

Expert Insights into Collecting Data with Mobile Applications

An Interview with Lukas Otto and Mareike Wieland

Publication date: June 27, 2024

Mobile devices are ubiquitous, and a large number of people have them with them almost all of the time. This provides huge opportunities for computational social science research. On the one hand, mobile data allows for 'in-the-moment' measures of self-reports, using mobile experience sampling or digital diaries. On the other hand, smartphones and other mobile devices produce sensing and tracking data that can be used to measure people's behavior in a very fine-grained and unobtrusive way. In combination, mobile designs and mobile digital behavioral data might answer questions in social science research that are hard to tackle otherwise. Despite this potential, the landscape for mobile research software is still scattered – and scientifically tested, stable, sustainable, and accessible solutions are scarce.

We talked about this with Mareike Wieland and Lukas Otto. Both are researchers in the Computational Social Science Department at GESIS – Leibniz-Institute for the Social Sciences in Cologne. Lukas leads the team “Designed Digital Data”. He holds a PhD in Psychology from the University of Koblenz-Landau. His main research interests are political communication, media effects, mobile methods, and longitudinal data analysis. Mareike is the service manager for the [GESIS AppKit](#), a data collection infrastructure for smartphone survey and sensor data that is currently developed by GESIS. She holds a PhD in journalism and communication studies from the University of Hamburg. Her research focuses on the use, perception and processing of (political) information in automated and mobile news environments.

The interview was conducted by Indira Sen and Leon Fröhling on May 11, 2023. The interview has been edited for clarity and length.

Keywords: data collection, (mobile) experience sampling method (ESM), ecological momentary assessment (EMA), designed data, intensive longitudinal designs, mobile survey, research software development, GESIS AppKit

GESIS: Hello Mareike and Lukas, thank you for the opportunity to talk about your research and service provision. What are you working on now, particularly in the context of mobile-based data collections?

Mareike Wieland: I am working on the intersection between mobile devices and social media, and I am interested in how so-called social media habits are affecting information processing on a cognitive and psychological side, and how this might affect political information and political knowledge. For that purpose, I made use of a research app that is able to track usage and to trigger experience sampling questionnaires based on these usages. In short, it automatically detects what is happening and then automatically triggers questions.

Lukas Otto: From a substantive point of view, I am mostly working on information processing and the processing of media information and how this affects emotional and attentional processes. The smartphone is a very good tool for that, because you can capture these short-lived, dynamic processes. From a methodological point of view, I am mostly interested in modeling longitudinal processes that you can capture with smartphones, for instance those resulting from smartphone sensors or intensive longitudinal designs.

GESIS: How did you end up working on the topic of mobile-based data collections?

Mareike Wieland: When I was doing literature research and observed my environment, I noticed that smartphones and mobile media were becoming more and more important, and that it was important to capture the architecture and the technology behind them in our models and theories. This was my starting point. Then I got contacted by Murmuras, which is a commercial company providing research apps. Since I was struggling a lot with my experiments at that time, I decided to jump on the idea of experience sampling and adapted my research design to what Murmuras could offer. This might not be the most scientific answer, but it is based on my observation that mobile sampling was going to be important, and then it was just a window of opportunity that opened up for me.

Lukas Otto: Up until 2014 or 2015, I was only doing experiments because they seemed to be the only way to capture direct responses to media. Then, in another department at the university where I worked at that time, they bought around 50 smartphones and installed a research app on them, so that people could be pinged during the day. It was only then that I realized that this would be another way to capture what I was investigating. I first started doing this in 2015, not with smartphones that had the app preinstalled but with people's own smartphones. Back then, it was still an issue that people did not have enough mobile data volume, so we needed a solution that first stored the data on the phone before transmitting it to the server.

GESIS: Looking at all the work that you have been doing in the field of mobile-based data collections, were there any major inflection points, moments in which something that was unclear before suddenly became clear?

In experiments, you would do two or three trials and then basically know what to do. Mobile-based data collections are much more complex technically.

Lukas Otto: The whole modeling part of the data became much clearer in the last years. When researchers first captured these data, it was not very clear how to analyze it. This has become much clearer only in the last years and is still developing. Other than that, it was more of a constant learning of what to do and what not to do. For mobile sampling, you are always making mistakes and learning from them, that is very different from experiments. In experiments, you would do two or three trials and then basically know what to do. Mobile-based data collections are much more complex technically, in terms of sampling for example, which makes the learning process more incremental.

Mareike Wieland: I would add that it is not only the modeling, but also the theorizing part. This became evident to me when I was thinking about the fact that our models and theories are mostly not capturing the in-person variance of behavior. Many models and theories want to explain stable personalities using personality traits and their tendency of behavior, but not how these might vary within the same person and in their everyday life. First, I was interested in the method and then I became aware that the theories are way behind. While the modeling part is really improving right now, the theory building is still an issue.

GESIS: Why do you think that mobile-based data collections are an important topic in computational social science research these days?

Lukas Otto: There are multiple reasons. One of them is that the smartphone is not only a tool for surveys and self-reports, but that it can also capture mobile digital behavioral data. This is one of its big advantages: you can integrate digital behavioral data and survey data with no other device. The second thing concerns prevalence. The average person checks their phone 56 times before lunch, every day. Almost everybody checks the phone in the first 15 minutes after waking up. When you talk about phenomena like social media, they are almost purely happening on the smartphone. This methodological part is important for computational and social science people, and then there is the prevalence and societal impact part that makes the smartphone important.

GESIS: What types of tools are available for this work? What types of functionalities do they offer, which types of data do they collect, and what types of measurements can be constructed from these data and these tools?

Mareike Wieland: There is a whole bunch of experience sampling survey tools. One variety is having an app installed on your smartphone that triggers short surveys. The early approach was to have an SMS reminding someone to take a survey, or to directly send a link to a questionnaire. Then, there are tracking apps like Murmuras, which also integrate app usage tracking. Once we talk about digital behavioral data, a big question is whether the app is running on Android or iOS. While survey tools are often covering both operating systems, I do not know of any tool that captures behavioral data, let's say app usage data, and runs on both systems. Finally, there are a lot of self-developed apps resulting from individual research projects that capture behavioral data like GPS data.

While we have some taxonomy of apps, it is not easy to get an overview of all that is going on in this space.

While we have some taxonomy of apps, it is not easy to get an overview of all that is going on in this space. We are trying to do a review right now on that issue, but the problem is that many apps are only used for one single project and are not being maintained anymore – you may still have the repository on GitHub, but the tool is just not usable anymore.

Lukas Otto: There was a review, three to four years ago, about tools for mobile data collection, and that covered around 40 different apps that were able to do mobile data collection in one way or the other, but it was still far from being complete. There are a lot of apps, that are developed for only one purpose. For example, imagine there is a project where somebody wants to know if people feel better when out in nature. This app would have to capture GPS and combine it with a short questionnaire afterwards, something simple like 'How do you feel?'. The researcher would probably just go to a developer, who would easily develop that app for this one specific data collection. There are commercial providers that are usually around for longer, but the non-commercial providers would mostly last only for one or two years.

Mareike Wieland: What is still missing are apps that cover a broader range of sensors for more than one research topic, and especially ones that combine sensor and survey data; they are probably missing because it is not that easy to set this up. A third type of apps are 'diary apps'. This is a special form of survey, an app that invites participants to document what they are doing, driven by their own behavior and not so much by survey questions.

GESIS: What are problems inherently associated with mobile-based data collections?

Mareike Wieland: On a more abstract level, I would again mention the theories that do not cover situation dynamics. On a more technical level, it is difficult to get a first overview. If you are not coming from an institute with a lot of people doing similar research, where do you start? There is nothing like a landing page on the available research tools. What also makes it difficult is that there are a lot of different names or labels for the same method, depending on your subfield. Even if you have an overview on what tools you can use in principle, it will still be difficult to determine which skills and levels of technical expertise are needed to set up the tool and run the study. And so far, we are not even speaking about the type of data that is coming out of the tool.

Lukas Otto: Contributing to these difficulties is that research with smartphones is still too niche to be part of the social sciences or even computational social science education. There are a few universities where this is taught, but this is then mainly driven by people that are doing this type of collection themselves. I would say 80 to 90 percent of the social sciences, psychology and computational social science curricula still do not cover mobile research at all.

The second thing is that you are still depending on tech companies. In computational social science, when we think about dependency from tech companies, we typically think about access to data from Facebook or Twitter/ X. But the dependency is also there for mobile-based research; there are many things that we would like to do that are not possible on iOS, for example. And there is no way to get around it, because the only way of getting apps to Apple phones is the Apple App Store, so you must play by their rules. For Android it is better, but it could also be that Samsung decides tomorrow that certain things will not be possible anymore, taking away even more freedom to do research with smartphones. In our field, we are very dependent on what the companies behind the smartphone operating systems decide.

GESIS: At GESIS, you are currently creating an infrastructure for enabling smartphone-based data collections. What are the main decisions you are thinking about? Maybe we could start with a brief overview of the type of infrastructure that you have in mind and then go into the technical details in a bit.

What we have in mind is an 'AppKit' for setting up and conducting mobile experience sampling studies using a web interface plus a research app that participants install on their smartphones.

Mareike Wieland: What we have in mind is an 'AppKit' for setting up and conducting mobile experience sampling studies using a web interface plus a research app that participants install on their smartphones. We plan to start with the survey component,

because we think you first need to know why people are doing something before you can make use of their behavioral data. Afterwards, we want to add mobile sensor and tracking data, deciding which sensors to prioritize according to the needs within the research community. Ideally, we want to allow researchers to set up and manage their studies within one single web interface, being able to seamlessly combine both types of measures. The last feature we would like to integrate is to use input data from the smartphone, for example GPS, to trigger certain surveys and ask the participants to answer them. This would allow us to study why people behave in a certain way the exact moment they are doing it.

Lukas Otto: I would like to comment from a more strategic perspective why GESIS is now adding to this jungle of apps. We think there is still a need for a system that combines the advantages of commercial and non-commercial software. The non-commercial software, as we just talked about, is often outdated very quickly and requires a lot of programming skills to set up, adapt, and run. This makes non-commercial software for many scientists – especially for many social scientists – not something they would go for.

The commercial software, on the other hand, comes with higher usability and stability, because it is actively maintained and constantly updated. Using these services costs a lot of money, and not all researchers can afford that. Sometimes it is not transparent how the data is gathered, how certain processing steps are done, especially mobile digital behavioral data. For example, what exactly is registered as the opening and closing of an app. This is something that you cannot know when you are using commercial software.

We basically want the best of both worlds: something that is transparent and open, updated continuously while still being easy to use even without programming skills.

We basically want the best of both worlds: something that is transparent and open, updated continuously while still being easy to use even without programming skills. This is the goal and, hopefully, the unique selling point for the GESIS AppKit, in competition with the many other apps we talked about.

GESIS: *Maybe you can walk us a bit through the development process. What were some of the first steps that you took, what were the major challenges that you encountered, and what are some of your lessons learned?*

Mareike Wieland: When you want to develop something new, the first thing you do is some research on whether there is anything comparable. In our case, there was something very useful, which was already at a stage of being more than just some lines of Python code, already out and being tested. Basically, a mobile survey tool. We then had to decide whether to start from scratch or take an existing solution and build upon this solution. We

decided for the latter, which had some disadvantages and some main advantages. We were able to make progress very fast because there were some decisions already made, but on the other hand, we also bought some pitfalls.

When it comes to providing a service, we have to think about legal aspects as well.

For us the decisions were always about finding solutions that would result in a minimal viable product useful for the community as soon as possible, while still being able to further develop and maintain it on a limited budget. Something I learned is that developing costs are high and can scale up very fast if you have to change something more fundamental.

When it comes to providing a service, we have to think about legal aspects as well. We have to figure out what we are actually providing. For example, whether we are only providing software as a service, which means that we are not taking responsibility for the data collection itself from a legal perspective. However, this might also affect the reputation of GESIS, because GESIS stands for data quality. But if everyone can use our service and we do not have any control over the studies that are actually conducted with our tool, does this still align with GESIS' reputation for high data quality? There are a lot of questions that have to be answered.

Lukas Otto: From a conceptual level, there are always a lot of decisions that need to be made for a project like this. For instance, how broad or how specific should its scope be. We could make a very good app that is not yet out there, for example for health interventions. That would be a rather specific app. Or we build something very broad that many social scientists can use in different areas.

We included the community through a survey and a workshop relatively early on.

We basically went with the second option, wanting to start with a product that many people can work with. The integration of surveys was the logical first decision. Likewise, the decision for a sensor will not be a specific measure, like temperature of the smartphone processor, but something with many use cases in in the social sciences. This also always means having an ear on what is happening in the community. We included the community through a survey and a workshop relatively early on.

GESIS: Moving to a slightly different topic for mobile-based data collections, what do the respondents in this type of data collection look like? Are they similar to the

populations that typically answer surveys? Are you hoping to reach populations that are traditionally hard to reach with this new type of data collection?

Lukas Otto: For all the intensive longitudinal data collections, either surveys or sensors, you will hardly reach the same sample sizes in terms of participants that you would reach for bigger surveys.

For smartphone-based studies, however, the sample size is not only defined by the number of participants, but also in the situations and in the measurement points. So even with 100 people you can easily get up to 20,000 to 30,000 measurement points, because sensor data is almost constantly measured, continuous data. Even for surveys, when participants answer three to five surveys per day, you end up with many data points – much higher than the sample size. You just think about the people that you study in a very different way, and this also relates to what Mareike said earlier, about those within-person processes. Smartphone-based studies are much more an individual level idiosyncratic approach than the big surveys that have thousands of people, or with the big data collections in computational social science that have millions of independent measurement points.

Another advantage of the smartphone as a device for web surveys in general is that you can get people to participate that you normally do not get for a traditional survey.

Another advantage of the smartphone as a device for web surveys in general is that you can get people to participate that you normally do not get for a traditional survey. Some people are not reachable through traditional modes of data collection in the social sciences, be it via telephone, face to face or via computer, but almost everyone owns a smartphone. There are already innovative studies, for example with refugees or men recently released from prison [1] that make use of the smartphone to survey hard-to-reach audiences.

Mareike Wieland: Another major issue is willingness to participate in smartphone-based studies. It is not only about reaching audiences of potential participants, also about getting them to successfully install and run the research app. Usually, only between 30 and 40 percent of people who are in principle willing to participate in a smartphone-based study are downloading and running the app. There are a lot of people who opt out at this early stage, and we just do not know why.

Initial research on this issue is being done right now, to find out whether it is for technical issues or whether people simply change their mind. Knowing this is essential. Also for GESIS it is important to improve the onboarding process, to be able to give advice on how to make it as easy as possible to get as many people as possible within the app itself, to

help avoiding systematic dropouts, or at least being able to say who is systematically not reached by the app in the end.

GESIS: How do you recruit participants and pay them differently compared to normal surveys? Connected to that, how do you handle issues of consent and privacy?

Mareike Wieland: Taking part in an intensive longitudinal study is burdensome and commitment of participants over time is essential. Typically, you would provide a flexible pay schedule. That means that participants get paid only if they answer a certain minimum percentage of all surveys carried out over the course of the study. Another approach is to incentivize every single submitted questionnaire with some additional small payment. Or you can have additional bonus points or money for installing the app and for adhering to the study until the end.

There is a lot of research needed to find out whom we are successfully incentivizing, and whom not.

There is a lot of research needed to find out whom we are successfully incentivizing, and whom not. For my dissertation, I have done a lot of quality checks within my data, so I found out that the response rate was higher for students and retired persons who have more availabilities during daytime, but it was not depending on income. The good message here is that participation is probably, at least in my study, not solely driven by monetary incentives, but rather by how much time people have. You could also think about why you send out a questionnaire over using a sensor to track behavior. If you can simply track behavior instead of asking about it, you should probably go for tracking it. But there will always be questions that cannot be tracked or observed, and for those you should then use a questionnaire.

Lukas Otto: There are certain rules of thumb of how you keep people participating in these studies, which is something that I wish I knew when first starting in the field. Flexible incentives are good at keeping participants in a study. Ethical and privacy issues are big, especially when thinking about collecting so much data from people. Through GPS data, app usage, and some of the survey data, people are easily identifiable. This is a huge issue when working with mobile data. And of course, when you think about tracking, you can have very private and sensitive information that could potentially be shared by the participants and measured through your app.

Ethical and privacy issues are big, especially when thinking about collecting so much data from people. Through GPS data, app usage, and some of the survey data, people are easily identifiable.

We are thinking a lot about this, together with the legal officer at GESIS. For the service itself, we are giving the responsibility to the researchers. While we are the providers of the infrastructure, researchers need to contact their own Institutional Review Boards (IRB) and legal officers to make sure that their research is ethically and legally correct. However, this is something we also think about and that we want to make easier for researchers to handle.

Mareike Wieland: One additional thought on incentivizing or compensating participants: In these intense longitudinal studies, it is even more important to build a viable research alliance or a research partnership with the participants, making it really clear that they understand how important their participation is. My experience is that if you really are in contact with your participants and really value their time and effort, then data quality increases a lot. I think this is underestimated because a lot of researchers simply want to have data. I would advocate for more respect for the time of the participants, thinking about this partnership as another type of incentive strategy.

GESIS: Do you see any sort of misperceptions or misunderstandings with smartphone-based data collections? Lukas, you mentioned IRBs and that maybe sometimes they do not have the best grasp of exactly what these collections are doing and how to judge them. Do you see such types of misperceptions in the research community?

Lukas Otto: People overestimate what is possible, especially when it comes to tracking. While theoretically a lot may be possible, practically getting content out of apps is really hard. This is a common misperception of participants, but sometimes also with researchers that are not so close to the smartphone-based research community.

Mareike Wieland: For us, it is important that social scientists with a more traditional survey approach realize that the smartphone is not just a new device for web surveys, but that it enables – and to a certain extent demands – completely new research designs. We will not solve the problem of small samples and high incentive costs. However, the idiosyncratic approach opens up granular and ecologically valid insights into people's thoughts and behaviors that we did not have before. Research traditions from survey research, psychology and, with behavioral data, computational sciences flow together here. This means that a lot of groundwork is ahead, but it also provides incredible opportunities.

For me, this is a missed potential: thinking about worldwide data collections.

Lukas Otto: I think especially for the computational social science community there is a 'missed opportunity'. When we think about whom we have studied in the past, we have mostly studied Western and American samples. Since the smartphone is used so broadly in

the whole world, it is no longer impossible to do, e.g., research on farmers in Uzbekistan – and I am not making this example up, it is a real example. This has been hard for traditional social science research. But even with computational approaches; how do you get a farmer from Uzbekistan to Twitter? For me, this is a missed potential: thinking about worldwide data collections.

GESIS: If you had the chance to really focus on a very big, substantive research question that you think could be answered using smartphone-based data collections, what would that be?

Mareike Wieland: I always dream about – and we already started with that – the intervention-based approach, using smartphones or mobile research not only to measure what people are doing, but also to influence how they behave. This is where new challenges for scientific research arise. How much influence should we have on people or their behavior? Relatedly, I would really like to see a deeper integration of the digital world, represented through media or social network usage, with the physical world, with locations and spaces in it. Up to a point where we could do actual citizen science, using the smartphone as an entry point for studying real world problems, based on the situative and spatial needs of people. One way of doing this would be to let certain behaviors, certain locations people are at, or a specific time be the trigger for a specific research regime. Taking it a step further, participants could even inform us about on what they would want to have research and how they would like to participate. This is really broad and vague, but it is always something that is my life goal – to do research with people and not about people.

Going beyond surveys and the usual suspects for sensors and tracking: something like eye movements, implicit measures, and reaction times – that would be cool.

Lukas Otto: My first answer would also have been on the interventional part. The ‘match made in heaven’ is not only about combining the mobile digital behavioral data with the survey data, but also about integrating interventions into mobile digital behavioral data. We could have moment- or situation-aware interventions like, when you leave the house, your phone would ask ‘yeah, well, did you close the window?’, or if you go to the supermarket and your phone goes ‘remember you wanted to do the veggie week, why did you buy a salami?’ – I think this is a really good idea for future research.

The second thing I am thinking about is trying to figure out what the next big thing is going to be. We have mobile surveys that have certain advantages. We have mobile digital behavioral data, like app usage data, like GPS data. What is the next big thing? Maybe five years from now, we have even more integration of seemingly old methods. Just one

example: eye movements – every phone has a camera that is constantly pointing towards you. This could be used for some implicit measures of reaction times. This has already been done a little bit, where people measured reaction times and then could say something about the mental processes that were going on while using the phone. This is something – half dream, half realistic – that I would really like to integrate into research. Going beyond surveys and the usual suspects for sensors and tracking: something like eye movements, implicit measures, and reaction times – that would be cool.

GESIS: Are there any types of resources, for example books, packages, libraries, tutorials, but also unconventional stuff like Twitter threads, that you would suggest for people to investigate, so that they may learn more about mobile-based data collections?

Mareike Wieland: There is a freely available handbook that provides a holistic and up-to-date overview [2]. I also can recommend one specific paper from the communication science community because it is suggesting reporting standards for different types of (mobile) experience sampling studies [3]. Another paper that I think is really helpful – especially when designing a mobile study that makes use of both survey and sensor data – is one by Niels van Berkel and Vassilis Kostakos [4].

Lukas Otto: There is one book that is the standard for experience sampling research. It is a bit older – from 2013 – but I think it is really good [5]. Some of our questions are rather new, so they are not yet covered in books, but what I always recommend is following some of the institutes that are doing research in the area, for example the Tilburg Experience Sampling Center.

GESIS: Thank you very much for the interview, Mareike and Lukas!

References

- 1 Sugie, N. F. (2018). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods & Research*, 47(3), 458-491. <https://doi.org/10.1177/0049124115626176>
- 2 Myin-Germeys, I., & Kuppens, P. (Eds.). (2022). *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed.). Leuven: Center for Research on Experience Sampling and Ambulatory Methods Leuven.

- 3 Schnauber-Stockmann, A., & Karnowski, V. (2020). Mobile devices as tools for media and communication research: A scoping review on collecting self-report data in repeated measurement designs. *Communication Methods and Measures*, 14(3), 145-164.
- 4 van Berkel, N., & Kostakos, V. (2021). Recommendations for conducting longitudinal experience sampling studies. In E. Karapanos, J. Gerken, J. Kjeldskov, & M. B. Skov (Eds.), *Advances in longitudinal HCI research* (pp. 59-78). Springer. https://doi.org/10.1007/978-3-030-67322-2_4
- 5 Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Suggested citation

Otto, L., Wieland, M. (2024): *Expert Insights into Collecting Data with Mobile applications. An Interview with Lukas Otto and Mareike Wieland* (=GESIS Guides to Digital Behavioral Data, 18). Cologne: GESIS – Leibniz Institute for the Social Sciences.

Series editors

Danica Radovanović, Maria Zens, Katrin Weller, Claudia Wagner

Publisher

GESIS Leibniz-Institute for the Social Sciences, Cologne

License

Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)