

Katharina T. Paul*

Collective organization of discourse expertise using information technology – CODE IT!

A citizen science case study of the Human Papilloma Virus (HPV) vaccination debate

<https://doi.org/10.1515/itit-2017-0022>

Received August 30, 2017; revised November 13, 2017; accepted January 28, 2018

Abstract: This paper offers a short report on a participatory citizen science project and offers some reflections on the lessons learned. In particular, we report on our aims and methods, and the development and use of a web application that we designed to enable a collective analysis of press releases allowing a high number of users. Specifically, we give a brief account of the HTML- and PHP-based platform, which was used to analyze and review press releases on a controversial vaccine.

Keywords: citizen science, social science, coding, crowdsourcing, collaboration

ACM CCS: Human-centered computing → Collaborative and social computing → Collaborative and social computing theory, concepts and paradigms → Computer supported cooperative work

1 Introduction

This paper reports on the development and use of a web application that we designed to enable a collaborative digital analysis of press releases ($n=486$) allowing a high number of users in a social science project. More specifically, the purpose of the research project was to trace and analyze the course of a debate on vaccination over a period of several years by involving citizens with no professional training in social science methods. The research design presented here was driven by our perceived need to increase citizen participation in a very expert-led policy debate and our own desire to explore the policy debate under consideration from a different perspective.

We proceed as follows: First, we situate the project design in the field of citizen science, in which social scientists are thus far underrepresented. Then, following a dis-

cussion of our research aims and methods, we give a brief account of the HTML- and PHP-based platform, which was used to analyze and review press releases on a controversial vaccine. In line with the experimental nature of this project, we also report on aspects of the study that remained underdeveloped as they could be instructive and relevant for future projects. We conclude that interaction and collaboration between social scientists, schools, and IT specialists may help enhance research in the age of ‘big data’, but may also soften boundaries between science and society and promote trust in science.

2 Project design

2.1 Human computation as participation

Human computation often entails taking a (research) problem that appears unmanageable for any one person and redesigning it “into smaller, more manageable pieces that can be delegated to many people” [12]. In this very basic and applied definition, our citizen science project may well speak to those working in human computation, too. Yet as [1] points out, citizen science and human computation can mean different things to different scholarly disciplines and different models of collaboration or “cyberscience” can thus coexist [12]. Likewise, we would add, computation itself takes many different shapes and meanings across disciplines. In our own discipline, that of political science, for example, the divide between manual and computational methods is one of many, and with regard to including non-professional expertise, Prainsack [17] reports substantial resistance in political science and development studies. Disregarding this diversity, human computation generally lacks a discussion of the political – rather than merely technical – value of participation, while recent commentaries on citizen science assesses citizen science precisely through this lens [7].

Our own view is informed by recent experience, rather than conceptual definitions, and we see this special issue as a rare opportunity for social scientists to speak to other disciplines, and to show but one way of how computer

*Corresponding author: Katharina T. Paul, Department of Political Science, Faculty of Social Sciences, Universität Wien, Universitätsstrasse 7, Vienna, A-1010, Austria, e-mail: katharina.t.paul@univie.ac.at

infrastructure can facilitate and mediate between trained scientists and untrained participants in citizen science.

2.2 Social science and citizen science

Citizen science offers the possibility to engage non-experts in the scientific process, to raise their understanding of scientific work, and to carry out research that would otherwise not be possible [21]. Citizen science methodology has primarily been used in the natural sciences, where technical infrastructures for quantitative observations are often readily available or where add-on functions for the sake of citizen participation (to deliver additional data) are developed in a comparatively straightforward fashion. The social sciences, however, have very little experience with citizen science methods: as social scientists typically study *the social*, it seems logistically, ethically, and scientifically challenging to involve citizens directly, apart from participatory action research [10]. The project reported on here is thus fairly experimental.

With the widespread availability and use of internet infrastructures, the past decade has seen a remarkable increase in successful web-based citizen science projects (e. g. FoldIt, GalaxyZoo, PatientsLikeMe) using a variety of human computation approaches. While these have highlighted the advantages of citizen involvement in science, they have also triggered debate [18]. Legal and ethical questions have been raised regarding data security [26], the risk of harming participating citizens [23] or using them only as a free labor for scientific research projects. In the present research project, we partnered with a local school to assess the suitability of web-based citizen science in a controlled setting. Informed by social science protocol, we obtained written informed consent from our participants and their caregivers. In addition, to remedy ethical challenges, we did not use any financial incentives, but instead informed participants about the nature of the exercise, its potential educational value to them, and their valuable role in this research project as well as the policy debate more generally. We shall report here on this pilot project, but less so on its results than on its aims, methods, and challenges.

2.3 Aims and methods

The research project discussed here was concerned with an analysis of a political debate surrounding the introduction of a vaccine against the sexually transmitted Human Papilloma Virus (HPV), which can cause cancer in women

and men but was originally targeted only at young women. The policy debate in Austria was conflict-ridden and polarized, leading to delays in introducing the vaccine and to a debate on immunization more generally between 2007 and 2013. State-sponsored vaccination programs have frequently triggered political conflicts [3], but the HPV debate was specific in mobilizing divergent expert opinions and thus political stances. As previous research suggests [22], the media had an important role in the policy debate, yet the role of divergent expert opinions remains insufficiently explored. The research presented here sought to fill this gap by asking: How, and based on what arguments did expert institutions – including policy actors, scientists, and commercial actors shape the policy discourse through media?

This newly developed methodology was motivated by a strong desire to move beyond conventional political science methods, such as expert interviewing [15] for two reasons: first, our previous research had focused on experts themselves [16], and we wanted to explore a “different way of knowing” [1] about HPV in this otherwise very expert-centered policy field. We did so by tapping into the perception and skills of those typically not involved in either policy research or policy debate: target groups of particular policies. In the case of the HPV vaccine under consideration here, the target group consisted of young adults, or adolescents. A second motivation for this project was to test the suitability and feasibility of citizen-science based content analysis [11, 24], where possibilities for online techniques have recently gained attention [14].

2.4 Participatory computing

While human computation by definition involves some linkage between machines, on the one hand, and individuals or groups, on the other hand [12, 13], these linkages have rarely been captured as political in nature in HC debates. Recent commentaries on citizen science projects in the biomedical sciences have reviewed ongoing research in this light, pointing out that “non-professionally trained people (...) participate in the governance, regulation, and translation of science, as well as in some of the core activities of science itself [7].

In light of their experience as the target group of HPV campaigns, we considered these to be “experts” in their own right, or “untrained experts” [4]. Drawing on the sociology of expertise, we posit that their recent experience of vaccination campaigns – including the conversations these triggered in schoolyards, social media, and at family dinner tables – makes this target group *experts* in the sense

of holding “experiential knowledge”. We sought to involve this target group as citizen scientists by offering a web-based platform that is user-friendly and offers trained researchers the possibility to validate findings. For students, the proposed project offered the opportunity to acquire research skills, to engage with a policy topic of key importance, and to train their skills in observing and analyzing media content. Finally, working with a local school allowed us to pilot our methodological approach before extending it to a wider public. Earlier citizen cyber science work [9] has demonstrated the usefulness of such a strategy.

We recruited three cohorts of 16-year old high school students to study and analyze a dataset of 486 press releases issued between 2007 and 2013. In total, we had 75 active participants. The press releases were available to us for download at the Austrian Press Agency (APA) by means of keyword searches (*HPV*). The press releases were manually categorized according to sender (actor category), as this was considered important for further analysis. We differentiated between the following categories: politics, industry, research, media, NGOs, and other. Doing so was to allow for a comparative analysis of content across actor categories at a later stage. The leading research question was to be addressed by means of coding press releases, thus tracing shifts and continuities in the political debate: How, and based on what arguments did expert institutions – including policy actors, scientists, and commercial actors shape the policy discourse through media?

2.5 Coding as method

In the analysis of text – spoken or written words – coding is a fairly common and well-established method in the social sciences, but it can mean different things. For social scientists, coding typically “involves taking text data or pictures gathered during data collection, segmenting sentences (or paragraphs) or images into categories, and labeling those categories with a term” [5].

A range of concepts and coding methods exist [2, 20], and manual coding is usually a laborious, careful, and iterative process where the researcher moves back and forth between data and codebook. Manual coding is often preferred for in-depth interpretive analysis, while computer software and even automated coding [8] is often used for large datasets (such as party manifestos, e. g. [19]). In such instances, coding generally entails transforming (in our case, textual) data into a form that is understandable by computer software. Information and data is thus first clas-

sified, making it available for processing with statistical software. It is not uncommon, but typically resource intensive, to employ multiple coders working independently on the same data – this increases reliability, but also enhances the depth of analysis.

Different applications – mostly commercial – are available for coding (Atlas.ti; MaxQDA, Nvivo), yet none of them allow for large numbers of users (coders), nor are they available in an open science format. Adhering to open science principles was important to us for two reasons: first, we sought to encourage colleagues to experiment with citizen science with similar applications. Second, the democratization of science and expertise was not only a topic-based concern, but a procedural one, too. In light of the resource-dependent compromises we often make in textual analysis, web-based citizen science, we propose, offers opportunities for social scientists thus far undiscovered.

As figure 1 below indicates, our coding method included three elements. First, the code book derived in the explorative coding phase [15] served as the base for deductive coding (“vorhandene codes”) second, and the tool “Code hinzufügen” offers the possibility of adding new codes, known as “inductive coding”. The choice of existing codes or the creation of new codes had to be linked to text passages that were copied and pasted into the relevant fields. When participants failed to provide text passages, they received an error message. The image shown in Fig. 1 illustrates the basic structure of the application.

The final part of the application (“Bewertung”) allowed participants to subjectively assess the relevance of individual texts, their credibility, to ability of the text to speak to their own interests, and the intelligibility of the text. In a sense, the project participants were then both coders (researchers) and, to a much lesser extent, objects of analysis, as we explored their interpretations of the HPV policy debate, too. We introduced this function with some hesitation, as we did not want to misuse participants as research objects. Ultimately, this function proved extremely relevant in two ways: first, it conveyed to our participants (or co-researchers) that their subjective *assessment* as much as what they perceived to be an “objective” *analysis* was heard and taken seriously. Second, it led to an important methodological innovation for our research project: Participants remarked that the application should also include questions on what is missing in the text to be analyzed. We acted on this recommendation in the further development of the application, and consider it a prime example of participant-led innovation that is relevant to content and discourse analysis more generally.

Content: text of press release → http://www.ots.at/presseaussendung/OTS_20130812_OTS

Possibility to add new code (inductive) → **Code hinzufügen!**

Deductive coding e.g. „vaccination is for girls“; „critical view of vaccination“ etc. → **Vorhandene Codes**

Assessment of press release (e.g. relevance, trust etc.) → **Bewertung**

Figure 1: Structure of the application.

3 Developing a citizen science web application

3.1 Citizen science in action

Given the aims and principles discussed above, the requirements for the organization of this “collective discourse expertise” – hence the acronym of the project – were as follows. First, the application had to be able to accommodate a very high number of users working on the same texts simultaneously. Second, coders had to be able to assign text passages to particular existing codes (known as *deductive coding*) listed in the application, and had to be able to create new codes based on text passages (*inductive coding*). The existing codes were derived from an earlier research project [15, 16] and were listed with short single sentence narrative explanations. These codes were, in essence, statements about the text, that could be indicated as absent (red button) or present (green button) (see above). The number of assigned text passages was limited to three, while the number of possible new codes was dynamic and infinite, following feedback from users. Indeed, the latter modification of the tool is a good example of how we involved users – or participants, rather – in the further improvement of the coding application.

The overall research project lasted from September 2016 to June 2017, but work on the application and website was commenced well in advance in spring and summer 2016. We introduced our 75 participants to the appli-

cation in a three-hour workshop at the partnering school and held a total of nine workshops – three per each cohort – in addition to a number of walk-in Q&A sessions. In these workshops, we first familiarized students with citizen science more generally, the rationale of the research project, and our research questions. In the second workshop, we introduced students to the web application and offered support and advice in a third workshop, where participants worked individually on their coding exercises. We found that we were able to instruct our participants in basic coding very effectively, not least because they felt involved and appreciated as critical readers of texts that concerned a political and scientific controversy.

3.2 Intercoder reliability and data quality

Given the experimental nature of this project, an additional necessary element was the need to validate and trace the data entries of participants. Intercoder reliability checks were performed in three ways. First, and most importantly, the PI examined data entries in the early stage of the project (see section 3.3. for technical details). The PI’s expertise concerning the HPV debate allowed her to do so most expediently. Second, the research design (working with three cohorts and with interactive workshops) allowed us to use intercoder reliability checks in a more iterative and personal fashion than would typically feature in web-based citizen science. Finally, following the assignment of codes, a trained research assistant validated a



randomly selected sample (10 % of the total dataset) and found the results to be in line with participants' coding. In light of common concerns regarding data quality in citizen science [25], especially when working with younger citizens [6], we thus assessed the degree of concordance and discordance between findings of citizens and trained researchers and found it insubstantial in terms of the end result. We did, however, detect a significant difference in understanding one particular code: while we had asked participants to indicate (i. e. code) instances where the vaccine was discussed critically, the participants marked text passages where there was criticism of other actors instead (regarding their stance on HPV). The participants thus interpreted our code differently, but simultaneously created a new code without intending to do so. This is but one example of how citizen science must leave room for creativity and experimentation, in order for such projects to become more than just low-cost data gathering. Indeed, it is this space for interpretation that allows for learning and important methodological lessons both for non-trained experts and trained experts, leading to "interactive expertise" [4].

3.3 Technical features

We created an HTML- and PHP-based platform, which was used to analyze and review press releases and for participants to analyze content – for instance, information about the free availability of the vaccine to boys and girls, and information on the sexual transmission of the virus. The gathered information was saved in a MySQL database. We chose MySQL as a readily available data management system suitable for database design that is able to link information provided by users (i. e. their analysis of texts) and information provided *about* users themselves (gender, age). We deemed this relevant to be able to draw conclusions about the suitability of the platform for the purpose of citizen science, and to explore inter-reader variability in texts that often contain messages appealing to gender and generation. With an additional interface, administrators were able to upload new press releases and to assign them to a data (actor) category as input for further discourse analysis. Additional information, such as the date of creation and issuer of the press release, was added in the filename. Doing so was desirable for researchers as this is common practice in digital libraries.

The database can be exported to an xml-file in the admin-view, to make it anonymous and readable for STATA, a statistical software package commonly used in the social sciences. To provide anonymity, usernames were hashed in the export-process. Additionally, we set up a

real-time statistics view for admins to get a fast overview. This included, among other things, information on the frequencies of entries of specific codes, omissions made, and data on which users were adding codes, and on what subject. We made use of this function particularly to assess the effectiveness of our didactical methods in the coding workshops, and to improve insights regarding intercoder reliability, as discussed above. Moreover, our project required the possibility to report errors, and for reports to become readily available in the admin-view. Finally, the data had to be easily extractable for administrators (social scientists). Buttons were thus created for administrators to be able to convert and download data into an XML file.

3.4 Challenges

The tool was hosted on the university server and had to operate within the given technical constraints. The research institution offered a website equipped with Content Management Software (TYPO 3) that made it possible for researchers to create and update information for users easily, thus saving resources. The web application itself was equally hosted by the university, but on an Apache server, and was linked by means of i-frame.

There were no resources available for usability testing and the tool had to be adaptable to new and unexpected developments as the research progressed, due to the experimental character of the project. The design of the application was a largely interactive process between project leader, research assistant, and programmer. The short duration of the research project demanded flexibility of all parties involved, as well as elasticity as far as the technical infrastructures was concerned: we had to give way to demands considered easily feasible by participants and researchers, but which were ultimately time-consuming adaptations from a programming perspective. For instance, the transformation of a static to a dynamic number of codes and text passages to be added was more time-intensive than the original creation of the tool itself. In another instance, the dependence of the tool on a readily available internet connection proved difficult: the coding workshops were conducted at the local secondary school, and the internet connection proved too slow on some machines. With our first cohort of participants, we created an offline, local version of the database on a local machine, but researchers were not aware that some users were already accessing the application on their smart phones. The two databases (online and offline) were then merged afterwards. In subsequent workshops, we provided hotspots.

This instance points to a set of particular and indeed unexpected challenges. First, our participants were less proficient with software use than we had expected. While often labeled as the “digital generation”, they were much less flexible than we had hoped when it came to using applications that were not only new to them, but also, in essence, work in progress. Due to budgetary constraints, no budget had been allocated to, for instance, piloting the application with peers or improving its design. To those who seek to work with web-based citizen science, we strongly recommend doing so in future projects. Second, we had opted not to invest in developing an app suitable for smartphones or tablets, which proved frustrating for some of our participants who would have preferred to contribute to the project in a mobile fashion. While we handled these challenges in an *ad hoc* fashion and more or less effectively, we report on them to ensure that future projects – and funding agencies – take these needs into account, for instance, in social science and humanities, where an increasing need for digital applications is being discussed.

4 Outlook and conclusions

In sum, this project broke new ground by testing and further developing citizen science methods in political science. Besides its substantive results, the study had great didactical value for all involved parties. As social science researchers, we learned to communicate and work with IT specialists more effectively, a skill we can expect to be useful in the age of ‘big data’ research. For our participants, the collective analysis of data offered training in basic social science methodology and basic statistics. Finally, and above all, the interactive and dynamic nature in which different worlds and experiences came together in this project – teachers, students, researchers, and programmers – was unique in its improvement on our mutual understanding and the promotion of transparency and openness in this risky and conflict-ridden research topic of vaccination policy. As such, we conclude that innovative – online and offline – encounters of this kind may also be effective in other controversial policy areas and may be of use in enhancing scientific transparency as well as mutual trust.

Acknowledgment: The author would like to thank all participants at AHS Rahlgasse (1060 Vienna), specifically those in groups 6A, 6B, 6C, and 6D in 2016-17. The project team is indebted to Ulrike Randl-Gadora for important

feedback on our research design and methodology. The author is grateful to Thomas Palfinger for his invaluable research assistance and to Nikola Szucsich for his patience as a programmer.

Funding: This work was supported by the Austrian Science Fund (FWF) Top Citizen Science Grant #14.

References

1. Brown P. (1992). Popular Epidemiology and Toxic Waste Contamination: Lay and Professional Ways of Knowing. *Journal of Health and Social Behavior* 33(3), 267–281 URL: <http://www.jstor.org/stable/2137356> Accessed: 04-02-2016 18:53 UTC.
2. Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Thousand Oaks: SAGE Publications.
3. Colgrove, J. (2006). *State of Immunity: The Politics of Vaccination in Twentieth century Americ*. Berkeley: University of California Press.
4. Collins, H. M. and Evans, R. (2002). The third wave of science studies of expertise and experience. *Social studies of Science* 32(2), 235–296.
5. Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. 3rd ed. Thousand Oaks, CA: SAGE Publications.
6. Delaney, D. G., Sperling, C. D., Adams, et al.(2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions* 10(1), 117–128.
7. Del Savio L., Prainsack B., Buyx A. (2016). Crowdsourcing the Human Gut. Is crowdsourcing also ‘citizen science’? *Journal of Science Communication* 15(3).
8. Grimmer, J. and Stewart B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3), 267–297.
9. Heigl F., Zaller J. G. (2014) Using a Citizen Science Approach in Higher Education: A Case Study Reporting Roadkills in Austria. *Human Computation* 1(2), 165–175.
10. Kindon, S., Pain R., and Kesby M. (2007) *Participatory Action Research Approaches and Methods. Connecting people, participation and place*. New York: Routledge.
11. Mayring, P. (2010): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 11th Edition, Weinheim: Beltz.
12. Michelucci, P. (2013). *Handbook of Human Computation*. New York, Springer.
13. Newman, G. (2014). Citizen CyberScience: New Directions and Opportunities for Human Computation. *Human Computation* 1(2), 103–109.
14. Popping, R. (2017). Online tools for content analysis. In *The SAGE Handbook of online research methods*, (pp. 329–343). Thousand Oaks: SAGE Publications.
15. Paul KT (2016). ‘Saving lives’: Adapting and adopting HPV vaccination in Austria. *Social Science & Medicine* 153, 193–200.
16. Paul, K. T., Wallenburg I., Bal R. (2017). Putting public health infrastructures to the test: Introducing HPV vaccination in



- Austria and the Netherlands. *Sociology of Health and Illness* DOI: 10.1111/1467-9566.12595.
17. Prainsack, B., Schicktanz, S., Werner-Felmayer, G. (2014). *Genetics as social practice: transdisciplinary views on science and culture*. Farnham, Ashgate Publishing.
 18. Riesch, H. and Potter, C. (2014). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science* 23(1), 107–120.
 19. Ruedin, D. (2013). The Role of Language in the Automatic Coding of Political Texts. *Swiss Political Science Review* 19 (4), 539–545.
 20. Silverman, D. (2008). *Interpreting qualitative data: A guide to the principles of qualitative research*. Thousand Oaks: SAGE.
 21. Silvertown, J. (2009). A New Dawn for Citizen Science. *Trends in Ecology and Evolution* 24(9), 467–471.
 22. Stöckl, A. (2010). Public Discourses in Policymaking: The HPV Vaccination from the European Perspective. In Wailoo, K., et al. (Eds) *The HPV Vaccine Controversies: Cancer, Sexual Risk, and Prevention at the Crossroads*. (pp. 254–269). New Brunswick: Rutgers University Press.
 23. Vayena, E. and Tasioulas, J. (2013). The ethics of participant-led biomedical research. *Nature biotechnology* 31(9), 786–787.
 24. Welker, M. & Wunsch, C. (2009). Die Online- Inhaltsanalyse. Forschungsobjekt Internet. Köln: von Halem.
 25. Wiggins, A., Newman, G., Stevenson, R. D. and Crowston, K. (2011). Mechanisms for data quality and validation in citizen science. In *2011 IEEE Seventh International Conference on e-Science Workshops (eScienceW)*, (pp. 14–19). IEEE.
 26. Woolley, J. P., McGowan, M. L., Teare, H. J., Coathup V., Fishman J. R., Settersen Jr., R. A., Sterckx, S., and Kaye, J. (2016). Citizen science or scientific citizenship? Disentangling the uses of public engagement rhetoric in national research initiatives. *BMC Medical Ethics* 17(1), 33.

Bionotes



Dr Katharina T. Paul

Department of Political Science, Faculty of Social Sciences, Universität Wien, Universitätsstrasse 7, Vienna, 1010, Austria
katharina.t.paul@univie.ac.at

Katharina T. Paul is a senior researcher and lecturer at the University of Vienna, Faculty of Social Sciences, Department of Political Science. She holds a PhD in Social and Political Sciences from the University of Amsterdam (The Netherlands) and has previously held a position as assistant professor of comparative policy analysis at Erasmus University Rotterdam (The Netherlands). Specialized in health policy, Paul joined the University of Vienna in 2013 and has received several grants from the Austrian Science Fund FWF (grants M1477, TCS 14, VA651, InScide).