

# Towards using focus groups to identify software developer's interests regarding their development process

Lynn von Kurnatowski, Annika Meinecke, and Adriana Rieger

Institute for Software Technology

German Aerospace Center (DLR)

Cologne & Weßling, Germany

Email: {lynn.kurnatowski,annika.meinecke,adriana.rieger}@dlr.de

**Abstract**—The assurance of quality, reliability, and trustworthiness of software systems is a basic requirement in software development. Therefore, it is necessary to have a comprehension and detailed understanding of a software project. To make the understanding of a complex software system more accessible, a set of tools that analyze and visualize complex software systems could be introduced. Such a software analysis and visualization tool has to meet challenging requirements to be suited for the specific needs of users. To fulfill these requirements and to provide an intuitive way to understand the software projects with visual analysis, we use the human-oriented method of focus groups. We developed a design for focus groups to identify topics of interest regarding the analysis of software development processes. To test and improve our focus group design we conducted a pilot study with research software developers.

**Index Terms**—focus group, human-oriented, visual analytics, program analysis

## I. INTRODUCTION

Software processes as well as software solutions are complex and, like all creative work, depend on human decisions and thus are prone to errors. However, quality, reliability, and trustworthiness are basic requirements in software development. In safety-critical software systems, such as in the fields of medicine, transport, and space, these requirements are especially important. As a result, the need to make existing knowledge about a software project more accessible to developers and derive insights into quality, reliability, and trustworthiness of a software product intensifies.

Visual analysis [1] is an opportunity to easily and quickly verify a software project, since the human brain is an amazing pattern matching machine. This allows developers to get an overview of a software status regarding its quality, reliability, and trustworthiness. For this, different metrics can be visualised using appropriate visualisation methods, like graph visualizations, metrics visualizations (e.g., bar charts), time-oriented visualizations (e.g., Sankey charts), etc.. Then, these visualisations can be integrated into a web-based dashboard [2], using the Python framework DASH<sup>1</sup> and Python library PLOTLY.

<sup>1</sup><https://dash.plotly.com/>

For the purpose of examination and analysis, our main goal is to analyze the software development process and provide software developers an intuitive way to understand the different metrics using visual analysis. Thus, we want to support developers in their work process and assist them in identifying abnormalities occurring during development. In order to accomplish this task, we need to collect and analyze information about software projects on the one hand, and on the other hand, we need to identify which aspects of the software development process are relevant for developers during their development process. However, identifying the interests of a software developer regarding the analysis of software development processes is not a task easily accomplished.

To identify these interests and needs we decided to adopt methods from social sciences and human-computer interaction in our research. Different forms of qualitative and quantitative methods can be used to identify human aspects in various areas [3]. We decided to adopt focus groups [4] as a qualitative method for the identification of different requirements which we can use to compile visual analysis.

Focus groups have been an established and cost efficient way to gain information on a specific topic from multiple participants at the same time [5]. Therefore, we want to determine:

Are focus groups an appropriate method to discern aspects of interest of a software developer regarding the analysis of software development processes?

With this knowledge as foundation we want to create a tooling perfectly adapted to the requirements of different developers and individual software development processes.

In this paper we describe our current status towards using focus groups as a method to identifying the interests of a software developer regarding the analysis of software development processes in the following structure:

- First, we give a brief insight into our tooling how to generate the data sets (Section II).
- Afterwards, we describe how we want to use focus groups as a method to identify the relevant requirements (Section III).

- Following, we describe how we conducted a pilot study and present the findings with regard to improving our focus group (Section IV).
- Finally, we summarize the major findings and indicate future work directions (Section V).

## II. DATA EXTRACTION

As mentioned in the introduction (I), to be able to make an assurance about quality, reliability, and trustworthiness of a software product, it is necessary to have a comprehensive and detailed understanding of a software project. One approach to analyze processes and to verify the resulting data products is the usage of provenance.

As stated by Moreau in 2010 in “The Foundations for Provenance on the Web”:

Provenance is becoming an important concern for several research communities in computer science, since it offers the means to verify data products, to infer their quality, to analyze the processes that led to them, and to decide whether they can be trusted. [6, Page 3]

Therefore, we collect and analyze the provenance of software artifacts and the provenance of software development processes.

Provenance describes the people, institutions, entities, and activities involved in creating, processing, or providing data [7]. Provenance can be formally expressed in different ways. We use the W3C specification PROV [8], which among other definitions defines the Provenance Data Model (PROV-DM). PROV-DM relies on the model class elements *entities*, *activities*, and *agents*. Agents are involved in producing a piece of data or artifact. Relations, such as *wasGeneratedBy*, *used*, *wasAssociatedWith* and *wasAttributedTo*, are used to relate these class elements (Figure 1) [9].

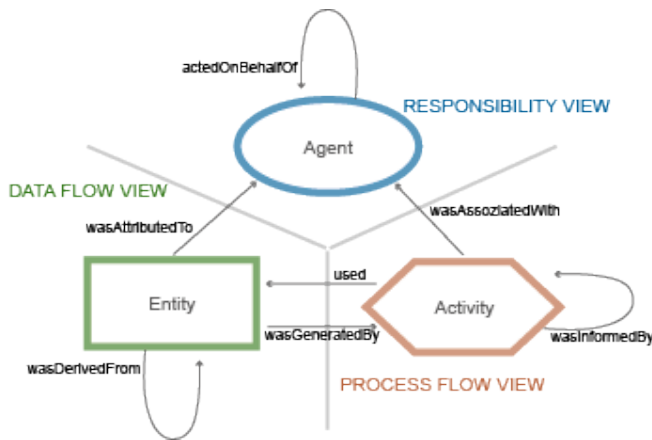


Fig. 1. PROV core concepts including the class elements entity, activity, agent, and relations.

### A. Process Extraction Tool

For software development processes based on git repositories, our tool GitLab2Prov [10], [11] extract the complete

provenance of all process steps through data mining on these repositories; in the case of GitLab<sup>2</sup>, this also includes provenance information from the respective issue trackers and CI systems. The resulting provenance data then contains all activities (e.g., commit creation, issue changes, release events), the generated or changed entities (e.g., source code files or issues), and involved agents (e.g., developers, testers, or users) along with their relations. For further analysis the provenance information can be imported as a labeled property graph into the graph database Neo4j<sup>3</sup>, using the process store tool [11].

## III. FOCUS GROUP DESIGN

The aforementioned increase of complexity in software solutions results in the need of more extensive development processes. Thus, resulting in very large and complex provenance graphs (see Section II), which are difficult to manually understand, interpret, and verify. Visual analysis is an opportunity to easily and quickly verify a software project. In order to provide an intuitive way to understand the software projects with visual analysis and to identify the interests of a software developer, we use the human-oriented method of focus groups.

We plan to conduct six focus groups [12], with each lasting about 90 minutes, to identify topics of interest regarding the analysis of software development processes. Participants of our focus groups will be research software developers experienced in the development of research software. In research software development we find many developers with different backgrounds, which gives us the opportunity to build diverse focus groups and gain insights from different perspectives while still keeping the focus on the main topic of software development processes. We aim to build focus groups with participants from diverse backgrounds working on different projects to profit from the different experiences each participant brings to the table. The focus groups will be conducted online using common video conference tools.

We decided on a smaller group size of up to 6 participants for each focus group to ensure a lively group dynamic in which everybody has the opportunity to participate with each other. Furthermore, each session will be recorded for analysis.

The focus groups will be accompanied by two researchers. One moderator and one as observer and note taker, as well as support for the moderator regarding organisational and technical aspects. After considering the findings of our pilot study (described in Section IV) the focus groups will be organized as follows:

- 1) Introduction.
- 2) Discussion of individual software development processes.
- 3) Discussion of specific topics of the development process. Based on word clouds generated prior to the focus group session.
- 4) Conclusion and general feedback.

<sup>2</sup><https://about.gitlab.com/features/>

<sup>3</sup><https://neo4j.com/product/neo4j-graph-database/>

During introduction all participants have the opportunity to introduce themselves and their personal background. In addition, the moderator will introduce the topic of the session.

The second part of the focus group will concentrate on software development processes. All participants will have the opportunity to describe their individual methodologies and processes in detail. Through this, we want to introduce the topic and focus participants on the related mindset. In addition, we, as well as the participants, will gain an overview of different approaches to the software development process. After stating their processes, participants have the opportunity to discuss similarities and differences, as well as advantages and disadvantages of their individual processes.

After getting an overview of the general aspects of the software development processes in our focus group, we want to identify developer's interests on specific topics of the development process. From our existing experience with research software developers we identified different topics of interest in the development process. Since we only have a limited amount of time, we decided to concentrate on six topics during the session:

- Team Interaction,
- Milestones,
- Review,
- Issues,
- Merge Requests,
- Commits.

Prior to the focus group, participants should write down their associations for each topic. From their responses we build word clouds, which are used to give a statistical overview on emerging topics by correlating word frequency with font size. We decided on generating word clouds from feedback of participants on the topics the discussions should focus on to give participants the opportunity to form their thoughts on these topics and share them with us prior to the group discussion. This gives us the opportunity to gain an overview of individual interests which are not influenced by opinions emerging in the group discussion. Furthermore, it gives us the opportunity to identify reemerging themes. We will present the generated word clouds to the focus group to help them start the discussion on the topics and focusing them on the topic to be discussed. An additional advantage of using word clouds to focus the discussion is, that we can minimize the interaction of participants with the moderators, thus reducing the possibility of introducing a potential moderator bias, while upholding a lively discussion between participants on each topic. During the conclusion of the focus group, participants can address topics they missed during the workshop, ask questions as well as give general feedback regarding the session.

#### IV. PILOT STUDY

Our main goal for the pilot study was, to identify problems in our study design and improve our design and procedure accordingly.

##### A. Procedure of the pilot study

To test our method, we conducted a pilot study. We invited 5 experienced software developers, who work on different software projects. All participants have a background in computer or data science and work for the same research facility where they develop scientific software. All participants knew each other; however, they all work on different projects.

We used the video conferencing tool BigBlueButton<sup>4</sup> to conduct the online focus group. The session was scheduled to last one hour, participants were not asked to prepare anything in advance.

The pilot session was organized as follows:

- 1) Introduction.
- 2) Discussion of individual software development processes.
- 3) Individual work in which participants were asked to write down word association to the different topics.
- 4) Discussion of specific topics of the development process, based on the generated word clouds built from the associations in the previous part.
- 5) Conclusion and general feedback.

##### B. Findings of the pilot study

In the following section we will concentrate on our findings to improve design and procedure and not on the findings regarding the topics discussed.

Since all participants were experienced with video conferencing tools, the session went without any technical difficulties. Overall we identified several points on how to improve our procedure. First, while all our participants knew each other, they were not familiar with each others work and the context in which they develop software. Since the context has an impact on their individual process, in future sessions we will put a special focus on the individual differences during introduction. This will be even more important since participants will not know each other during the following focus groups. We feel this will help participants to understand each other better during discussion and therefore improve the quality of their exchange.

While participants talked about their development processes we saw a lot of variation. One of the main aspects which had an impact on the discussion was, if participants develop software alone or in a team. While we were aware of this issue before, the pilot focus group shows that building different focus groups for developers working in teams or alone might be beneficial for the discussion. This would also help us identifying differing requirements for each group.

Building word clouds from participant's associations with different topics gives us an insight in individuals interests without the influence of group opinions. Furthermore, we experienced that presenting the word clouds to participants centered their discussion on the topic, without much input by the moderator. Nonetheless, building the word clouds during the session led to them being presented without processing the

<sup>4</sup><https://bigbluebutton.org/>

data first. Therefore, for our main study we decided to create these word clouds from participant's responses obtained prior to each session.

Overall we experienced that participants led a lively discussion in which participants were able to voice their opinion on each topic. The group size of five participants ensured that each participant could be included in each topic discussion, without them fading into the background. During the concluding feedback participants voiced their enjoyment during the session and all were interested in a follow-up. They felt that the one hour session we scheduled was too short since it did not allow them to dive deeper into each topic.

## V. CONCLUSION AND FUTURE WORK

To enable assessment of the quality, reliability, and trustworthiness of a software product, it is essential to have a comprehensive and detailed overview of the respective software project. For this purpose, our main goal is to analyze the software development process and provide software developers an intuitive way to understand the different metrics using visual analysis. But such a software analysis and visualization tool has to meet challenging requirements to be suited for the specific needs of users. To fulfill these specific requirements, we first need to identify topics of interest regarding the analysis of software development processes.

We decided on exploring focus groups as a method to determine the requirements of software developers since they are an effective and cost efficient way to gain insight on developers' interests. We developed a design on how to conduct the focus group involving word clouds to center the focus of the participants on specific predefined topics (see Section III). In addition we conducted a pilot study to improve our method (see Section IV).

In the near future, we plan to test our method by conducting multiple focus groups as described above. All focus groups will be transcribed and analysed using qualitative content analysis. The insights we gain will be used to develop concepts on how to provide software developers an intuitive way to analyze their development process. Our concepts will be evaluated by potential users through user experience testing.

After collecting first requirements through focus groups, we also want to explore how different roles in a software development process influence the requirements regarding an analysis tool.

While we decided on using focus groups to conduct our research, in the future we want to explore different methods like contextual inquiries or interviews as well and evaluate which methods are best suited to identify user requirements of research software developers.

## REFERENCES

- [1] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, Eds., *Mastering the information age : solving problems with visual analytics*. Goslar : Eurographics Association, 2010. [Online]. Available: <https://diglib.org/handle/10.2312/14803>
- [2] A. Schreiber, L. Kurnatowski, A. Meinecke, and C. Boer, "An interactive dashboard for visualizing the provenance of software development processes," in *2021 Working Conference on Software Visualization (VISSOFT)*. Los Alamitos, CA, USA: IEEE Computer Society, sep 2021, pp. 100–104. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/VISSOFT52517.2021.00019>
- [3] C. Kothari, *Research Methodology: Methods and Techniques*. New Age International (P) Limited, 2004. [Online]. Available: <https://books.google.de/books?id=hZ9wSHysQDYC>
- [4] D. L. Morgan, "Focus groups," *Annual Review of Sociology*, vol. 22, no. 1, pp. 129–152, 1996. [Online]. Available: <https://doi.org/10.1146/annurev.soc.22.1.129>
- [5] J. Kontio, J. Bragge, and L. Lehtola, *The Focus Group Method as an Empirical Tool in Software Engineering*. London: Springer London, 2008, pp. 93–116. [Online]. Available: [https://doi.org/10.1007/978-1-84800-044-5\\_4](https://doi.org/10.1007/978-1-84800-044-5_4)
- [6] L. Moreau, "The foundations for provenance on the web," *Foundations and Trends® in Web Science*, vol. 2, no. 2-3, pp. 99–241, 2010. [Online]. Available: <http://www.nowpublishers.com/article/Details/WEB-010>
- [7] L. Moreau and P. Missier, "Prov-dm: The prov data model," Online, 2013, retrieved: 03.10.2021. [Online]. Available: <https://www.w3.org/TR/prov-dm/>
- [8] L. Moreau and P. T. Groth, *Provenance: An Introduction to PROV*, ser. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan Claypool Publishers, 2013.
- [9] L. Moreau, P. Groth, J. Cheney, T. Lebo, and S. Miles, "The rationale of PROV," *Journal of Web Semantics*, vol. 35, pp. 235–257, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1570826815000177>
- [10] A. Schreiber, C. de Boer, and L. von Kurnatowski, "Gitlab2prov—provenance of software projects hosted on gitlab," in *13th International Workshop on Theory and Practice of Provenance (TaPP 2021)*. USENIX Association, Jul. 2021. [Online]. Available: <https://www.usenix.org/conference/tapp2021/presentation/schreiber>
- [11] C. de Boer and A. Schreiber, "prov2neo," Dec. 2021. [Online]. Available: <https://github.com/DLR-SC/prov2neo>
- [12] G. Guest, E. Namey, and K. McKenna, "How many focus groups are enough? building an evidence base for nonprobability sample sizes," *Field Methods*, vol. 29, no. 1, pp. 3–22, 2017. [Online]. Available: <https://doi.org/10.1177/1525822X16639015>