Contribution ID: **307** Type: **Poster**

Development of a data search navigation tool to support data linkage information for comparative effectiveness research –a service provided by Taiwan Gateway to Health Data

Monday, October 13, 2025 7:10 PM (20 minutes)

The Taiwan Gateway to Health Data (GHD TW) is a government-funded project that collaborates with all primary data custodians and data controllers in Taiwan. We establish a data portal for various data users, including industrial and academic researchers, to promote community health and clinical and biomedical research. Our primary responsibility is to provide services that enhance the findability, accessibility, interoperability, and reusability (FAIR) of our data partners. For comparative effectiveness research, the search for fit-for-purpose data often involves linking data from multiple sources to obtain complete information on outcomes, exposures, and confounders. In Taiwan, data linkage could be a challenge to data users. Since different data controllers may govern various data sources, the availability for linkage is generally subject to various restrictions. It is usually time-consuming for data users to determine whether a fit-for-purpose collection of data is available. We decided to develop a data search navigation tool to support data linkage information, which is available in both traditional Chinese and English.

The framework is built upon a MeSH-based synonym. We first convert English MeSH keywords into traditional Chinese and utilize a bilingual (Chinese-English) synonym dictionary and a hierarchical tag classification system, specifically designed to enhance data discoverability and relevance in Taiwan's national health and biomedical research context. We then leverage over 130,000 curated synonym entries and 182 concept tags derived from and extended beyond MeSH. We designed a search engine tailored for domestic datasets. Each dataset is annotated with weighted relevance scores across standardized tags, enabling the system to recommend associated dataset bundles based on term input and relevance ranking. The framework enables users to identify suitable datasets quickly, comprehend their fit-for-purpose structure, and assess their interoperability based on the feasibility of data linkage and the characteristics of the data. In addition to building the search interface, we incorporated user feedback mechanisms to dynamically adjust relevance scores, thereby improving recommendation precision over time. This architecture not only supports rigorous and reproducible research but also aligns with FAIR principles. Our approach demonstrates how domain-specific, language-sensitive design can bridge the gap between global data standards and local research needs.

Primary authors: Dr LEE, Hsiu-An (National Health Research Institutes); Mr TSENG, Wen-Chang (National Health Research Institutes); YANG, Yi-Hsin (National Health Research Institutes)

Presenter: Dr LEE, Hsiu-An (National Health Research Institutes)

Session Classification: Poster Session

Track Classification: SciDataCon2025 Specific Themes: Rigorous, responsible and reproducible science in the era of FAIR data and AI