Extreme Requirements Elicitation: Lessons Learnt from the COVID-19 Case Study

Anna Bernasconi

Dipartimento di Elettronica, Informatione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133, Milano, Italy

Abstract

[Context and motivation] The outbreak of the COVID-19 pandemic has highlighted the general unpreparedness of the research community to face biological data challenges in a timely manner. In these dramatic times, healthcare providers and virology researchers would have largely benefited from sound support from the information systems and database communities, able to produce systems that allow to explore data, compute statistics, and discover knowledge on the virus. **[Question/problem]** The research progress in emergency times is hindered by the lack of experience in organizing effective interaction between parties, possibly supported by specific tools. Streamlining the domain knowledge exchange through an appropriate requirement engineering practice is of paramount importance, not only in health crises but also in disaster management and cases of extreme weather hazards. **[Principal ideas/results]** To this end, we report on the lessons learnt by applying an agile interviewing method during the COVID-19 outbreak, involving several experts spanning from clinicians and geneticists, to biologists and virologists. We have preliminarily tested our approach for the creation of an integrative database and a search system for viral sequences; foreseen extensions of the system target visualization, question-driven data warehousing, and statistical reporting. **[Contribution]** We present a number of practical suggestions aimed at conducting interviews under extreme conditions, where the conventional requirement elicitation methods are not applicable. Specifically, our proposal lays the foundations to i) address the complexity of Extreme Requirements Elicitation (ERE) sessions in the context of interdisciplinary fields; ii) design timeliness-critical systems, essential to fight pandemics.

Keywords
Domain Knowledge, Extreme Requirements Elicitation, Interdisciplinarity, Virology, COVID-19

1. Introduction

In March 2020, the world was hit by the COVID-19 epidemic, shaking the foundations of healthcare and medical research. Virologists were confronted with an unknown virus – showing extraordinary transmissibility – and clinicians with the related disease, causing symptoms of different severity in humans, based on yet obscure rules. Several data initiatives around the world started collecting information on virus sequences and clinical parameters of COVID-19 patients [1]. Motivated by the general call to arms against the COVID-19 pandemic and by the previous knowledge in human genomics [2], the GeCo research group (www.bioinformatics.

✉ anna.bernasconi@polimi.it (A. Bernasconi)

🆔 0000-0001-8016-5750 (A. Bernasconi)

deib.polimi.it/geco/) engaged in producing systems to aid virology research. Within the project, the author of the paper has contributed defining the objectives and orchestrating a series of interviews to quickly approach the domain of viral data, in the role of *requirements analysis designer*.

Essentially, our team was able to build a powerful search engine, ViruSurf, http://gmql.eu/virusurf/ [3], in the short time frame spanning from April to July, 2020. The work, as described in Section 2, proceeded continuously, in an agile setting that alternated data and software design sessions with interviews to experts in the domain, to understand their requirements in what we call an *Extreme Requirements Elicitation* (ERE) process. In Section 3, the lessons learnt from this experience are reported, characterized by the urgency of the pandemics and the domain knowledge gap with the virology experts, addressed with a "show, don't tell" approach to transfer our need of understanding. ViruSurf is just an intermediate result of a long agenda of ongoing and planned research tasks, targeting not only virologists, but also clinicians, for empowering research in clinics, hospitals, and intensive animal farmings (that are also subject to viral epidemics). Section 4 sketches the plan to continue the elicitation process from here on, by abstracting the encountered challenges in the ViruSurf case, to systematize it for future application; finally, conclusions are drawn in Section 5.

2. Our Experience from ViruSurf

At the beginning of the pandemic, the most important databases for viral sequences (https://www.gisaid.org/; https://www.ncbi.nlm.nih.gov/genbank/; https://www.cogconsortium.uk/) started collecting and making available to the public sequences for SARS-CoV-2, the virus responsible for the COVID-19 disease. We engaged in the challenge of making such open data integrated and useful to the research community. First, we represented the domain of viral sequences (i.e., strings of A, C, T, G nucleotides with describing metadata) with a conceptual model [4], then we employed it to drive an integration pipeline and the computation of homogeneous nucleotide/amino acid variants associated to their impacts at the protein, phenotype, and disease levels. The built database is queried by the ViruSurf search engine [3], our initial attempt to enable virologists' data exploration for biological discovery; it allows to express complex queries in a simple way, matching sequences with combined conditions on attributes and variants.

The ViruSurf development team included six members – holding mixed expertise in Computer Science and Bioinformatics – led by a Principal Investigator, two senior researchers, and three PhD students (including the author of this paper). Given the author's background in interviewing experts for understanding requirements of human genomics exploratory software – during collaborations with the European Institute of Oncology (www.ieo.it) and the Italian Foundation for Cancer Research (www.ifom.eu) – and in evaluating their usability with empirical studies [5], she has been responsible for the requirement elicitation design and the direction of the methodological framework.

To inform the conceptual model and ViruSurf design, twenty researchers were interviewed, knowledgeable in fields, related to the virus from different points of view: the specific mechanisms of its biology, the pathogenesis, its interaction with human/animal hosts. Interviewees

ID	#Interv.	Title/Expertise Level	Research Area/Specialization	Location	Institution Type
I1	3	Associate Professors/Clinicians + Researcher	Emergency medicine/Hepatology	IT	Hospital
I2	1	Full Professor	Virology	US	University
I3	1	Researcher	Molecular Biology	IT	University
I4	1	R&D Manager	Bioinformatics	IT	Private Company
I5	1	Full Professor	Molecular Microiology	US	University
I6	1	Senior Researcher	Epidemiology	UK	Research Center
I7	2	Researchers	Veterinary Virology/Veterinary Medicine	IT	Research Center
I8	1	Researcher	Bioinformatics	US	Private Company
I9	1	Researcher	Veterinary Virology	IT	Research Center
I10	2	Full Professor + Associate Professor	Medical Genetics	IT	University/Hospital
I11	1	Researcher	Virology	US	University
I12	1	Research Associate	Medical Genetics	BE	University/Hospital
I13	2	Full Professor + Researcher	Molecular Biology	IT	University
I14	1	Full Professor	Molecular Virology	HK	University
I15	1	Full Professor	Computational Biology	SI	University

Table 1
List of interviewees groups with their characterization.

had different levels of expertise and were sourced from public/private institutions (Table 2). All participants were interviewed individually except for I7, I10, I13 (in pairs) and I1 (in three).

We conducted lightly structured interviews; the typical setting included an introductory 1.5 hours video call session where the participants were interviewed jointly by the Principal Investigator, the requirements analysis designer, and other two team members, supported by slide presentations. All such sessions were approximately divided in three parts of 30 minutes each. First, we explained our interest in the topic, showing the progress of our systems. Then, we made questions to gather participants' attention towards: In some cases, participants showed particular interest in receiving follow ups, offering their availability for successive sessions, organized with a more "hands-on" angle. In particular, with I2, I3, I7, I8, and I11 we were able to perform other two sessions of 1 hour each, joined by other members of the development team: (1) a "mock-up session", using prototypes of the system or slides with simulated user workflows, allowing a "satisfied/non-satisfied" answer from participants, with an open discussion to formulate variations with respect to the proposed design; (2) a "demonstration session" for the implemented features, with the presentation of small application use case, and finally request for feedback on the compliance with previously formulated requirements and on the usefulness of the functionality in real case scenarios.

3. Lessons Learnt

Our exercise did not follow predefined requirement elicitation frameworks [6], due to initial lack of expertise about RE techniques in the team. However, it was of great importance to learn a series of lessons: in spite of our expertise in developing software for genomics, we were lacking a substantial amount of domain knowledge; moreover, requirements were collected rapidly and in emergency, to build a tool whose power we only fully understood while collecting information from experts. We summarize the main understandings in the following.

Deal with diversity. Domain-knowledge understanding has been addressed in general scenarios [7] and in contexts specifically related to bionformatics [8, 9] or genomics [10]. However, the domain experts involved in COVID-19 research encompass an unprecedented plethora of different specializations: biologists, clinicians, geneticists, virologists – focusing on philogenesis, epidemiology, veterinary, etc. It is important to capitalize on the diversity of interviewed people

and also on the fact that a similar diversity is present in the many stakeholders who could take advantage of our system. Some features may, for example, be useful to both virologists and clinicians: however, each single functionality is described by experts in different ways, and each type of expert must be approached differently. Diversity characterizes not only disciplines of specialization but also the level of expertise and personal talents (e.g., how to address young researchers and senior professors). For example, during interviews we observed that young researchers appreciated prototypes/mockups-driven presentations (allowing them to gain insight in the details of the user interaction and implementation), while professors appreciated high-level presentations, as they were able to highlight completely new directions from more abstract descriptions.

Investing in short, just-in-time pre-interview meetings. While requirements engineering usually features a long and controlled process of steps – including the interpretation, analysis, modeling, and validation of requirements' completeness and correctness [11] – we argue that, in the specific setting described in this work, exceptions should be allowed. Domain experts have limited time available; they allow it for interviews if they see potential in the proposal, so interaction should be as productive as possible. For successfully conducting the interviews, we understood that organizing quick preparatory meetings preceding the call was essential to: i) refresh the foundational aspects of the specific research expertise of the person to be interviewed[1]; ii) discuss it within the team; iii) quickly agree a clear scheme for the interview; and iv) identify/plan roles for each member of the team.

Being use-case driven. When speaking different domain languages, one quite successful approach is that of "show, don't tell" [12] (typical of design thinking mindsets [13]), e.g., to bridge the knowledge gap between computer scientists and virologists. Interactions should be driven by specific problems: proposing examples, small prototypes, showcasing alternatives, and mimicking a live interface – by using slides with animations on specific modules and objects – led to promising results of transferring our need of knowing or understanding.

Being curious. Using all possible means to facilitate discovery. While collecting requirements, we were continuously reasoning about *new resources* and *use cases* that could broaden our views; this has produced new interesting data sources and data analysis methods, not envisioned at the beginning of the project, now integrated within our data ingestion and curation pipelines. In parallel, we activated a continuous monitoring of literature and news of interesting facts, so as to enrich our interviews with every novel aspect that could create stimulating interaction niches (e.g., the emergence of a given important mutation, the publishing of a study with broad societal impact...). Finding *new experts* was essential: in our case next contacts were often suggested during previous interviews.

Being aware that a second opportunity of interaction is worthwhile. A challenging aspect was to identify the right moment in which developments had made a big enough step for the next interview: a second round of discussion with the stakeholders was used to show progress, ask feedback, and proceed with following requirements, embracing a similar approach to the one of agile requirements analysis [14].

---

[1]While in some cases prior domain knowledge may introduce negative bias in interviews [7], in this context requirements analysts must have (flexible) experience in the domain; note that most of the interviewed experts interacted under the implicit assumption that explanations of basic concepts of their discipline was not needed.

4. Research Agenda

The research proposed in this preview ultimately aims to develop a systematic interviewing method driven by three principles: i) overcoming strong initial differences in the understanding of a given domain; ii) converging towards a solution in a short time; iii) conducting ERE sessions in parallel with design and development, neglecting the typical precedence between phases and enforcing cycles of ERE sessions that instrument design and vice versa. After experimenting these concepts during the first COVID-19 pandemic wave for ViruSurf, in our agenda we are planning to produce a tool suite for supporting virology research. We envision many significant additions, including the support for: queries for testing the stability of viral regions (useful for vaccine design), emerging knowledge about the mutations' impact, effective visualization for data analysis, and ad-hoc services for covering needs of hospitals and other parties who are not interested in publishing their sequences.

To support all these directions, the ERE design shall pursue several research questions: **RQ1.** *How could we train groups of system designers and developers to manage extreme requirements for virology-related software?* Scenarios which are driven by urgency can be simulated or reproduced by means of training exercises, where urgency is artificially created by simulating a disaster and then instructors play the role of domain experts, also setting artificial goals that represent the need for rushing towards the definition of requirements, with strict submission deadlines and fixed time schedules. **RQ2.** *Could it be worth applying the proposed method – conceived to manage developing knowledge in a biology interdisciplinary setting – within a known disciplinary context?* Out first ERE exercise aimed at covering two kinds of emergencies, besides urgency we were also driven by the need of dynamically creating enough knowledge within the group in order to manage the forthcoming interviews one after the other. We should conceive a structured ERE method for collecting requirements within a known domain, shared between designers and interviewed experts, and compare its performance with a structured and conventional plan of interviews. **RQ3.** *How will we diversify our portfolio of ERE use cases in the forthcoming design tasks?* After the lockdown period we managed to hire new people in the group, including a biologist, a communication designer and a UI expert. We will test if our performance in dealing with extreme requirements will increase or decrease in the context of this new group.

5. Conclusion

In this research preview we have described the experience developed in gathering information from life science practitioners to build timeliness-critical software. We experimented ERE while we designed and developed the ViruSurf system, but we aim to specify the approach precisely and tune it within the team in order to support the design of several extensions for the pandemic: on diagnosis, vaccine design, co-occurring mutations trends, and clinical applications. This experience is not isolated; several are the fields, in life sciences as well as in emergency management [15], in which we expect ERE to be necessary. Our technique, consolidated over time, will allow to collect new requirements while verifying the ones that are progressively translated into software, short-circuiting times and quickly enlarging the impact of a system.

Acknowledgments

References

[1] A. Bernasconi, A. Canakoglu, M. Masseroli, P. Pinoli, S. Ceri, A review on viral data sources and search systems for perspective mitigation of COVID-19, Brief. in Bioinf. (2020).

[2] A. Canakoglu, A. Bernasconi, A. Colombo, M. Masseroli, S. Ceri, GenoSurf: metadata driven semantic search system for integrated genomic datasets, Database 2019 (2019).

[3] A. Canakoglu, P. Pinoli, A. Bernasconi, T. Alfonsi, D. P. Melidis, S. Ceri, ViruSurf: an integrated database to investigate viral sequences, Nucleic Acids Research 49 (2020) D817–D824.

[4] A. Bernasconi, A. Canakoglu, P. Pinoli, S. Ceri, Empowering virus sequence research through conceptual modeling, in: Conceptual Modeling, Springer, Cham, 2020, pp. 388–402.

[5] A. Bernasconi, A. Canakoglu, S. Ceri, Exploiting conceptual modeling for searching genomic metadata: A quantitative and qualitative empirical study, in: Advances in Conceptual Modeling, Springer, Cham, 2019, pp. 83–94.

[6] I. Inayat, S. S. Salim, S. Marczak, M. Daneva, S. Shamshirband, A systematic literature review on agile requirements engineering practices and challenges, Computers in Human Behavior 51 (2015) 915 – 929.

[7] I. Hadar, P. Soffer, K. Kenzi, The role of domain knowledge in requirements elicitation via interviews: an exploratory study, Requirements Engineering 19 (2014) 143–159.

[8] D. Bolchini, A. Finkelstein, V. Perrone, S. Nagl, Better bioinformatics through usability analysis, Bioinformatics 25 (2009) 406–412.

[9] W. Ritzel Paixão-Côrtes, V. Stangherlin Machado Paixão-Côrtes, C. Ellwanger, O. Norberto de Souza, Development and usability evaluation of a prototype conversational interface for biological information retrieval via bioinformatics, in: Human Interface and the Management of Information. Visual Information and Knowledge Management, Springer, Cham, 2019, pp. 575–593.

[10] P. Crovari, S. Pidò, F. Garzotto, Towards an ontology for tertiary bioinformatics research process, in: International Conference on Conceptual Modeling, Springer, Cham, 2020, pp. 82–91.

[11] B. Nuseibeh, S. Easterbrook, Requirements engineering: a roadmap, in: Proceedings of the Conference on the Future of Software Engineering, 2000, pp. 35–46.

[12] B. Willers, Show, Don't Tell, Springer, London, 2015, pp. 3–22.

[13] T. Brown, Design thinking., Harvard business review 86 (2008) 84–92,141.

[14] E.-M. Schön, J. Thomaschewski, M. J. Escalona, Agile requirements engineering: A systematic literature review, Computer Standards & Interfaces 49 (2017) 79–91.

[15] L. Yang, R. Prasanna, M. King, GDIA: Eliciting information requirements in emergency first response, Requirements Engineering 20 (2015) 345–362.