Studies: A Managed Solution for Standardized, Scalable, Privacy-First Smartphone Sensor Data Collection for Indoor Positioning Research

Kelvin Tsz Hei Choi[1,*,†]

[1]Sensor Logger Email, sensorlogger@tszheichoi.com

Abstract

Smartphone-based indoor positioning research demands high-quality, time-synchronized sensor streams under diverse environmental and hardware conditions, yet many researchers face challenges in their data collection campaigns owing to device heterogeneity, sampling inconsistency, consent management, and privacy compliance. We introduce Studies, a fully managed, end-to-end Software as a Service (SaaS) framework built on top of the popular Sensor Logger app (iOS, Android, WatchOS, WearOS). Studies formalizes roles for investigators and participants, automates sensor configuration distribution across platforms, orchestrates data capture, and integrates built-in consent and customizable questionnaires. Sensor Logger and Studies have already supported numerous research efforts presented at previous IPIN conferences, including pedestrian trajectory reconstruction via bidirectional Kalman filtering [1] and magnetometer calibration during SLAM [2]. In this paper, we present the system architecture, cross-platform implementation, managed workflows, and real-world case studies, demonstrating how Studies enables researchers to collect the right data at the right time—while ensuring participant privacy and data consistency.

Keywords

smartphone-based, positioning, crowd-sourcing, reproducible research, privacy compliance, managed data workflows, data collection

1. Introduction

Indoor positioning research relies on rich, multi-sensor datasets to achieve high accuracy in complex environments. However, researchers frequently encounter challenges including device heterogeneity and sampling variation [3], consent and privacy concerns [4], and manual workflow overhead [5], especially in large-scale studies. To address these issues, we introduce Studies, a service-oriented framework within the Sensor Logger app that helps researchers specify precise sensor suites, orchestrate recordings, and automate secure data uploads, allowing focus on experimental design and analysis. Studies has already gained popularity and recognition in the IPIN community, including enhanced pedestrian trajectory reconstruction [1] and SLAM magnetometer calibration [2], demonstrating its adaptability and impact.

In Section 2, we first outline the key challenges researchers face in data collection campaigns. In Section 3, we review some existing solutions for data logging in research context. Section 4 introduces the Sensor Logger app, and from there we dive into the Studies architecture in Section 5 to demonstrate how it helps reduce friction for researchers. Finally, in Section 6, we showcase how Studies have been used in practice across academia already.

2. Challenges for Researchers

Collecting reliable, large-scale smartphone sensor data involves overcoming numerous technical, logistical, and ethical hurdles. Prior tools lack scalable, integrated workflows for crowd-sourced sensor

IPIN-WCAL 2025: Workshop for Computing & Advanced Localization at the Fifteenth International Conference on Indoor Positioning and Indoor Navigation, September 15–18, 2025, Tampere, Finland

*Corresponding author.

data collection with privacy control [6]. Existing crowd-sourcing platforms also often require custom development, which hinders reproducibility [7]. In the following, we review the principal challenges researchers face.

2.1. Heterogeneous Devices

Smartphones differ widely in sensor quality, sampling APIs, and coordinates / units definitions, leading to non-uniform data that undermines reproducibility and comparability, especially in large-scale studies [12, 13]. Researchers often resort to procuring a uniform device pool, adding tens of thousands of dollars in hardware costs, to ensure consistency, or they resort to writing custom code to mitigate differences, taking time away from valuable research efforts[12].

2.2. Custom Sensing Logic

Many data collection campaigns involve custom sensing logic – whether it is geo-fencing or rules based on time of day or day of week. Incorporating custom sensing logic often necessitates bespoke application development, increasing the technical burden on research teams.

2.3. Ad-Hoc Data Transfer

Crowd-sourced studies often rely on ad-hoc uploads to Google Drive folders or Dropbox links, burdening participants with manual steps [14, 15] and researchers with reminder overhead, let alone keeping track of files and their naming conventions. Moreover, these methods frequently lack robust privacy protections, raising concerns about the secure handling of sensitive participant data.

More advanced research groups may use cloud object stores (e.g., Amazon S3) with buckets via SDKs or REST APIs [16]. However, this comes with engineering overhead [16] and working with devops tools that researchers may not be familiar with. Indeed, managing tens of participants is perhaps manageable with manual oversight, but studies with thousands of contributors will have unbearable clerical bottlenecks, leading to delays in research schedule [17].

2.4. Privacy, Consent, and Ethical Compliance

Smartphone sensor streams (GPS, microphone, Bluetooth) can reveal sensitive personal behaviors, creating high privacy risks [25]. Institutions enforce diverse ethics reviews—e.g., University of Washington's IRB, Stanford's Human Subjects Committee – each with unique consent form requirements, data retention policies, and anonymization standards [20].

2.5. Capturing Rich Contextual Data

While raw sensor data (e.g., accelerometer, magnetometer, GPS) captures physical phenomena, it often lacks essential contextual information such as user intent, environmental conditions, or semantic labels. These are often critical to researchers. To address this, many frequently supplement sensor streams with participant-reported data via additional surveys, daily diaries, or momentary assessments. However, most resort to separate tools like Google Forms or Qualtrics to administer surveys [22], introducing significant friction for participants who must switch between platforms or apps. This added complexity can reduce response rates, break temporal alignment with sensor events, and increase dropout [23]. Furthermore, merging survey data with sensor logs typically requires post-hoc synchronization based on timestamps, which is error-prone—especially when sensor and survey data are collected on separate threads or apps with differing clock sources.

3. Existing Solutions

A number of mobile sensing platforms have been developed to facilitate research using smartphone and wearable data. We briefly review representative examples and highlighting potential shortfalls with respect to challenges outlined in the previous section.

3.1. AWARE Framework

The AWARE Framework [14] provides a comprehensive SDK and server backend for Android devices, enabling passive data logging of sensors such as accelerometers, location, and app usage. While AWARE includes plugin support and centralized dashboards, it lacks seamless cross-platform support (very limited iOS capabilities), and consent flows must be managed externally.

3.2. RADAR-base

RADAR-base [28] is an open-source platform originally developed for health-related sensing in clinical studies. It offers integration with wearable sensors (e.g., Fitbit, Biovotion) and collects active and passive data using a modular architecture. However, it requires significant DevOps setup, including Kafka and PostgreSQL servers.

3.3. Beiwe

The Beiwe Research Platform [26] emphasizes privacy-preserving smartphone sensing in clinical trials. It supports customizable surveys and sensor capture but is primarily Android-focused and requires institutional review board (IRB) configuration before deployment.

3.4. SensingKit and OpenSensing

Other toolkits such as SensingKit [27] and OpenSensing aim to simplify mobile sensor data access for developers but remain limited in deployment management, standardization, and participant oversight. They serve more as low-level libraries than managed platforms.

3.5. GetSensorData

GetSensorData is an Android-based sensor acquisition tool used in the IPIN community, particularly for competition datasets [29]. It enables straightforward collection of raw smartphone sensor data but lacks integrated participant management, and does not support cross-platform harmonization or built-in privacy and consent workflows.

4. The Sensor Logger App

Sensor Logger is a free, versatile, cross-platform mobile application designed to simplify sensor logging from smartphones, tablets, and wearables. As shown in Figure 1, it supports a comprehensive array of on-device sensors—including accelerometers, gyroscopes, magnetometers, barometers, GPS, audio, and camera streams—as well as device metadata such as battery level, network state, and screen brightness. Of note to the IPIN community is support for WiFi and Bluetooth beacons.

Figure 1 also shows the various post-hoc export formats and live data streaming options via HTTP and MQTT. However, the focus of this paper is the Studies subsystem (highlighted with bolder borders), and will be detailed below.

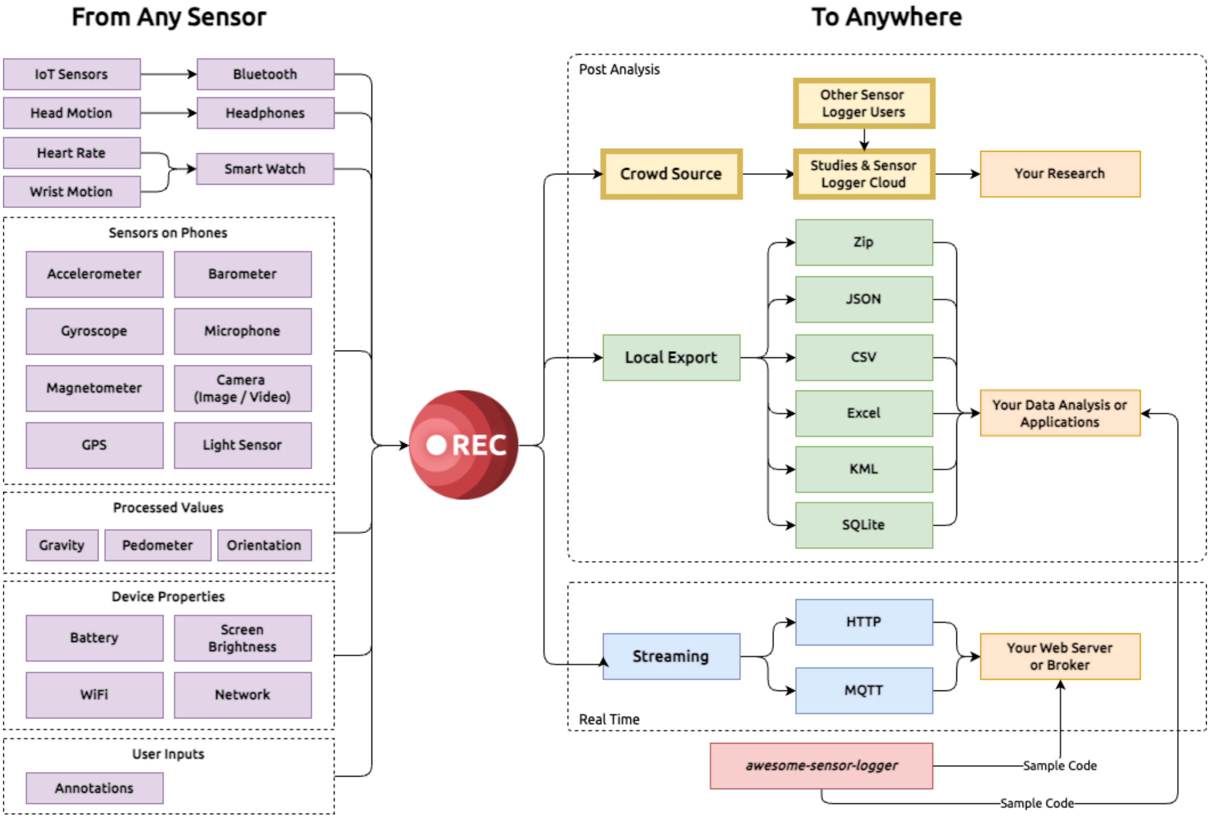

Figure 1: The overall ecosystem of Sensor Logger, illustrating what types of data it can log and where the logged data can be exported or streamed. In particular, the Studies sub-system is highlighted in bold boxes.

5. Studies Architecture

The Studies sub-system is designed to tackle the challenges of running data collection campaigns as outlined in Section 2. Figure 2 shows the outline of the Studies system, which will be explained below.

5.1. Participant–Investigator Model

The corner stone of Studies is the participant-investigator model. In this context, researchers play the role of an Investigator, and Participants contribute data to the research. All parties use the same Sensor Logger app on their respective smartphone throughout the life-cycle of a Study.

As shown in step (A) of figure 2, investigators begin by defining a Study directly within the Sensor Logger app, controlling exactly how and what data should be collected. Each Study configuration includes the following:

- Defined Sensor Suite: Accelerometer, gyroscope, magnetometer, barometer, GPS, audio, camera, microphone, Bluetooth beacons, Wi-Fi and any custom sensor decoder plugins.
- Defined Sampling Rules: sampling frequencies, time based or geo-fence/location based conditional recording, sensor based triggers. Investigators can test out these rules on their own device and iterate until satisfaction.
- Export Formats: JSON, CSV, Excel, KML, SQLite etc that suits the researcher's downstream analysis scripts.
- Metadata & Privacy: privacy statements, approved consent text, contact details. To help investigators, templates are provided as part of the creation workflow.

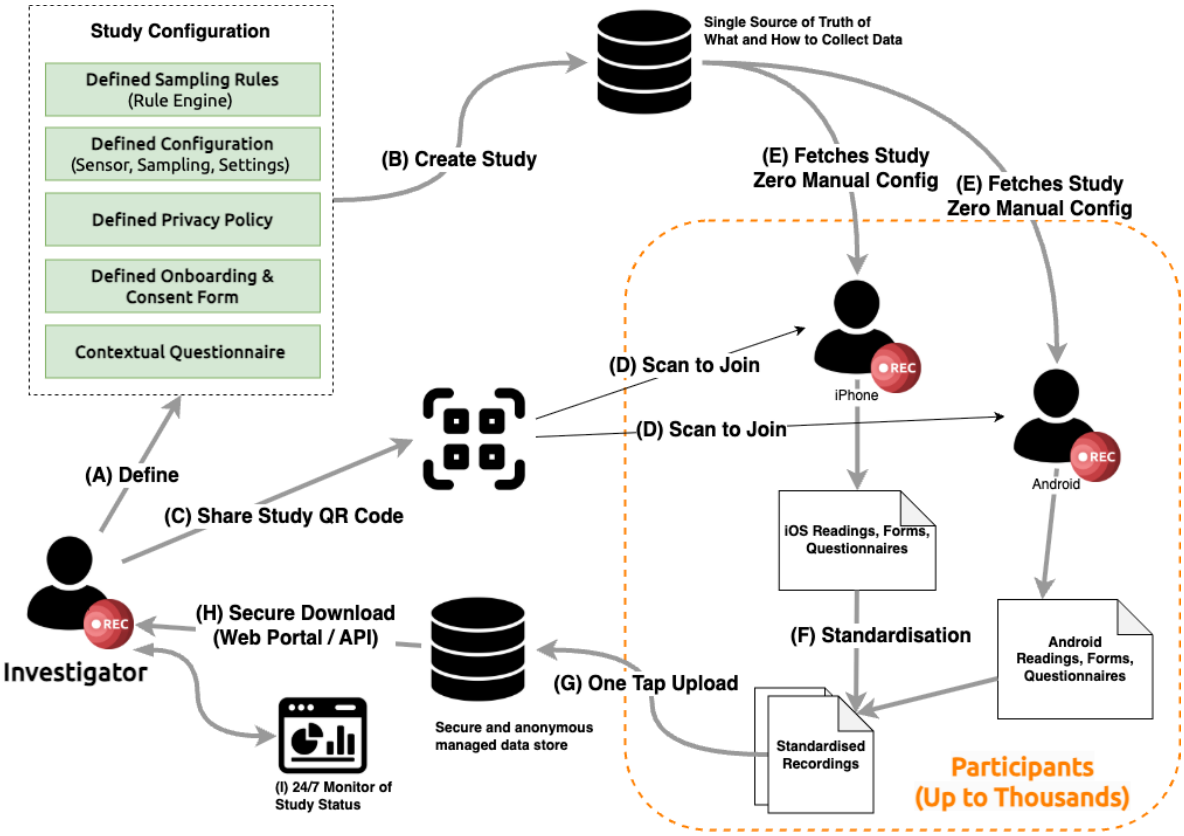

Figure 2: Outline of the Studies system within Sensor Logger, demonstrating the workflow and the roles of Investigators and Participants.

- Contextual Questionnaire: Study-specific questionnaire templates (text, numeric, multiple-choice, signature fields), which can be shown on joining a study and on the ending of each recording session. See figure **??** in the appendix.

The investigator submits this Study configuration to a remote server, also known as Sensor Logger Cloud, as shown in step (B) of figure 2. A backend then validates the Study configuration to ensure integrity and compliance. Thereafter, a single source of truth defining what and how data should be collected exists in the cloud.

5.2. One Tap Join

Once a Study is created, a simple QR code (or link) is generated, shown as step (C) and (D) of figure 2. This QR code points to the aforementioned created Study. Participants can then enroll seamlessly by simply scanning with their own phones. To ensure privacy compliance, all participants can first preview exactly what they will be collecting on behalf of the Study investigator before joining. Depending on the Study configuration, participants will be asked to complete a consent form before joining. Once joined, shown as step (E) of figure 2, each participant automatically receives the exact configuration as defined in the previous section, regardless of platform. They can also join, pause, or switch studies in-app, filter recordings by study, and remain anonymous by design. It is also noted that once joined, the participant's devices do not require further internet connection until they are ready to share their recordings back with the investigator – this is important for participants with limited cellular data.

5.3. Standardization & Harmonization

Since participants may use a variety of recording devices, the raw data collected can differ in units and coordinate systems. For example, as shown in Figure 3, Android's SensorManager and iOS's Core Motion

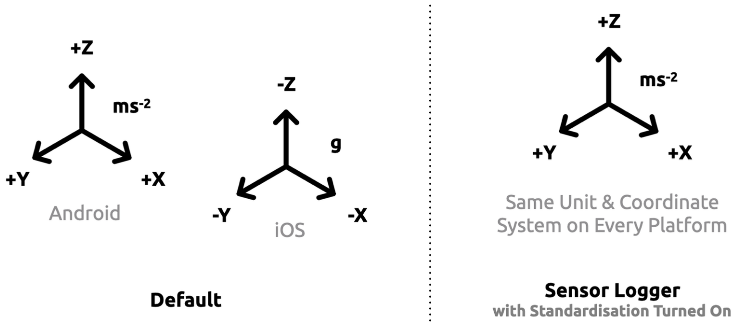

Figure 3: Studies are fully compatible with both iOS and Android devices, ensuring broad participation regardless of the platform. With built-in cross-platform functionality, Sensor Logger automatically resolves any data discrepancies between devices, allowing you to focus on analysis without worrying about format inconsistencies.

define coordinate systems for acceleration data in opposite directions. In particular, note that iOS does not follow the standard right-handed coordinate system due to legacy reasons. Additionally, Android and iOS apply different strategies for location fusion, each making distinct trade-offs between network-based and GNSS-based positioning. Studies employs a unified standardization mechanism to resolve these inconsistencies. This standardization process (Step F in Figure 2) is performed automatically in the background and requires no configuration or understanding from participants. Standardization includes:

- **Unit Normalization**: Converts all measurements to SI units (m, m/s^2, rad/s, μT, hPa); GPS coordinates in decimal degrees.
- **Coordinate System Alignment**: Unified definition of x, y and z coordinate systems for all vector-valued sensor readings.
- **Absolute Reference Frames**: All GPS and fused location data are aligned to the WGS84 geodetic reference system.
- **Timestamp Synchronization**: To resolve misalignment across sensors (e.g., gyroscope vs. GPS) caused by internal platform scheduling jitter or buffer differences, Studies interpolates or aligns samples to the nearest common sampling grid as defined by the investigator. All timestamps are referenced to UNIX epoch in milliseconds and are corrected for known device clock skew where possible.
- **Metadata Standards**: For traceability and reproducibility, all exported data is bundled with device identifiers (model, manufacturer, OS version), sensor vendor details, and platform-specific flags (e.g., fused vs. raw GNSS, estimated accuracy).
- **BLE Decoding**: A core component of Sensor Logger is its open-source initiative to provide a unified interface for Bluetooth Low Energy (BLE) data acquisition and decoding. BLE plays a crucial role in indoor positioning due to its widespread availability and fine-grained proximity sensing capabilities. However, BLE data collection varies significantly between platforms, with differences in scanning APIs, advertisement formats, and signal reporting. Studies fully leverages this unified BLE interface to ensure consistent decoding [30].

For reproducibility, all transformations performed as part of the Standardization are documented in detail in [11]. This ensures full transparency for reproducibility, debugging, and peer review.

5.4. One Tap Upload

With a single tap, participants can upload their recordings securely to Sensor Logger Cloud, via an encrypted connection (Step (G) of figure 2). The backend infrastructure is hosted on industry-grade providers such as Cloudflare and Backblaze, offering distributed object storage with configurable redundancy, automatic deduplication, and secure TLS channels for data in transit and at rest.

Data is stored in the cloud as per a configured data retention policy. During this period, the investigator can download all contributed recordings securely from a web portal or via an API as shown in step (H) of figure 2. Throughout the Study life cycle, the investigator can monitor the progress of the Study through the app (Step (G) of figure 2) – such as how many participants have enrolled and how many recordings have been uploaded.

While the current implementation emphasizes asynchronous, privacy-first uploads to preserve participant control and minimize cellular data usage, Studies has been designed with real-time streaming capabilities in mind. Preliminary internal testing supports low-latency MQTT-based data relay for scenarios requiring continuous monitoring (e.g., infrastructure-aware SLAM, live mobility tracking).

5.5. Data Parsing and Analysis Tools

To facilitate downstream data analysis, Studies exports sensor data in widely-used formats including JSON, CSV, Excel, KML, and SQLite. Recognizing the importance of easy data ingestion for researchers, we provide open-source parsing libraries and example scripts compatible with popular scientific computing environments. For Python users, a repository is available on GitHub [11], which offers utilities for loading, filtering, and synchronizing sensor streams, including handling of the standardization metadata and timestamps.

5.6. Software as a Service

The principle of Studies is Software as a Service (SaaS). In this paradigm, researchers don't configure and manage storage, data transfer and communication infrastructure with participants. Researchers simply declare a study, and pay for the resources (such as storage) they consume [12]. This subscription-based model reduces upfront costs and lowers barriers to entry for smaller labs or pilot studies – often free for sufficiently small studies.

6. Case Studies & Adoption

Studies has been adopted by leading institutions (MIT Media Lab, ETH Zurich Robotics Systems Lab, KU Leuven, University of Duisburg-Essen, University of Cambridge) for applications including:

- Pedestrian trajectory reconstruction with bidirectional Kalman filters [1].
- Retail footfall analytics using BLE/RSSI logs [8].
- SLAM magnetometer calibration across devices [2].
- Mental health biometrics in longitudinal clinical trials [9].
- Agricultural incident detection via multi-sensor wearables [10].

7. Conclusion

We present Studies as a comprehensive, privacy-first service that automates end-to-end sensor data collection for indoor positioning research. The benefits of Studies compared to typical manual workflow for researchers are summarized in table 7. By formalizing roles, standardizing configurations, and offering dynamic scalability, Studies accelerates reproducible experiments. Future work includes extending Studies to real-time streaming contributions, and even more throughout cross-platform harmonization, such as unified sensor fusion algorithms. Anyone can try Studies for free in the Sensor Logger app. Learn more at https://www.tszheichoi.com/studies.

Table 1
Comparison: Typical Manual Workflow vs. Studies Workflow with Sensor Logger

Typical Manual Workflow	Studies Workflow
Researcher writes custom scripts or apps per device/OS.	Investigator declares study parameters in single app.
Create separate consent forms on external platforms (e.g., Google Forms, Qualtrics).	In-app consent flows and questionnaires embedded in the study configuration.
Participants install and configure manually.	One-tap enrollment via QR code or short link applies exact configuration.
Participants record data manually; start/stop reminders require manual follow-up.	Automated background recording with smart rules (time, geofence, motion) and on-device reminders.
Participants upload files via email, Google Drive, or custom S3 scripts.	One-tap, encrypted uploads to cloud; automated retry and reminder logic.
Researcher merges, cleans, and standardizes heterogeneous exports offline.	Data delivered in normalized, SI-unit JSON/CSV; metadata and transformations documented automatically.
Maintain custom server infrastructure or ad-hoc storage buckets.	Fully managed SaaS backend with dynamic scaling, SLAs, and REST/MQTT APIs.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

[1] Y. Dong et al., "Enhanced Pedestrian Trajectory Reconstruction Using Bidirectional Extended Kalman Filter and Automatic Refinement," in *IPIN 2024*, 2024.
[2] I. Vallivaara et al., "Magnetometer Calibration During SLAM," in *IPIN 2024*, 2024.
[3] J. Smith and A. Lee, "Sensor Heterogeneity in Mobile Devices," *Sensors*, 2024.
[4] M. Jones, "Privacy and Consent in Crowdsourced Data," *Sci. Direc.*, 2016.
[5] K. Nguyen, "Manual Workflows in Sensor Data Collection," *MDPI*, 2020.
[6] P. Barsocchi et al., "Challenges in Crowdsourced Indoor Positioning," *Springer*, 2020.
[7] L. Zhao et al., "Reproducibility Barriers in Mobile Sensing," *Eurasip J. Adv. Signal Process.*, 2024.
[8] R. Yonetani et al., "RetailOpt: Opt-in Easy-to-Deploy Trajectory Estimation from Smartphone Motion Data and Retail Facility Information," in *ACM ISWC '24*, 2024.
[9] N. Loecher et al., "Assessing the Efficacy of a Self-Stigma Reduction Mental Health Program with Mobile Biometrics," in *IEEE FG '23*, 2023.
[10] A. J. Etienne et al., "Testing the Feasibility of Commercial Wearables in Agricultural Incident Detection," *J. Agric. Safety Health*, 2024.
[11] T. Choi, "Awesome Sensor Logger," GitHub repository, 2025. [Online]. Available: https://github.com/tszheichoi/awesome-sensor-logger.
[12] O. Vallivaara *et al.*, "Comparative Assessment of Multimodal Sensor Data Quality ...," *Sensors*, 2024.
[13] S. Joo *et al.*, "Ensemble Models for Human Activity Recognition Independent of Device Configuration," *Springer*, 2024.
[14] AWARE Framework, "Use AWARE," 2025. [Online]. Available: https://awareframework.com/use-aware/.
[15] B. Struminskaya *et al.*, "Augmenting Surveys With Data From Sensors and Apps: Opportunities and Challenges," 2020.
[16] M. K. Boehnke *et al.*, "Collecting and Analyzing Smartphone Sensor Data for Health," *ACM*, 2020.
[17] R. Harari *et al.*, "Using Smartphones to Collect Behavioral Data in Psychological Studies," *Psychol. Methods*, 2022.

[18] Wikipedia, “Crowdsensing,” 2024. [Online]. Available: https://en.wikipedia.org/wiki/Crowdsensing.
[19] C. West *et al.*, “Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Nonparticipation Bias,” *Res. Methods*, 2024.
[20] N. Harari *et al.*, “Lessons from First-Generation Smartphone Sensing Studies,” *Social Sci. Comput. Rev.*, 2022.
[21] J. D. Biersdorfer, “Google Drive Privacy,” *New York Times*, 2017.
[22] Struminskaya, B. et al., “Augmenting Surveys With Data From Sensors and Apps: Opportunities and Challenges,” in *Social Science Computer Review*, 2020.
[23] West, C. et al., “Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Nonparticipation Bias,” in *Research Methods*, 2024.
[24] Harari, G. M. et al., “Lessons From First-Generation Smartphone Sensing Studies,” in *Social Science Computer Review*, 2022.
[25] Cisco, “What Is SaaS?,” 2025. [Online]. Available: https://www.cisco.com/c/en/us/products/software/what-is-software-as-a-service-saas.html :contentReference[oaicite:1]index=1
[26] J. Torous, M. V. Kiang, J. Lorme, and J.-P. Onnela, “The Beiwe platform: a high-throughput smartphone-based digital phenotyping tool,” *JMIR Ment. Health*, vol. 3, no. 1, p. e6434, 2016.
[27] C. Perera, P. Jayaraman, A. Zaslavsky, P. Christen, and D. Georgakopoulos, “SensingKit: Evaluating the Sensor Power Consumption in iOS Devices,” in *Proc. IEEE Int. Conf. Intell. Sensors, Sensor Networks Inf. Process. (ISSNIP)*, pp. 1–6, 2015.
[28] M. Van Velthoven, N. Mastellos, and A. Majeed, “RADAR-base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices,” *J. Med. Internet Res.*, vol. 23, no. 7, p. e21830, 2021.
[29] J. Blanco, E. Garcia, and M. L. Perez, “GetSensorData: A Smartphone Sensor Data Acquisition Tool for Indoor Positioning,” *Pervasive and Mobile Computing*, vol. 77, 2022. Available: https://www.sciencedirect.com/science/article/pii/S2352711022001121
[30] K. T. H. Choi, “Sensor Logger BLE Decoder,” GitHub repository, 2025. [Online]. Available: https://github.com/tszheichoi/sensor-ble

A. Appendix

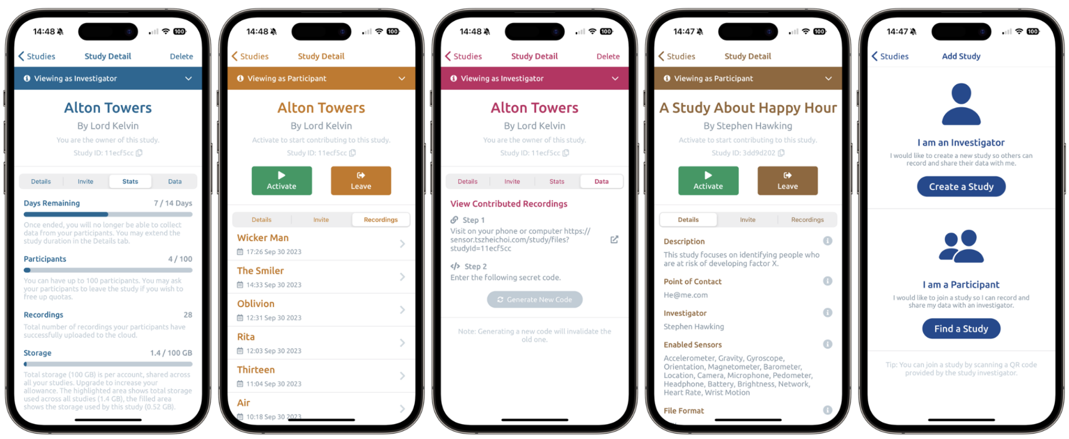

Figure 4: Example screenshots of how Studies look like for Investigators and Participants.