



Contribution ID: 118

Type: **Presentation**

Creating an integrated Electronic Health Record data platform for revolutionising healthcare research

Monday, October 13, 2025 2:41 PM (11 minutes)

Background

The digitalisation of Electronic Health Record (EHR) data has unlocked unique opportunities for research. Unlike administrative datasets, EHRs provide granular clinical data, real-time updates within systems, and access to detailed clinical notes. Despite these advantages, EHR data—primarily collected for operational purposes—remains siloed, lacks standardisation between systems, suffers from poor interoperability, and contains large amounts of unstructured text. Consequently, EHR data is often not curated to the standard required for research, which hampers its optimal use in healthcare studies. The National Centre for Healthy Ageing (NCHA), a partnership between Monash University and Peninsula Health (comprising four hospitals and over ten outpatient and community services), has developed a unique EHR-derived Data Platform. Its primary objective is to integrate multi-site healthcare data from an entire geographic region to support translational research focused on healthy ageing across the life course.

Methods

Our approach involved establishing a core set of EHR data suitable for research from the sole public health provider within a defined geographic region. Relevant items were identified based on published literature and consensus processes, then curated within a specialised research data warehouse. The curation process included data validation, quality assessment, internal linkage using a patient data spine, and data harmonisation/merging. Semi-automated data extraction processes were developed for approved research projects. End-users were engaged in defining the platform's content, and consumer workshops were conducted to understand community perspectives on data management and governance. Further efforts to expand the platform's content included implementing an AI pipeline to extract concepts from clinical notes, routinely collecting Patient Reported Outcome Measures, and linking to a variety of local, state, and national datasets (e.g., primary care, medication, aged care, hospital, and mortality datasets). Publicly available datasets were scoped for inclusion, and collaborations with local councils were established to incorporate community data.

Results

The platform's research data warehouse contains curated data for over one million patients, collected over ten years and updated weekly. A total of 131 core data items from 11 research-relevant datasets across the four hospitals and ten community/outpatient services have been identified for inclusion. Data access, extraction, and release processes are guided by the Five Safes Framework. A natural language processing (NLP) pipeline has been implemented and trained to detect dementia in clinical notes. A framework for the routine collection and integration of patient-reported outcome measures has been developed. Linked state and commonwealth data for 179,089 residents aged >60 years (January 2010–May 2021) has been obtained, achieving a linkage accuracy of 98.4%. Environmental (greenery, air pollution, walkability) and Census data have been incorporated at the neighbourhood level (Statistical Area 1). Over 50 use-case projects have tested data access, extraction, and release into a Monash-hosted Trusted Research Environment, covering topics such as dementia, residential aged care, medication use, homelessness, environmental impacts on health, ageing, and health service redesign.

Implications

The NCHA Data Platform provides an international exemplar for leveraging linked EHR data to advance population health research. In addition to delivering high-quality, research-grade data to clinicians and researchers, the platform serves as critical infrastructure to underpin data-driven innovation across multiple domains.

Primary author: ANDREW, Nadine (Monash University)

Co-authors: Prof. BEARE, Richard (Monash University); Prof. SRIKANTH, Velandai (Monash University)

Presenter: ANDREW, Nadine (Monash University)

Session Classification: Presentations Session 2: Data and Research & Data Science and Data Analysis

Track Classification: SciDataCon2025 Specific Themes: Infrastructures to Support Data-Intensive Research - Local to Global