

Toward Standardized Representation of Wearable and Patient-Generated Health Data in OMOP-CDM: A Framework for Community Discussion

Somayeh Abedian¹

Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria
Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Canada

Background

Patient-generated health data (PGHD) collected through wearable devices, mobile health applications, patient-reported outcome measures (PROMs), and patient-reported experience measures (PREMs) are increasingly contributing to observational research and real-world evidence generation(1). These data provide longitudinal information regarding health status, symptoms, daily activities, treatment outcomes, and health behaviors that are often unavailable in conventional healthcare records.

The Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) has become a widely adopted framework for harmonizing healthcare data and enabling large-scale observational analytics across heterogeneous data sources (2). While OMOP-CDM has demonstrated strong support for clinical and administrative healthcare data, representing wearable and patient-generated data remains challenging. Continuous sensor streams, device metadata, provenance information, data quality indicators, and contextual information associated with patient-reported data introduce requirements that are not fully addressed by existing OMOP structures (3). While large-scale initiatives like the *All of Us Research Program* have successfully incorporated certain consumer wearable data (e.g., Fitbit) and participant surveys into custom OMOP implementations, standardizing continuous, high-frequency device streams and complex survey provenance remains a critical gap across the broader OHDSI ecosystem (4,5).

In May 2026, we published a conceptual extension model examining these representational gaps and proposing a modular framework for integrating wearable and patient-generated health data into OMOP-CDM (3). The work introduced a set of complementary entities intended to support provenance-aware, traceable, and analytically useful representation of PGHD. Rather than reproducing the published work, this submission presents the framework as a discussion artifact and invites feedback from the OHDSI community regarding future approaches for representing wearable and patient-generated health data within OMOP-CDM.

Methods

Requirements were derived from a literature review and three major PGHD initiatives: the OHDSI PGHD Working Group (2018–2022), the IMI-FAITH project, and the NIH All of Us

¹Somayeh Abedian, Ludwig Boltzmann Institute for Digital Health and Prevention, Austria; E-mail: Somayeh.Abedian@lbg.ac.at

Research Program. The framework originated from a structured assessment of OMOP-CDM version 5.4 and its suitability for representing wearable and patient-generated health data (2,3) . The analysis focused on architectural requirements encountered when integrating wearable-device outputs, continuous sensor measurements, PROMs, PREMs, and other forms of patient-generated information into secondary-use research infrastructures.

A gap analysis was conducted to evaluate the ability of existing OMOP structures to support longitudinal observations, device traceability, provenance preservation, data-quality monitoring, and reproducible ETL processes (3). Particular attention was given to requirements associated with high-frequency sensor streams, wearable device lifecycle information, and contextual metadata needed to support interpretation and reuse of PGHD.

Based on the identified requirements, a modular extension framework was designed to complement existing OMOP structures while preserving compatibility with established OHDSI analytical workflows and tools (Figure 1).

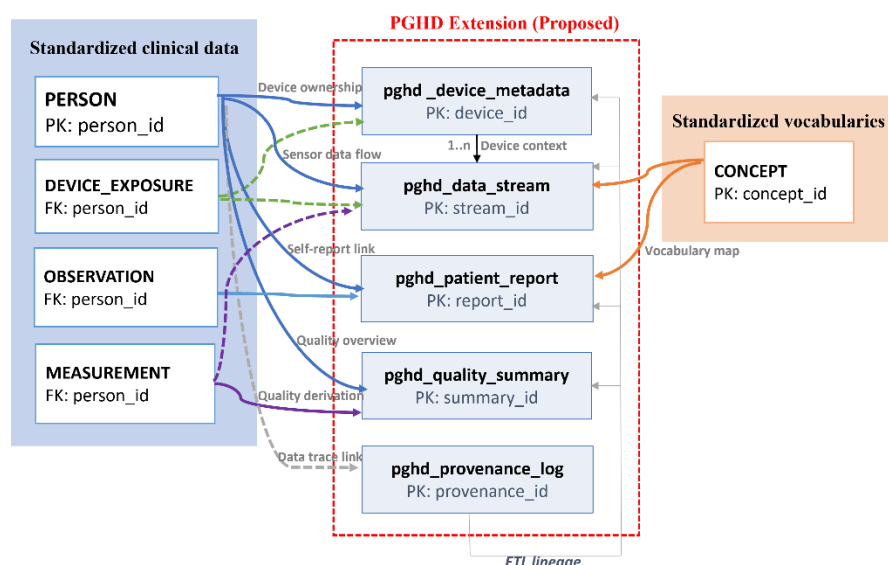


Figure 1. Proposed extension framework for representing wearable and patient-generated health data in OMOP-CDM. Source: Abedian and Hussein (3)

Results

The analysis identified several recurring representational gaps affecting wearable and patient-generated data within OMOP environments (3). Existing OMOP structures provide robust support for standard clinical measurements and observations but offer limited mechanisms for preserving stream-level structure, device lifecycle and calibration information, transformation provenance, and PGHD-specific data-quality context. To address these challenges, the framework introduces five complementary components (3):

- Device Metadata, supporting representation of device characteristics, manufacturers, firmware versions, calibration information, and acquisition methods.

- Data Streams, supporting representation of longitudinal and high-frequency measurements generated by wearable devices and mobile applications.
- Patient Reports, preserving contextual information associated with PROMs, PREMs, and other patient-reported instruments.
- Quality Summaries, documenting completeness, validity, and quality indicators associated with collected PGHD.
- Provenance Logs, supporting transparent recording of ETL lineage and transformation history.

The framework is intended to complement rather than replace existing OMOP structures. Summary-level measurements and observations remain within standard OMOP workflows, while the extension retains additional PGHD context when this information cannot be adequately preserved in the core model. For example, in the case of wearable heart-rate data, the detailed time-series data can be retained in the Data Streams component, while a derived summary measure can remain in MEASUREMENT. Information about the device, including manufacturer, firmware version, calibration, and acquisition method, can be retained in Device Metadata. Related data-quality information and transformation history can be recorded in Quality Summaries and Provenance Logs, respectively. For patient-reported data, individual responses can remain within standard OMOP structures such as OBSERVATION, while the additional context associated with the PROM or PREM instrument can be retained in Patient Reports. Quality information related to the collected responses can be recorded in Quality Summaries, and the associated transformation history in Provenance Logs. Table 1 illustrates this mapping with concrete field-level values for a representative heart-rate stream. Thus, these components provide additional traceability and context for PGHD while supporting FAIR (Findable, Accessible, Interoperable and Reusable)-aligned secondary use.

Table 1. Illustrative example: representation of a single wearable heart-rate stream across existing OMOP structures and the proposed PGHD components.

Component	Example field(s)	Illustrative value
MEASUREMENT	measurement_concept_id, value_as_number, unit_source_value	3027018 (Heart rate), 68, beats/min - 24h mean
pghd_device_metadata	device_id, person_id, device_type, manufacturer, firmware_version, acquisition_method, calibration_date	DEV_077, P_10432, wrist- worn optical sensor, Fitbit, 3.2.1, Fitbit API, 2026-02-15
pghd_data_stream	stream_id, device_id, person_id, measurement_concept_id, start_time, end_time, sampling_rate, compressed_data, summary_statistics, provenance_id	STR_1024, DEV_077, P_10432, 3027018, 2026-03-01 00:00, 2026-03-02 00:00, 0.033 Hz, [JSON/BLOB], [JSON summary statistics], PRV_2201
pghd_quality_summary	quality_id, person_id, data_date, missing_rate, valid_records, derived_quality_index	Q_5591, P_10432, 2026-03- 01, 5.8%, 2712, 0.94

Component	Example field(s)	Illustrative value
pghd_provenance_log	provenance_id, source_system, etl_process, etl_version, etl_datetime, checksum, etl_operator	PRV_2201, Fitbit API v1.2, raw-to-aggregate, v1.4, 2026-03-02 09:00, SHA256:8f4cxxx, Somayeh Abedian

Extending existing OMOP structures such as MEASUREMENT or FACT_RELATIONSHIP was considered as an alternative to introducing new tables. For example, at a sampling frequency of 1 Hz, a single day of heart-rate monitoring would generate 86,400 measurements per person. Representing each sample as an individual MEASUREMENT row could therefore substantially increase row counts, particularly when scaled across participants and longer observation periods. Unlike a row-per-sample approach, pghd_data_stream stores high-frequency signals as compressed time-series blocks with derived summary statistics, reducing the number of rows stored in MEASUREMENT compared with a row-per-sample approach. FACT_RELATIONSHIP can express associations between OMOP facts, but it does not provide a structured representation for device-specific attributes, sampling characteristics, or transformation provenance. Representing these elements would therefore require additional facts or implementation-specific conventions, reducing consistency across implementations. The proposed components therefore isolate this high-volume, structurally distinct content in dedicated tables while retaining a single derived summary value in MEASUREMENT, preserving compatibility with existing OHDSI tools that query MEASUREMENT directly.

The framework has so far undergone a preliminary conceptual review of referential integrity and semantic consistency against LOINC, SNOMED CT, and UCUM vocabularies. Practical, operational examples and implementation results will be discussed in more detail in forthcoming publications. The framework's design preserves the existing OMOP core structure so that standard OHDSI workflows can continue to operate on the summary-level data. Broader challenges, including cross-device interoperability, PRO instrument harmonization, consent governance, and semantic comparability, remain outside the scope of the current framework and require further community discussion.

Conclusion

The increasing adoption of wearable technologies and patient-generated health data creates new opportunities for observational research while simultaneously introducing new representational requirements for common data models (1). The proposed extension framework represents one possible approach for addressing architectural gaps related to provenance management, device metadata, longitudinal observations, patient-reported data context, and data-quality traceability within OMOP (3).

The framework is being explored in the context of PGHD interoperability initiatives involving wearable data integration, FHIR-based data exchange, and secondary-use research infrastructures. Many of the challenges associated with PGHD extend beyond data structures

alone. Representing wearable and PGHD within OMOP requires not only structural solutions, but also shared approaches for semantic harmonization, device interoperability, governance, provenance management, and the meaningful use of PGHD in research and clinical contexts. There is still no broadly accepted approach for addressing many of these issues across different digital health initiatives.

This submission is intended as a community discussion paper rather than a finalized specification. By bringing the framework to the OHDSI community, we invite OHDSI working groups to evaluate it as a concrete starting point for discussion around PGHD representation, CDM evolution, and vocabulary governance. The proposed tables should be viewed as one candidate design rather than a recommendation for immediate CDM modification. A key question for community evaluation is whether these requirements justify dedicated structures or can be addressed through existing OMOP domains, conventions, and vocabulary extensions without compromising interoperability or analytical usability.

References

1. Azodo I, Williams R, Sheikh A, Cresswell K. Opportunities and Challenges Surrounding the Use of Data From Wearable Sensor Devices in Health Care: Qualitative Interview Study. *J Med Internet Res*. 2020 Oct 22;22(10):e19542. doi:10.2196/19542.
2. Pyper E, McKeown S, Hartmann-Boyce J, Powell J. Digital Health Technology for Real-World Clinical Outcome Measurement Using Patient-Generated Data: Systematic Scoping Review. *Journal of Medical Internet Research*. 2023 Oct 11;25(1):e46992. doi:10.2196/46992
3. Abedian S, Hussein R. Proposed Schema Extensions and ETL Pathways for Integrating Wearable and Patient-Reported PGHD into OMOP-CDM for Secondary Use. *Stud Health Technol Inform*. 2026 May 21;336:1885–9. doi:10.3233/SHTI260565 PubMed PMID: 42175235.
4. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, et al. The *All of Us* Research Program: Data quality, utility, and diversity. *Patterns*. 2022 Aug 12;3(8):100570. doi:10.1016/j.patter.2022.100570
5. Master H, Kouame A, Marginean K, Basford M, Harris P, Holko M. How Fitbit data are being made available to registered researchers in *All of Us* Research Program. *Biocomputing 2023*. 2022 Nov;19–30. doi:10.1142/9789811270611_0003