PROCEEDINGS OF SCIENCE

Innovation Data Sprint in Citizen Social Sciences

Sophia Segler[a,*] **and Julia Gantenberg**[a]

[a] *University Bremen,*
Celsiusstraße 2, Bremen, Germany

E-mail: segler@uni-bremen.de

The concept of *data sprint* is tested and implemented in the citizen social science project "Exploring Society Together" (GINGER), to explore the issue of how to provide co-interpretation of big data sets with citizen scientists and what infrastructure is needed to design research in dialogue with citizens. Therefore we tested in GINGER the participatory research format of a public data sprint in which big data sets are analyzed and interpreted in cooperation between researchers and citizen scientists according to scientific standards. Public data sprints are innovative formats in the transitive field of science communication and citizen sciences. The innovation of public data sprints in citizen social science formats lies in the short and intensively research collaboration between citizens and scientists who explore a research topic and data set together within just multiple consecutive days. Each data print focuses on a specific topic, in this case digital social and political network analysis and polarization tendencies of politicians on X (formerly Twitter). The data and digital infrastructure in this discussed format is provided by „Exploration of Political Information Networks“ (EPINetz). With the support of extremely low-threshold applications to analyze data digitally, as by EPINetz, data sprints provide excellent opportunities for broadening perspectives in social science disciplines by having many people working and thinking in a concentrated way on one or more small topic-related problems or puzzles of social phenomena analyzed, based on big data information.

Austrian Citizen Science Conference 2023, ACSC2023
19 - 21 April 2023
Linz, Austria

*Speaker

PoS(ACSC2023)019

1. Introduction

The process of datafication has reached the citizen social sciences as well and levels the access to co-participatory interpretation of large data sets in the dialogue between professional scientists and citizen scientists. In the citizen social science project "Exploring Society Together" (GINGER) [1] at the University of Bremen, we address the issue of how to offer adequate infrastructure and what interdisciplinary expert knowledge is needed to be able to jointly evaluate large data sets with citizen scientists. A barrier is posed above all by knowledge about literacy and evaluation of large data sets, which cannot be taken for granted when working with citizen scientists. In the GINGER project, we have therefore tested the innovative data sprint format in a citizen social science context in cooperation with the digital learning and research platform "Exploration of Political Information Networks" (EPINetz) [2] and would like to share our initial experiences and reflections. We therefore introduce the public data sprint format and relate the format to questions around the integration of computational citizen social sciences and interdisciplinary cooperation with EPINetz. In the main body, we outline the process and best practices of data sprints, as well as emerging hurdles and limitations of co-interpretation of big data sets. By introducing the public data sprint format to citizen social sciences, we want to continue to spark and inspire the important dialogue around big data and data literacy in the citizen social science community.

2. The Format, Features and Methods to Conduct Citizen Social Science with Big Data

2.1 The Format: Public Data Sprint

The public data sprint format is inspired by "hackathons" from internet activists and defenders of the free world wide web and open digital speech, as well as, in contrast therefore, from a market-driven logic in which data sprints are used to collaboratively promote solutions for complex data or innovations in companies [3]. These two movements have given rise and inspiration to the data sprint format in the social sciences, which the Royal Danish Library [4], for example, offers annually on changing topics and data-sets. Depending on interest, data sprints can cover all research steps, such as finding research questions, developing hypotheses or refining and strengthening them, jointly evaluating data sets, or, for example, developing creative ideas to prepare existing research results and scientifically inaccessible data bases for a broad audience in a target group-specific way. The most exciting approach to data sprints is to involve people who do not bring a scientific expert's view to the format and the respective topic, but rather their expert views from diverse socialization and everyday experiences and talents. Here data sprints become *public* data sprints. It is the dialogue between expert scientists and citizen scientists that makes public data sprints particularly inspiring and extremely intensive research formats. Therefore, emphasis is placed on presenting the results of the data research and their discussion at the end of each data sprint and also making the results publicly available afterwards. Thematically, data sprints can be either tightly structured or completely open with the stipulation of a guiding theme. According to the research question or the search for a solution, data sprints can be designed conceptually and in terms of content. Furthermore, the data sprint results can either be used for integration into existing research, for the

start of new (participatory) research or, in the case of extensive data interpretation, even for citizen science publications, which will be elaborated after the sprints.

2.2 The Database and Feature for Data-Analysis: Exploring Political Communication (EPINetz)

For the performance of the data sprint in the citizen social science project GINGER, we were able to cooperate with a partner that is ideal for the format: "Exploring political information networks" (EPINetz), a digital learning and research platform, provided by an interdisciplinary team of researchers in the field of computer science (University Heidelberg) and computational social science (University Hildesheim) [2, 5]. Methodologically, EPINetz offers the analysis of big data sets on the basis of hashtags or keywords according to time series, metadata analyses that can also serve for content analyses as well as the unique and innovative function of low-threshold network analyses of words, as well as German politicians and their co-occurrences. These features enable political discourse mapping of heterogenous networks [5]. EPINetz thus offers a suitable, innovative and extremely low-threwshold infrastructure and widgets for participatory research on big data with non-experts and without the need for knowledge on programming languages, usually essential for analysis of large data sets. We worked with GINGER citizen scientists during the public data sprints using metadata from X (formerly Twitter), which EPINetz was able to provide updated daily until early June 2023. In the meantime, the application programming interface (API) to Twitter data became more expensive. Nevertheless, EPINetz continues to offer highly relevant data via links to German daily newspapers that are also digitized and analyzable via EPINetz's features [2]. The innovation of EPINetz is that the extensive information based on big data, which is updated continuously, leaves room for a multitude of new knowledge, which is made easily accessible by the low-threshold application of the platform. In particular, the highly accessible analysis of complex information networks in EPINetz is a blueprint for citizen social science and co-interpretation of big data.

3. Performance of the Public Data Sprint and preliminary lessons learned

3.1 Performance of the Public Data Sprint

We were able to gain initial experience in carrying out the data sprint format in cooperation with EPINetz within two seperated sessions over two days at the University of Hildesheim [6]. The public data sprint started with a short introduction of social science methods, such as text analysis and network analysis and their interpretation, as well as the introduction to theories of political communication and digital social media. In the next phase, the participants already started with the exploration of the EPINetz platform itself by formulating ideas and hypotheses of possible hashtags and starting with a first data description based on the features (1) time series (2) textdata as Tweets and (3) networks based on word or user coocurences. At the end of the first sprint session, participants presented their initial ideas, hypotheses, and research questions to the group. On the second day of the sprint, the participants got right back into data exploration, but now with a defined hashtag and the task to complete their data description and starting data analysis. The participants in the citizen social science project GINGER have, for example, worked on

the German hashtags "#Paschas", "#LetzteGeneration", "#Russland", "#Ukraine", "#Lambrecht", "#KI", "#Armutsbetroffen", "#Asyl", "#Chatkontrolle". At the time of the data sprints, in beginning of 2023, these hashtags were key terms for popular political-social discussions. In the next step, the data analysis had to be completed and the results had to be transferred into (online) posters. The results have been presented in a final presentation and discussion and have been discussed and classified in a dialogue with experts of the corresponding research discipline and the public. Some of the results can already be openly accessed [6], other results are still being processed and even written up and published by the citizen scientists. The public data sprint format can be extended or shortened to several days, weeks or even selectively and as a one-time event.

3.2 Essentials and limitations of the Data Sprint

In order to be able to carry out the public data sprint format successfully and to achieve useful results, some basic conditions are crucial. Thus, the cooperation of non-professionals and professional scientists is essential. Technical equipment, a ready-to-use data set as well as didacticists and science communicators are also of central importance to be able to offer such a complex format as a data sprint. Especially, the translation activities and the preparation of the data set in the form of text-based instructions require an intensive work phase before the data sprint. Last but not least, it needs interested citizens to participate and feel addressed, which in turn correlates with the work of science communicators. A data sprint is determined by the time limit of the format, in that results and rudimentary solutions, i.e. data interpretations, or at least impulses for research, should be forthcoming. Obviously, a data sprint does not substitute other, deeper research approaches and the essential work of text-based publications. However, the format data sprint is an extremely promising addition to co-productive knowledge, which ideally serves to let interested people from civil society participate in the creation of new knowledge with diverse and relevant life experiences. Likewise, data sprints offer scientists the opportunity to share their thematic and methodological expertise on the one hand, and on the other hand to come across previously unthinkable things and to be inspired in a new and productive way about certain research-relevant questions, thereby reflecting on and questioning previous interpretive stereotypes if necessary.

4. Conclusion

By presenting and sharing our experiences of the innovative format of public data sprints in the citizen social science, we aim to contribute to the enrichment of citizen sciences as well as the evaluation of a suitable format for co-productive data interpretation. While citizen scientists are often involved in the research process of data collection, collaborative data interpretation presents a greater constraint. Public data sprints are therefore very efficient formats to solve problems with the knowledge of the crowd in a very short time or, in the social sciences, to interpret topics of social phenomena jointly between citizen scientists and experts in the field in a more integrated and multi-perspective way. The digital learning and research platform EPINetz offers the technically low-threshold applications for analysis with several social science methods of large data sets. Of course, data-sets can also be used in conventional programming languages and statistical programs during public data sprints, but a low-threshold application and operation without a programming language makes it much easier for non-experts to get started and to co-research the social phenomena

in question. However, we also see the facilities offered by EPINetz as a decisive factor in enabling society to participate in dealing with our data-driven world in a scientific way. Large data sets, which are usually only readable and interpreted by expert knowledge, can be made accessible and readable by learning platforms like EPINetz directly and in a stimulating way to a broad interested public, even without statistics and programming knowledge. This opens up new worlds, not only for the citizen scientists themselves, but also for scientists, who can participate in formats of dialogue, such as the public data sprint, by the knowledge of the many and thus experience the exchange with different social groups and individuals. Therefore, public data sprints are characterized by the cooperation of intensive interdisciplinarity. Public data sprints also benefit from intensive disciplinary collaboration. For the public data sprints described here, specialist scientists from social and political science, computer science and science communication, and last but not least, didactic and educational science as well as political education worked together in order to be able to offer and carry out the formats in an appropriate way and with high quality. On the one hand, this interdisciplinarity is complex and time-consuming; on the other hand, an interdisciplinary team develops enormous creativity and the fusion of the knowledge of many experts in the preparation phase, which also plays a key role for success during the implementation of the public data sprint format in dialogue with citizen scientists. Last but not least, the public data sprint formats and the cooperation with EPINetz are suitable for use in schools and in the context of political education. EPINetz can be implemented in media education, political education and even more demanding language education. Finally, it becomes clear that offering high-quality public data sprints also requires the openness of scientists to open their research practices and spaces, as well as institutional funding and thus the structural security of participatory research. We hope that we can inspire the citizen science community to get involved in the adventure of public data sprints and we look forward for more exchange on this topic in the field of social sciences and citizen sciences.

References

[1] GINGER (Gemeinsam Gesellschaft erforschen), *Citizen Social Sciences Project*. Dr. Julia Gantenberg und Sophia Segler (Universität Bremen), Funded by the German Ministry of Research and Education, 2021 - 2024. https://www.uni-bremen.de/ginger/

[2] EPINetz (Exploration politischer Informationsnetzwerke), *Digitale Lern- und Rechercheplattform zur Exploration politischer Informationsnetzwerke*. Michael Gertz (Universität Heidelberg) und Wolf Schünemann (Universität Hildesheim). Funded by Klaus Tschira Stiftung, gemeinnützige GmbH, 2021 - 2024. https://epinetz.de//

[3] Venturini, T, A. Munk, A. Meunier, "Data-Sprinting: a Public Approach to Digital Research". In: C. Lury; R. Fensham; A. Heller-Nicholas (edit.): *Routledge Handbook of Interdisciplinary Research Methods*. (2018). https://dx.doi.org/10.4324/9781315714523-24

[4] Royal Danish Library, *About Data Sprint*. (no date). https://datasprint.kb.dk/en/about-data-sprint

[5] König, Tim; Schünemann, Wolf J.; Brand, Alexander; Freyberg, Julian; Gertz, Michael: The EPINetz Twitter Politicians Dataset 2021. A New Resource for the Study of the German

Twittersphere and Its Application for the 2021 Federal Elections. *Politische Vierteljahresschrift 2022* (online first) (2022). https://doi.org/10.1007/s11615-022-00405-7

[6] GINGER (Gemeinsam Gesellschaft erforschen), *Ergebnisse Public Data Sprint mit EPINetz, Hildesheim, Stand: 05. Mai 2023*. https://t.ly/ipwX_