Contribution ID: **20**   Type: **Regular Presentation**

New Project and Tools To Aid In DDI-based Variable Concordance and Harmonization

Wednesday, 29 November 2023 13:55 (25 minutes)

The current research data environment provides many opportunities for linking similar topical datasets and harmonizing extant common variables. The DDI Lifecycle standard supports documenting these linkages, but few software tools are available to facilitate the actual performance of this resource-intensive task. This project uses a DDI based framework to assemble richly-described datasets that are mapped against DDI represented and conceptual variables to identify equivalent concepts and variables. The tools use machine learning and advanced text analysis algorithms to guide the creation of concorded databases (variable crosswalks) that support harmonization and discoverability, both within and across statistical datasets and studies. Specifically, the tools use several human-in-the-loop algorithms to operate as a "recommendation engine" to guide the concordance of potentially equivalent or similar variables among multiple datasets. The goal of this project is to significantly decrease the labor, time, and resources required to create accurate and standardized concorded databases and store the results using the DDI standard.

Primary author: SMITH, Dan (Colectica)

Presenter: SMITH, Dan (Colectica)

Session Classification: Harmonisation

Track Classification: Proposed Topics of the Conference: Interoperability, Reusing and Sharing Metadata