WearMerge: An Interoperable Framework for Self-tracking Data Integration and Standardization

Dimitrios Panteleimon Giakatos*, Sofia Yfantidou†, Stefanos Efstathiou‡, and Athena Vakali§
Department of Informatics, Aristotle University of Thessaloniki, Greece
Email: *dgiakatos@csd.auth.gr, †syfantid@csd.auth.gr, ‡efstathiou@csd.auth.gr, §avakali@csd.auth.gr

Abstract—**Ubiquitous self-tracking technologies' (STTs) adoption has taken a quantum leap in recent years, leading to a rapid increase in terms of volume, variety, and variability of the generated data from their embedded sensors. Consequently, integrating data from different self-tracking devices for further exploration and analysis has become time-consuming. In addition, it requires advanced technical skills, hindering their widespread adoption in interdisciplinary scientific and industrial research. This paper introduces an extensible, open-source framework and tool called WearMerge that automates the integration and transformation into a common standard of STTs' data across different brands and models. WearMerge aims to help and ease practitioners and researchers on STTs' data analysis.**

Index Terms—**data integration, data standardization, data visualization, wearables**

I. INTRODUCTION

Today, more and more people use ubiquitous self-tracking technologies (STTs), such as wearables and mobile apps, and their embedded sensors have enabled a broad range of measurements regarding individuals' health and well-being. Such data, if extracted and analysed, can be particularly useful for a wide range of specialties across industry and academia [1, 2]. However, if one tries to understand and integrate data extracted from STTs, one will discover significant non-uniformity. The reason is that each manufacturer has different extraction and representation standards, making it arduous to consolidate all [3]. Naturally, the question that arises is how we can integrate and analyze data from multiple STTs' manufacturers, models, and sensors, given their diversity, enabling the interoperability between different devices.

This question can be further dismantled into four issues that beset the domain of interoperable STTs. Initially, there is a lack of understanding of how different manufacturers store and represent the collected data, rendering data exploration time-consuming (I1). On top of that, different manufacturers adopt different storage and export standards, such as JSON, CSV, or XML, and representation schemas (I2). For example, different manufacturers utilize different field names or units for the same measured quantity or provide distinct quantities not found in competitors. At the same time, to the best of the authors' knowledge, there is no free, easy-to-use application to integrate and visualize different standards and schemas, especially for users lacking technical skills (I3). Finally, there are no open-source tools, limiting their extensibility and community access (I4).

There have been some efforts to develop integration APIs or mobile and web applications for handling the issues above. After careful review, we have identified three domain-related categories: medical applications [4, 5, 1, 6], social and psychological applications [7, 8] and data analysis applications [9, 2]. However, the majority of these works do not use mobile health data interoperability standards for standardizing and harmonizing data from disparate sources, nor have they made their code open-source, while others support only a limited amount of manufacturers [4]. Also, none provides data schematics for the included STTs' models.

This paper tackles these limitations by introducing WearMerge, a framework and web application that explores, integrates, standardizes, and visualizes STTs' data from the top six manufacturers in terms of global market share (Apple, Fitbit, Garmin, Huawei, Samsung, and Xiaomi). Our solution utilizes the open-source standard Open mHealth (OMH) [10], promoting accessible digital health data through an open interoperability standard. Note that, to the best of the authors' knowledge, Open mHealth is the only standard that provides common data schemas focusing on the mHealth domain. Hence, exported data from WearMerge can be efficiently utilized by other applications using the same or compatible standard, such as HL7 FHIR. In summary, the contributions of this work are as follows:

Data Schematics: We analyze, and visually depict via open-source UML diagrams [11] the representation schemas utilized by different manufacturers, considerably facilitating and accelerating data exploration for future work (I1).

Data Integration & Standardization: We adopt the open-source standard OMH to integrate and standardize data from different manufacturers, models, or sensors into a common format, enabling interoperable data analysis for different export and representation standards (I2).

Online Interactive Tool and API: We build an online tool with an intuitive interface, and a REST API, rendering our interoperability framework accessible for users of diverse levels of technical skills and disciplines (I3).

Extensible Open-source Software: We open our code to the community, making our work easily extensible for additional models and manufacturers of STTs (I4).

II. THE WEARMERGE INTEROPERABILITY FRAMEWORK

This section presents the data pipeline of the WearMerge framework (Section II-A), as seen in Figure 1, and the respec-

tive open-source tool as proof of concept (Section II-B).

A. Data Validation, Integration & Transformation Principles

Integrating data from systems, sensors, and people into mashup applications is an increasingly common phenomenon. The interoperability problem starts as soon as different technological platforms, query languages, and data standards co-exist. The WearMerge framework takes a set of steps for the cleaning, validation, integration, and transformation of STTs' data, as illustrated in Figure 1 and explained below, to address this problem.

First, each STTs' manufacturer has a unique export standard for their data. In the **Data Cleaning** stage, WearMerge ensures the compatibility of the exported file's structure with the data parser, i.e., the algorithm that reads the files and sends raw data to the next stage, and filters out all but CSV, JSON or XML files. Then, in the **Data Validation** stage, WearMerge's validation algorithm is executed for each sample coming from the previous stage. Specifically, given that all selected STTs' data can be verified by their representation schema, as identified through our UML diagrams [11], the validation algorithm recognizes the manufacturer through a schema-manufacturer matcher enhanced by regular expressions. If a match is possible, the data move to the next stage; otherwise, the sample is discarded, and the algorithm continues.

Based on past works, we move forward with the **Data Integration** stage by creating a database that includes unified data, i.e., a data warehouse. This "global schema" approach is better suited for systematic or ad-hoc analyses. Specifically, WearMerge utilizes a global representation schema, as our use cases are not time-sensitive. For the interoperability of STTs' data, we follow literature guidelines to resolve common data integration issues [12, 13] as follows:

Schematic Harmony Evaluation: WearMerge assesses the schematic harmony between disparate data representation schemas. For example, we identify whether a schema contains a temperature field or not. To achieve this, we create and compare data schematics through UML diagrams for all studied devices' models and manufacturers.

Content-based Record Linkage: WearMerge then links records with the same content. For instance, the temperature field can be referred to as temperature or as daily_temperature in different representation schemas. However, it must be consolidated as it refers to the same measurable quantity.

Data Compatibility Handling: WearMerge also handles data compatibility with a compromise between records of different content. Specifically, for the data types that incorporate a unit of measurement, it converts them to the measurement unit adopted by the OMH standard (e.g., temperature in Celcius). Otherwise, it assumes the standard application unit. For example, the unit of temperature (°C vs. °F) must be consolidated.

Finally, in the **Data Transformation** stage, WearMerge transforms the integrated data into the open-source interoperability standard OMH representation schema to make human-generated health, physical activity, and well-being data more

TABLE I
COMPATIBILITY BETWEEN STTs' MANUFACTURERS' REPRESENTATION SCHEMATA AND OPEN MHEALTH DISTINCT SCHEMATA. THE ✓, ⌣ AND X SYMBOLS INDICATE FULL, PARTIAL AND NO MATCH RESPECTIVELY.

	Heart Rate	Step Count	Total Sleep Time	Pace	Physical Activity	Calories Burned	Body Weight	Body Height	Body Fat (%)
Xiaomi	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fitbit	✓	✓	✓	x	⌣	✓	✓	✓	✓
Garmin	✓	✓	✓	✓	✓	✓	✓	✓	x
Apple	✓	✓	✓	✓	⌣	✓	✓	✓	✓
Huawei	✓	✓	✓	✓	✓	✓	✓	✓	✓
Samsung	✓	✓	✓	✓	⌣	✓	✓	✓	x

interoperable. Table I shows the compatibility between the STTs' manufacturers' representation schemata and the selected OMH schemata.

B. Open-Source Tool Implementation & Uptake Scenarios

To demonstrate the functionality and utility of the WearMerge framework, we built a REST API and an extensible online tool[1], which consists of two parts: the frontend, developed with HTML5, CSS3, and JavaScript, and the backend, developed with JavaScript, opening up our wearables interoperability framework to the community.

The WearMerge tool is straightforward to use and implements the data pipeline described in the previous section. Namely, after the user's data upload, the tool proceeds with the cleaning, validation, integration, and transformation of the uploaded data samples. Once ready, the application notifies the user to download the interoperable data and access the produced interactive visualizations while it automatically deletes the raw input data. We also offer a REST API for more advanced users, which accepts calls from third-party applications upon agreement. For a preliminary evaluation of the functionality and integrity of the WearMerge framework, we compared various statistical features between the automatically and manually consolidated data, verifying the correctness of our pipeline. It is worth mentioning that, due to the nature of STTs data, the application design accounts for the issues of data velocity and volume, utilizing JavaScript's Streams API and automated deletion of deprecated data.

WearMerge can be of great value in interdisciplinary research. For instance, an indicative uptake scenario includes researchers from the Faculty of Education interested in exploring the correlations between students' stress levels and concentration and their sleep patterns. WearMerge enables them to integrate the collected data from the students' STTs and prepare them for analysis without any technical skills requirements. On a different note, a developer of a well-being mobile app that explores how social media usage affects users' mental health can take advantage of WearMerge API to incorporate various STTs as data sources in their app, broadening their potential user base without a significant time investment.

III. DISCUSSION AND CONCLUSIONS

Summing up, we implemented all four contributions to the community through the completion of the WearMerge backend

---
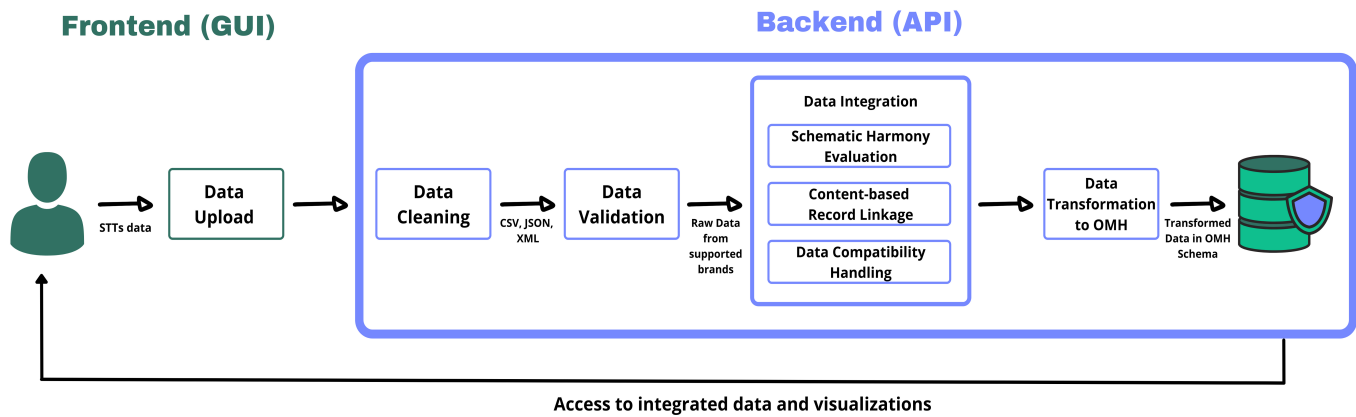
[1]https://wearmerge.csd.auth.gr/

Fig. 1. The WearMerge Data Pipeline

API and frontend platform. During application development, we identified several challenges related to data management and integration, including the existence of missing values, the incorrect structure of uploaded files, the conversion of some schemas to deprecated by OMH, and the absence of time zone (UTC), measurement units or physical activity details in certain schemas. Such issues can be critical during data analysis. To this end, we designed WearMerge to deal with such problems except those stemming from the lack of manufacturers' documentation. Additionally, there are certain future work directions we propose in order for WearMerge to become a full-fledged online integration tool. Specifically, a high-scale user evaluation is required to ensure the quality of our framework and tool. Also, given the open-source nature of our project, we encourage researchers and developers to incorporate more manufacturers and models into the framework, engage in the continuous maintenance of the selected representation schemata, and enrich the data exploration charts presented to the user.

REFERENCES

[1] R. K. Saripalle, "Leveraging fhir to integrate activity data with electronic health record," *Health and Technology*, vol. 10, no. 1, pp. 341–352, 2020.

[2] L. P. Prieto, K. Sharma, Ł. Kidzinski, M. J. Rodríguez-Triana, and P. Dillenbourg, "Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data," *Journal of computer assisted learning*, vol. 34, no. 2, pp. 193–203, 2018.

[3] A. Farseev and T.-S. Chua, "Tweetfit: Fusing multiple social media and sensor data for wellness profile learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[4] F. Sartori, "An API for Wearable Environments Development and Its Application to mHealth Field †," *Sensors*, vol. 20, no. 21,

p. 5970, Jan. 2020, number: 21 Publisher: Multidisciplinary Digital Publishing Institute.

[5] V. Gay and P. Leijdekkers, "Bringing Health and Fitness Data Together for Connected Health Care: Mobile Apps as Enablers of Interoperability," *Journal of Medical Internet Research*, vol. 17, no. 11, p. e260, Nov. 2015.

[6] M. Muzny, A. Henriksen, A. Giordanengo, J. Muzik, A. Grøttland, H. Blixgård, G. Hartvigsen, and E. Årsand, "Wearable sensors with possibilities for data exchange: Analyzing status and needs of different actors in mobile health monitoring systems," *International Journal of Medical Informatics*, vol. 133, p. 104017, Jan. 2020.

[7] K. Kutt, G. J. Nalepa, B. Giżycka, P. Jemiolo, and M. Adamczyk, "BandReader - A Mobile Application for Data Acquisition from Wearable Devices in Affective Computing Experiments," in *2018 11th International Conference on Human System Interaction (HSI)*, Jul. 2018, pp. 42–48.

[8] F. de Arriba Pérez, M. Caeiro Rodriguez, and J. Santos, "Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios," *Sensors*, vol. 16, p. 1538, Sep. 2016.

[9] F. Hantke and A. Dewald, "How can data from fitness trackers be obtained and analyzed with a forensic approach?" in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. Genoa, Italy: IEEE, Sep. 2020, pp. 500–508.

[10] D. Estrin and I. Sim, "Open mhealth architecture: an engine for health care innovation," *Science*, vol. 330, no. 6005, pp. 759–760, 2010.

[11] D. P. Giakatos, "Datalab-AUTH/BSc-2021-Giakatos-Dimitrios- Panteleimon-WearMerge-UML-Diagrams: Wearables' UML-Diagrams," Jan. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.5850311

[12] R. De Virgilio, F. Giunchiglia, and L. Tanca, *Semantic web information management: a model-based perspective*. Springer Science & Business Media, 2010.

[13] F. Daniel and M. Matera, *Mashups: concepts, models and architectures*. Springer, 2014.