form of principles (DQ principles [20]), processes (e.g. different data assessment process) and information models (e.g. various XML models, etc.)

Following the model on Figure 3 professional feedback was gathered in forms of individual interviews. Specifically each interview was as carried out as a short **discussion**. The duration of the discussions were approximately 40 minutes of which 20 minutes were devoted to familiarizing the expert bodies with the topic. Five experts' opinions were gathered so far with the intention to gather more as the artefact improves over time following the DS methodology cycle. Knowledge obtained from discussions was structured in a way of notes and tables in different subtopics. As recurrent process my plan is to continue gathering information from practitioners as it will expectantly improve the quality of the artefact.

The actual designing and developing of the artifact is done in the "information/process synthesis" stage. This is the phase, where obtained knowledge from the systematic literature review and discussions with practitioners in a form of methodologies, principles and semantic constraints of concepts (e.g DQM [14]) is fused. The result is presented in [22] using BPMN. This process can be referred to the Method engineering approach where the solution is conjunction of few available methods. At this stage some conceptual evaluations were made. A group of experts were asked to assess the artifact against the conceptual criteria stated in section 2, as well as general process quality factors accommodated [15]. A set of questions, based on the categories in the logic model of the artifact, was sent to each of the participants in the group. Although the results provided by using this framework are fairly subjective, it provides us with valuable feedback about the quality and the relevance of the concept. This allows us to further refine it before it has been applied and implemented in the real world. The evaluation framework presented in [16] shows that developed DQ methods mostly meet the conceptual requirements for quality process and thus

they will provide with promising output results. However, at this stage, some aspects of the conceptual evaluation are still on the go.

The "design in practice" is important step of the design process as this is the moment where a prototype, in form of software solution, is designed. This can be seen as an instantiation of the artefact. As it was mentioned, DS considers collaboration with practitioners as vital part of the research progress. In this case, the collaborator in not an industrial partner but rather academic. University of Granada is working on a project called Dynamic Open Home Automation (DOHA), which incorporates all Ubiquitous Computing principles [18]. DOHA is a SOA-based platform for the access, control and management of home-automated systems, composed of a set of lightweight and independent services (devices). In this case, the application of the artefact is a software extension application to DOHA middleware. This application is built by following the DQ process developed in the previous substrate

(4) Demonstration and (5) evaluation. Considering the nature of the research artifact, a case study method has been selected to demonstrate its utility. Case study is an empirical inquiry that investigates a phenomenon within its real-life context [24]. To objectively evaluate the DQ process (new software component) in an operational manner, data collection from executing different scenarios will be done following the general principles of conducting case studies. The output of this stage is prototypical software tool that will be evaluated trough execution of different case scenarios.

4 Conclusions

The relevance and suitability of Design Science as a research methodology has been argued in many publications. This article expounds the suitability of the generic Design Science Research methodology to the IS research in the context of Data Quality.

Systematic literature review, collaboration with practitioners and process synthesis were identified to play an important role in producing solutions to the Data Quality research in the Ubiquitous Computing context. Based on these three activities, we introduced and described an extended, more detailed model of the well-known DS research. Using this model, we were able to demonstrate the creation of a DQ process helps identifying faulty information within ubiquitous spaces. Moreover, we refined and evaluated this artifact thanks to the collaboration with our university partner.

Currently, the evaluation our artifact (DQ process) is based on only few case scenarios that we have conducted at the University of Granada. Therefore, in future, we plan to enlarge the scale of evaluation process, by extending the numbers of case scenarios. Another line of future investigation is to go beyond the practical evaluation, and to also consider reactions from the practitioners. Our expectations are that their feedback will provide us with valuable outcome that will contribute to grounding and consolidating the process for assessing data quality in ubiquitous computing.

The contribution of this paper is an application of Design Science to the domain of Data Quality and a further clarification of its process. This contribution, in turn, makes the design process more efficient by means of increased applicability. Moreo-

ver presented research methodology in this study would supply a base guideline to conduct similar research – research in the Data Quality domain.

Acknowledgements

This work was supported by the Irish Research Council (IRC) under the Postgraduate Scholarship Scheme. This work was also carried out in collaboration with the University of Granada.

References

1. Papazoglou, M., Willem-Jan, v.: Service oriented architectures: approaches, technologies., 389-415 (2007)
2. Weiser, M.: The computer for the 21st century., p.94–104. (1991)
3. Dey, A.: Understanding and using context. In : Personal and ubiquitous computing 5, p.4–7 (2001)
4. Kuechler, B., Vaishnavi, V.: On Theory Development in Design Science Research: Anatomy of a Research Project. In : European Journal of Information Systems, 17(5), pp.489-504 (2008)
5. Carlsson, S.: Socio-technical IS design science research: developing design theory for IS integration management. In : Information Systems and E-Business Management, 9, pp.109-131. (2011)
6. Alturki, A. .: A Design Science Research Roadmap.. In : Proceedings of the design science research in information systems and technology , Heidelberg (2011)
7. Ostrowski, L.: Detailed Design Science Research and its Impact on the Quality of Design Artefact.. In : Proceedings of the European Design Science Symposium, Leixlip (2011)
8. Okoli, C., Schabram, K.: A Guide to Conducting a Systematic Literature Review of Information Systems Research.
9. Egan, M., Petticrew, M., Ogilvie, D., Hamilton, V.: The Health and Social Impacts of Opening New Road. Protocol for Systematic Review, University of Glasgow, Glasgow (2001)
10. Petkov, P. .: A Methodology for Analyzing and Measuring Semantic Data Quality in Service Oriented Architectures. In : CompSysTech, Ruse, Bulgaria (2013)
11. Cassel, C., Johnson, P.: Action Research: Explaining the Diversity.. In : Human Relations 59(6) , pp.783-814 (2006)
12. Mizoguchi, R.: Tutorial on Ontological Engineering., 363-384 (2003)

13. Athena Consortium: Method Engeneering.. (2006)
14. Baskerville, R., Pries-Heje, J., Venable, J.: Soft Design Science Methodology. In : Proceedings of design science research in information systems and technology, Malvern (2009)
15. Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. (2007)
16. Hevner, A. R., March, S. T., Park, J.: Design Research in Information Systems Research. MIS Quarterly (2004)
17. Ostrowski, L., Helfert, M.: Reference Model in Design Science Research to gather and Model Information. AMCIS 2012 Proceedings. (July 29, 2012)
18. Petkov, P., Helfert, M.: BPEL Aided Framework for Constructing Monitoring Rules in Service Oriented Architectures. In : 9th International Conference on Web Information Systems and Technologies, Aachen (2012)
19. Rodríguez-Valenzuela, S., Holgado-Terriza, J.: Services composition model for home-automation peer-to-peer pervasive computing,. In : Proceedings of the Federated Conference on Computer Science and Information Systems, p.529–536 (2011)
20. Oracle, I.: Understanding Data Quality Management. In : Oracle® Warehouse Builder User's Guide. Oracle Inc. (2008)
21. Moody, D. L., Shanks, G. G.: Improving the quality of data models: empirical validation of a quality management framework. Information Systems(28(6)), 619-650 (2003)
22. Petkov, P., Rodríguez-Valenzuela, S.: Modeling Context-Awareness in a Pervasive Computing Middleware using Ontologies and Data Quality Profiles. In : PerCAM 2013 : 3rd International Workshop on Pervasive and Context-Aware Middleware, Dublin (2013)
23. Yin, R. K.: The Case Study Crisis: Some Answers.. Administrative Science Quarterly,(26(1)), 58-65. (1981)
24. Carlsson, S., Henningsson, S., Hrastinski, S., Keller, C.: Socio-technical IS design science research:. (2011)
25. Batini, C.: Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Application)). Springer, Berlin, Germany (1998)