

Unlocking the Power of Real-World Data: A Framework for Sustainable Healthcare

Ashkan PIRMANI^{a,b,c}, Yves MOREAU^a and Liesbet M. PEETERS^{b,c,1}

^a*ESAT, STADIUS, KU Leuven, Leuven, Belgium*

^b*Data Science Institute, Hasselt University, Diepenbeek, Belgium*

^c*Biomedical Research Institute, Hasselt University, Diepenbeek, Belgium*

ORCID ID: Ashkan Pirmani <https://orcid.org/0000-0003-4549-1002>

Abstract. Real-world data (RWD) has the potential to revolutionize healthcare by offering valuable insights into patient outcomes and treatment efficacy. However, leveraging RWD effectively presents challenges, including its inherent limitations, diverse stakeholders, and insufficient data management pipelines. A proposed framework advocates three essential elements: adherence to FAIR principles (Findable, Accessible, Interoperable, and Reusable), stakeholder engagement and education, and highlighting the need for inclusive, pragmatic federated hybrid pipelines. By employing these strategies, healthcare organizations can overcome obstacles to RWD utilization and foster sustainable progress in patient care.

Keywords. Real-World data, Healthcare data management, Federated data analysis

1. Introduction

Derived from various sources, Real-World Data (RWD) reflects the diverse nature of patient experiences and healthcare delivery. However, leveraging RWD to its fullest extent involves navigating challenges across technical, operational, and ethical domains.

High-quality, interoperable RWD is crucial but often fragmented within IT systems. Challenges arise when using this data in machine learning, including annotation, trustworthiness, and interpretability issues. Complicating matters further are heterogeneous data formats and privacy regulations. Additionally, the absence of standardized methodologies for integration hinders seamless utilization of RWD [1].

Integrating the diverse requirements of healthcare providers, regulatory bodies, research entities, and patients into a cohesive strategy is essential. However, communication and interpretation challenges often hinder effective healthcare data management. Divergent terminologies and technical language barriers can lead to misunderstandings and inefficiencies. Additionally, the lack of and varying levels of engagement among healthcare stakeholders have significant further impacts.

The architecture of data management pipelines plays a crucial role in handling and processing RWD [2]. Whether centralized, decentralized, or federated, each approach offers its own set of advantages and challenges. Consequently, the choice of architecture significantly impacts the feasibility, efficiency, scalability, and compliance of data processing activities. However, the lack of comprehensive frameworks and insufficient adoption of hybrid systems curtail pipeline efficiency [1].

¹ Corresponding Author: Liesbet M Peeters; E-mail: liesbet.peeters@uhasselt.be.

2. A Triangular Framework to Addressing the Challenges

Addressing healthcare data management challenges requires a comprehensive plan that integrates technological advancements with strategic, procedural, and policy-oriented changes. A holistic approach is needed to tackle the complex obstacles, combining technical innovation with practical considerations [3].

Implementing FAIR [4] principles across RWD initiatives can facilitate findability, accessibility, interoperability, and reusability, addressing some of the core challenges of RWD utilization. To overcome stakeholders' challenges, adopting an interdisciplinary approach that fosters collaboration and mutual understanding through ongoing education and training is crucial. Establishing interdisciplinary groups and organizing regular workshops, symposiums, and forums can help bridge gaps and promote knowledge-sharing among healthcare stakeholders. By prioritizing professional development, we can empower stakeholders to drive growth and sustainability within their organizations. To overcome challenges of healthcare data management pipelines, we advocate for a hybrid approach that integrates centralized and federated systems strategically. Federated data analysis, a key component of this model, allows for privacy-preserving analytics by processing information directly at decentralized sources, eliminating the need to transfer sensitive data. This approach transcends the centralized versus federated dichotomy. While centralized systems offer simplicity, federated systems can optimize efficiency and privacy. By employing a hybrid system, stakeholders can customize strategies based on their specific requirements, benefiting from the architecture's flexibility [2].

3. Conclusions

The effective utilization of RWD in healthcare is a multifaceted challenge that demands an integrative approach, combining technological solutions with a strong emphasis on community engagement and collaboration. Exploring the use of federated data analysis by adopting a hybrid data management architecture, operationalizing FAIR principles, and fostering a culture of continuous education, we can navigate the complexities of RWD, stakeholder engagement, and data management pipelines. This comprehensive strategy not only addresses the immediate challenges but also sets a foundation for sustainable, data-driven advancements in healthcare.

References

- [1] Rudrapatna VA, Butte AJ. Opportunities and challenges in using real-world data for health care. *J Clin Invest* 2020; 130: 565–574.
- [2] Pirmani A, De Brouwer E, et al. The Journey of Data Within a Global Data Sharing Initiative: A Federated 3-Layer Data Analysis Pipeline to Scale Up MS Research. *JMIR Med Inform* 2023; 11: e48030.
- [3] Hiramatsu K, Barrett A, Miyata Y. Current Status, Challenges, and Future Perspectives of Real-World Data and Real-World Evidence in Japan. *Drugs Real World Outcomes* 2021; 8: 459–480.
- [4] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016 3:1 2016; 3: 1–9.