

International Journal of Population Data Science

Journal Website: www.ijpds.org



Clean Data, Breakthrough Insights: The Importance of Standardizing Your Research Warehouse

Erik Youngson¹, Nathan Armani¹, Tara Whitten¹, Hude Quan², Bing Li¹, Ting Wang¹, and Jeffrey Bakal¹

¹Alberta Provincial Research Data Services, Edmonton, Canada

²University of Calgary, Calgary, Canada

Leveraging routinely collected health services data to conduct scientifically robust research requires a clear understanding and standardization of data from disparate clinical information systems which often change over time. Within Alberta Provincial Research Data Services (PRDS) a large team of analysts utilize a vast Enterprise Data Warehouse containing dozens of datasets with various update cycles and availability, generally spanning more than 20 years of data for a Canadian province of over 4 million people. We examine two recently developed standardized data marts for mortality and medical lab tests. These data marts are now updated daily using Snowflake Tasks to ensure timely data can be accessed as needed to support research studies. The mortality data are linked from three sources to ensure comprehensive coverage:

Vital Statistics, Provincial Registry, and Connect Care (the provincial Epic based electronic medical record) and resolves inconsistencies in personal health numbers and discrepancies in dates of death. The Lab data mart uses test level data from several different historical lab systems and the current Connect Care system and consolidates coding and definitions over time to create a clean, easy to use data mart. Both use cases create a single, validated and linkable source of truth that is easy to use for research studies, negating the need for each analyst to re-engineer from several different source tables each time. The process used here will be applied to other subject matters internally and could be similarly applied data warehouse environments in other jurisdictions.

