



Contribution ID: 16

Type: **Regular Presentation**

Bridging Portals and Formats: Transforming Metadata from DDI-C 2.5 to OmicsDI in BY-COVID Project

Wednesday, 29 November 2023 16:00 (30 minutes)

One of the primary objectives of the BeYond-COVID project was the integration of social sciences & humanities studies' metadata into the COVID-19 Data Portal. This presentation focuses on the technical work done to achieve this for studies found in CESSDA Data Catalogue (CDC).

Focus of the presentation is the implementation of XML transformation from DDI-Codebook 2.5 to extended OmicsDI format using an XSL template (XSLT). Other parts of the overall process are also covered. XSLT transformations are ran with SaxonC which fully supports XSLT 2.0 and even 3.0, version 2.0 being the one used in this project. Other necessary steps consist of things like harvesting metadata from CDC's OAI-PMH endpoint, filtering and XML validation.

All steps are combined into one fully automated pipeline that creates an XML file with filtered subset of metadata in OmicsDI format for the COVID-19 Data Portal to harvest and use. First version has been in production since September 2022. However, improvements will be made until the end of the project in September 2024.

The presentation also covers the encountered challenges, lessons learned and what can still be improved.

Primary author: TUOMINEN, Markus (Finnish Social Science Data Archive (FSD))

Presenter: TUOMINEN, Markus (Finnish Social Science Data Archive (FSD))

Session Classification: DDI & Other Standards

Track Classification: Proposed Topics of the Conference: Software and Tools