




# Beiwe: A data collection platform for high-throughput digital phenotyping

Jukka-Pekka Onnela<sup>\*1</sup>, Caleb Dixon<sup>2</sup>, Keary Griffin<sup>3</sup>, Tucker Jaenicke<sup>2</sup>, Leila Minowada<sup>2</sup>, Sean Esterkin<sup>2</sup>, Alvin Siu<sup>2</sup>, Josh Zagorsky<sup>2</sup>, and Eli Jones<sup>2</sup>

<sup>1</sup> Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University <sup>2</sup> Zagaran, Inc. <sup>3</sup> Rocket Farm Studios

DOI: [10.21105/joss.03417](https://doi.org/10.21105/joss.03417)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

**Editor:** [Frederick Boehm](#) 

## Reviewers:

- [@Tam-Pham](#)
- [@erik-whiting](#)

**Submitted:** 23 April 2021

**Published:** 12 December 2021

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Beiwe is a high-throughput data collection platform for smartphone-based digital phenotyping. It has been in development and use since 2013. Beiwe consists of two native front-end applications: one for Android (written in Java and Kotlin) and one for iOS (written in Swift and Objective-C). The Beiwe back-end, which is based on Amazon Web Services (AWS), has been implemented primarily in Python 3.6, but it also makes use of Django for ORM and Flask for API and web servers. It uses several AWS services, such as S3 for flat file storage, EC2 virtual servers for data processing, Elastic Beanstalk for orchestration, and RDS for PostgreSQL database engine. Most smartphone applications use software development kits (SDKs) that generate unvalidated behavioral summary measures using closed proprietary algorithms. These applications do not meet the high standards of reproducible science, and often require researchers to modify their scientific questions based on what data happens to be available. In contrast, Beiwe collects raw sensor and phone use data, and its data collection parameters can be customized to address specific scientific questions of interest. Collection of raw data also improves reproducibility of studies and enables re-analyses of data and pooling of data across studies. Every aspect of Beiwe data collection is fully customizable, including which sensors to sample, how frequently to sample them, whether to add Gaussian noise to GPS location, whether to use Wi-Fi or cellular data for uploads, how frequently to upload data, specification of surveys and their response options, and skip logic. All study settings are captured in a JSON-formatted configuration file, which can be exported from and imported to Beiwe to enhance transparency and reproducibility of studies.

## Statement of need

**Background.** The phenotype of an organism is a collection of traits, such as enzyme activity, hormone levels, and behavior. Many researchers are increasingly advocating a more substantial role for large scale phenotyping as a route to advances in the biomedical sciences ([Bilder et al., 2009](#); [Delude, 2015](#); [Houle et al., 2010](#); [Robinson, 2012](#)). Of the many different phenotypes, social, behavioral, and cognitive phenotypes are difficult to study due to their temporal and contextual dependence. The standard approach to learning about something as complex as human behavior is to use surveys, which are cross-sectional, subjective, and often burdensome. The ubiquity and capability of smartphones—when coupled with appropriate data analytic techniques—can be part of the solution. We coined the term Digital Phenotyping to refer to the “moment-by-moment quantification of the individual-level human phenotype *in situ* using

<sup>\*</sup>Principal Investigator and corresponding author.

data from personal digital devices, in particular smartphones" ([Onnela & Rauch, 2016](#); [Torous et al., 2016](#)). We developed Beiwe specifically for use in smartphone-based digital phenotyping research. In addition to enabling more objective measurement of existing phenotypes, the approach can also give rise to entirely new phenotypes.

**State of the field.** Social and behavioral phenotypes have traditionally been studied using either self-administered or investigator-administered surveys in research settings and self-administered or clinician-administered surveys in clinical settings. For example, the Amyotrophic Lateral Sclerosis Functional Rating Scale - Revised (ALSFRS-R) includes 12 questions (items) each scored on 0 (no function) to 4 (full function) scale, and it has been used for both diagnosing patients and for measuring disease progression ([Mora, 2017](#)). In observational studies and clinical trials, it may be administered every six weeks, with smaller within-subject standard deviation in the score if self-administered by the patient on the smartphone rather than administered by a clinician at the clinic on paper ([Berry et al., 2019](#)). To eliminate recall bias, some of these items may be actually measured objectively in free-living settings. For example, two of the items in ALSFRS-R are related to physical activity: walking (Item 8) and climbing stairs (Item 9), both of which can be estimated using smartphone accelerometer and gyroscope data ([Straczekiewicz et al., 2021](#)).

The development of Beiwe was driven by several key considerations: (1) Beiwe is open source software under a permissive 3-clause BSD license; (2) it has been developed and continues to be maintained by professional software engineers to ensure high quality of code base; (3) it supports both Android and iOS devices to allow access to nearly all smartphones; (4) it collects raw sensor data to allow reproducible research and pooling and re-analyses of data; (5) it is compliant with HIPAA for maintaining the privacy of individually identifiable information; (6) it enables full configurability to accommodate the data collection needs of different studies; (7) it emphasizes study replicability and reproducibility by capturing all data collection settings in a JSON configuration file that can be imported into / exported from Beiwe; (8) it uses scalable cloud-based architecture.

Digital phenotyping is related to sensing in computer science. There is a Wikipedia page that compares different mobile phone sensing platforms ([Mobile Phone Based Sensing Software, 2021](#)), but this page appears to be out of date. The comparison includes 26 pieces of software, but these platforms are mostly not intended for digital phenotyping, which among others requires continuous (or essentially continuous) collection and banking of raw passive data. In addition, for representative cohorts and equitable participation in research, it is important to support both Android and iOS users given the socioeconomic differences across their user base; 14 of these platforms support both. To support reproducibility and replicability, being open source is important, which leaves 8 software systems. Some of these systems have not been updated in years, which leaves us with 4 pieces of software: AWARE, Beiwe, EARS, and mindLamp. As far as we know, AWARE has not been used in biomedical research; EARS is not actually open source as it has significant portions of code redacted ("[EARS-iOS-Public, 2020](#)"), although the Wikipedia site has it listed as open source. Finally, mindLamp appears to have been released fairly recently; although it can collect some passive data, it appears to be geared mostly towards clinical use, e.g., it allows patients to track medication, keep a journal, and perform guided meditations.

An important development for the field has been the introduction of software development kits (SDKs) for smartphones, such as Apple's ResearchKit and Google's Research-Stack, which has facilitated writing of software for these devices. Use of prepackaged software however limits what types of data can be collected, which then limits the types of analyses that may be performed ([Onnela, 2021](#)). For example, Apple's ResearchKit does not support background sensor data collection ([ResearchKit, 2018](#)); Apple's HealthKit supports background sensor data collection for selected sensors only ([Apple Developer, 2020b](#)); and the Core Motion framework makes it possible to collect raw accelerometer data in the background but only for up to 12 hours at a time ([Apple Developer, 2020a](#)). The algorithms that underlie HealthKit metrics, such as step count ([Apple Developer, 2020c](#)), appear all to be proprietary. The

use of closed algorithms, which may change at any time without notice, makes it hard or impossible to compare data collected at different times (using possibly different versions of these algorithms) or data collected using different SDKs.

**Workflow.** The Beiwe platform is used primarily in health research settings to collect active and passive data. When using open-source Beiwe, the first step is to deploy the AWS-based system back-end. While historically it's been possible to use either a single server deployment or a scalable server cluster deployment, currently only the server cluster deployment is documented and supported. It's worth pointing out that the back-end deployment is non-trivial, and among other things it requires creating an AWS account, setting up a user with sufficient permissions, generating credentials for programmatic access, obtaining a domain name, and creating a Sentry account for monitoring errors. Once deployed, researchers can then login to the web-based Beiwe portal where they create a study, which includes specifying surveys and configuring passive data collection settings. Note that a single back-end can support tens or even hundreds of studies each with their own surveys and passive data collection settings. Investigators then create a collection of Beiwe user IDs, which are randomly generated 8-character strings (e.g., abcd1234), and distribute them to study subjects together with their temporary passwords. Subjects then download the Beiwe2 smartphone application from Apple App Store (iOS) or Google Play Store (Android) and enter their Beiwe ID, password, and the name of the study server (e.g., my.beiwe.org). The main goal of Beiwe is to collect high-throughput passive data in the background, and as such the idea is for the application to intervene as little as possible in the daily life of the subject. Therefore, the only time the user would actively use the Beiwe smartphone application is when entering surveys or contributing audio diary entries; at all other times the application is simply running in the background. Note that Beiwe never asks the subject for her name, phone number, or any other identifier; the only piece of information that links a user to the Beiwe account is the Beiwe user ID.

**Data elements.** The Beiwe platform can collect both active data (subject input required) and passive data (subject input not required). Currently supported active data types for both Android and iOS are text surveys and audio diary entries and their associated metadata. Passive data can be further divided into two groups: phone sensor data and phone logs. Beiwe collects raw sensor data and raw phone logs, which is absolutely crucial in scientific settings, yet this point remains underappreciated. Relying on generic SDKs for data collection is convenient but for many reasons ineffective: the data generated by different SDKs are not comparable; the algorithms used to generate data are proprietary and hence unsuitable for reproducible research; new data summaries cannot be implemented post data collection; the composition of the metrics collected by SDKs changes in time, making it difficult or impossible to make comparisons across subjects when they are enrolled at different points in time; and data cannot be pooled across studies for later re-analyses or meta-analyses. Currently supported passive data types are the following: accelerometer, gyroscope, magnetometer, GPS, call and text message logs on Android devices (metadata only, no content), proximity, device motion, reachability, Wi-Fi, Bluetooth, and power state. For each sensor, such as GPS, data collection alternates between an on-cycle (data collected) and an off-cycle (data not collected); logs are collected without sampling if their collection is specified. The investigators specify what data is collected based on the scientific question: for example, a study on mobility might choose to collect accelerometer and gyroscope data used in human activity recognition. The text that appears within the application is also customizable for each study. Data streams that contain identifiers, such as phone numbers in communication (call and text message) logs, are anonymized on the device; the “fuzzy” GPS feature if enabled adds randomly generated noise to the GPS coordinates on the device. Finally, study meta settings are also customizable, and include items such as frequency of uploading data files to the back-end (typically 1 hour) and duration before auto logout from the application (typically 10 minutes).

**Privacy and security.** All Beiwe data are encrypted while stored on the phone awaiting upload and while in transit, and they are re-encrypted for storage on the study server while at rest. More specifically, during study registration the platform provides the smartphone app with the

public half of a 2048-bit RSA encryption key. With this key the device can encrypt data, but only the server, which has the private key, can decrypt it. Thus, the Beiwe application cannot read its own data that it stores temporarily, and therefore there is no way for a user (or anyone else) to export the data. The RSA key is used to encrypt symmetric Advanced Encryption Standard (AES) keys for bulk encryption. These keys are generated as needed by the app and must be decrypted by the study server before data recovery. Data received by the cloud server is re-encrypted with the study master key provided and then stored on the cloud. Some of the collected data contain identifiers: communication logs on Android devices contain phone numbers, and Wi-Fi and Bluetooth scans contain media access control (MAC) address. If the study is configured to collect these data, the identifiers in them are anonymized on the phone, and only anonymized versions of the data are uploaded to the back-end server. Briefly, the Beiwe front-end application generates a unique cryptographic code, called a salt, during the Beiwe registration process, and then uses the salt to encrypt phone numbers and other similar identifiers. The salt never gets uploaded to the server and is known only to the phone for this purpose. Using the industry standard SHA-256 (Secure Hash Algorithm) and PBKDF2 (Password-Based Key Derivation Function 2) algorithms, an identifier is transformed into an 88-character anonymized string that can then be used in data analysis.

**Use cases.** At the time of writing, Beiwe is or has been used in tens of scientific studies on three continents across various fields, and there are likely several additional studies we are not aware of. Smartphone-based digital phenotyping is potentially very promising in behavioral and mental health (Onnela & Rauch, 2016), and new research tools like Beiwe are especially needed in psychiatry (Torous et al., 2016), where in the context of schizophrenia it has been used to predict patient relapse (Barnett et al., 2018), compare passive and active estimates of sleep (Staples et al., 2017), and characterize the clinical relevance of digital phenotyping data quality (Torous et al., 2018). The platform has also been used to assess depressive symptoms in a transdiagnostic cohort (AM Pellegrini & Onnela, 2021) and to capture suicidal thinking during the COVID-19 pandemic (R Fortgang & Nock, 2020). There is an increasing amount of research on the use of Beiwe in neurological disorders, such as in the quantification of ALS progression (Berry et al., 2019) and behavioral changes in people with ALS during the COVID-19 pandemic (Beukenhorst et al., 2021). The platform has been used in the context of cancer to assess postoperative physical activity among patients undergoing cancer surgery (Panda et al., 2021), to capture novel recovery metrics after cancer surgery (Panda, Solsky, Huang, et al., 2020), to enhance recovery assessment after breast cancer surgery (Panda, Solsky, Hawrusik, et al., 2020), and to enhance cancer care (Wright et al., 2018). Digital phenotyping and Beiwe have also been applied to quantifying mobility and quality of life of spine patients (Cote et al., 2019) and to study psychosocial well-being of individuals after spinal cord injury (Mercier et al., 2020).

## Acknowledgements

The Principal Investigator, Jukka-Pekka Onnela, is extremely grateful for his NIH Director's New Innovator Award in 2013 (DP2MH103909) for enabling the crystallization of the concept of digital phenotyping and the construction of the Beiwe platform. He is also grateful to the members of the Onnela Lab. The authors also wish to acknowledge the contributions of the following individuals at Zagar: Aaron Klein, Kevin Fan, Christopher McCarthy, Dor Samet, Alicia Fan, Rebecca Magazine Malamud, Jacob Klingensmith, and Benjamin Zagorsky.

## References

AM Pellegrini, P. S., EJ Huang, & Onnela, J. (2021). Estimating longitudinal depressive symptoms from smartphone data in a transdiagnostic cohort. *Brain and Behavior*.

- Apple Developer. (2020a). *Core motion documentation*. Accessed November 10, 2021 [Online]. <https://developer.apple.com/documentation/coremotion/cmsensorrecorder/1615987-recordaccelerometer>
- Apple Developer. (2020b). *SensorKit documentation*. Accessed November 10, 2021 [Online]. <https://developer.apple.com/documentation/sensorkit/srsensor#3681604>
- Apple Developer. (2020c). *SensorKit documentation*. Accessed November 10, 2021 [Online]. <https://developer.apple.com/documentation/healthkit/hkquantitytypeidentifier/1615548-stepcount>
- Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J.-P. (2018). Relapse prediction in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology*, 43(8), 1660–1666. <https://doi.org/10.1038/s41386-018-0030-z>
- Berry, J. D., Paganoni, S., Carlson, K., Burke, K., Weber, H., Staples, P., Salinas, J., Chan, J., Green, J. R., Connaghan, K., Barback, J., & Onnela, J.-P. (2019). Design and results of a smartphone-based digital phenotyping study to quantify ALS progression. *Annals of Clinical and Translational Neurology*, 6(5), 873–881. <https://doi.org/10.1002/acn3.770>
- Berry, J. D., Paganoni, S., Carlson, K., Burke, K., Weber, H., Staples, P., Salinas, J., Chan, J., Green, J. R., Connaghan, K., Barback, J., & Onnela, J.-P. (2019). Design and results of a smartphone-based digital phenotyping study to quantify ALS progression. *Annals of Clinical and Translational Neurology*, 6(5), 873–881. <https://doi.org/10.1002/acn3.770>
- Beukenhorst, A. L., Collins, E., Burke, K. M., Rahman, S. M., Clapp, M., Konanki, S. C., Paganoni, S., Miller, T. M., Chan, J., Onnela, J.-P., & others. (2021). Smartphone data during the COVID-19 pandemic can quantify behavioral changes in people with ALS. *Muscle & Nerve*, 63(2), 258–262. <https://doi.org/10.1002/mus.27110>
- Bilder, R. M., Sabb, F., Cannon, T., London, E., Jentsch, J., Parker, D. S., Poldrack, R. A., Evans, C., & Freimer, N. (2009). Phenomics: The systematic study of phenotypes on a genome-wide scale. *Neuroscience*, 164(1), 30–42. <https://doi.org/10.1016/j.neuroscience.2009.01.027>
- Cote, D. J., Barnett, I., Onnela, J.-P., & Smith, T. R. (2019). Digital phenotyping in patients with spine disease: A novel approach to quantifying mobility and quality of life. *World Neurosurgery*, 126, e241–e249. <https://doi.org/10.1016/j.wneu.2019.01.297>
- Delude, C. M. (2015). Deep phenotyping: The details of disease. *Nature*, 527(7576), S14–S15. <https://doi.org/10.1038/527S14a>
- EARS-iOS-public. (2020). In *GitHub repository*. GitHub. <https://github.com/C4DMH/EARS-iOS-Public>
- Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: The next challenge. *Nature Reviews Genetics*, 11(12), 855–866. <https://doi.org/10.1038/nrg2897>
- Mercier, H. W., Hamner, J. W., Torous, J., Onnela, J. P., & Taylor, J. A. (2020). Digital phenotyping to quantify psychosocial well-being trajectories after spinal cord injury. *American Journal of Physical Medicine & Rehabilitation*, 99(12), 1138–1144. <https://doi.org/10.1097/phm.0000000000001506>
- Mobile phone based sensing software*. (2021). *Wikipedia*. Accessed November 10, 2021 [Online]. [https://en.wikipedia.org/wiki/Mobile\\_phone\\_based\\_sensing\\_software](https://en.wikipedia.org/wiki/Mobile_phone_based_sensing_software)
- Mora, J. S. (2017). Edaravone for treatment of early-stage ALS. *The Lancet Neurology*, 16(10), 772. [https://doi.org/10.1016/S1474-4422\(17\)30289-2](https://doi.org/10.1016/S1474-4422(17)30289-2)
- Onnela, J.-P. (2021). Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology*, 46(1), 45–54. <https://doi.org/10.1038/s41386-020-0771-3>



- Onnela, J.-P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, 41(7), 1691–1696. <https://doi.org/10.1038/npp.2016.7>
- Panda, N., Solsky, I., Cauley, C. E., Lipsitz, S., Desai, E. V., Huang, E. J., Benjamin, E. M., Lubitz, C. C., Onnela, J.-P., & Haynes, A. B. (2021). Smartphone-based assessment of preoperative decision conflict and postoperative physical activity among patients undergoing cancer surgery: A prospective cohort study. *Annals of Surgery*. <https://doi.org/10.1097/SLA.0000000000004487>
- Panda, N., Solsky, I., Hawrusik, B., Liu, G., Reeder, H., Lipsitz, S., Desai, E. V., Lowery, K. W., Miller, K., Gadd, M. A., & others. (2020). Smartphone global positioning system (GPS) data enhances recovery assessment after breast cancer surgery. *Annals of Surgical Oncology*, 1–10. <https://doi.org/10.1245/s10434-020-09004-5>
- Panda, N., Solsky, I., Huang, E. J., Lipsitz, S., Pradarelli, J. C., Delisle, M., Cusack, J. C., Gadd, M. A., Lubitz, C. C., Mullen, J. T., & others. (2020). Using smartphones to capture novel recovery metrics after cancer surgery. *JAMA Surgery*, 155(2), 123–129. <https://doi.org/10.1001/jamasurg.2019.4702>
- R Fortgang, A. M., SB Wang, & Nock, M. (2020). Increase in suicidal thinking during COVID-19. *Clinical Psychological Science*. <https://doi.org/10.1177/2167702621993857>
- ResearchKit. (2018). *ResearchKit framework programming guide*. Accessed November 10, 2021 [Online]. <http://researchkit.org/docs/docs/Overview/GuideOverview.html>
- Robinson, P. N. (2012). Deep phenotyping for precision medicine. *Human Mutation*, 33(5), 777–780. <https://doi.org/10.1002/humu.22080>
- Staples, P., Torous, J., Barnett, I., Carlson, K., Sandoval, L., Keshavan, M., & Onnela, J.-P. (2017). A comparison of passive and active estimates of sleep in a cohort with schizophrenia. *NPJ Schizophrenia*, 3(1), 1–6. <https://doi.org/10.1038/s41537-017-0038-0>
- Strackiewicz, M., James, P., & Onnela, J.-P. (2021). A systematic review of smartphone-based human activity recognition methods for health research. *NPJ Digital Medicine*, 4(1), 1–15. <https://doi.org/10.1038/s41746-021-00514-4>
- Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, 3(2), e16. <https://doi.org/10.2196/mental.5165>
- Torous, J., Staples, P., Barnett, I., Sandoval, L. R., Keshavan, M., & Onnela, J.-P. (2018). Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *NPJ Digital Medicine*, 1(1), 1–9. <https://doi.org/10.1038/s41746-018-0022-8>
- Wright, A. A., Raman, N., Staples, P., Schonholz, S., Cronin, A., Carlson, K., Keating, N. L., & Onnela, J.-P. (2018). The HOPE pilot study: Harnessing patient-reported outcomes and biometric data to enhance cancer care. *JCO Clinical Cancer Informatics*, 2, 1–12. <https://doi.org/10.1200/cci.17.00149>