

International Journal of Population Data Science

Journal Website: www.ijpds.org



Implementation of the Frictionless data standard for linked data

Mike Jarrett¹

¹Population Data BC

Objective and Approach

Data linkage centres receive, hold and provision data in a wide variety of formats. Using a consistent, flexible and open metadata standard for data holdings can enable a center to clearly and unambiguously describe their data holdings and facilitate processing pipelines. One such framework is the Frictionless data standard. We present our experience at our data centre implementing this standard for both internal data management and for communication with external data providers and third parties.

Implications

Adopting the Frictionless data standard for metadata files has streamlined our internal processes and improved internal and external communication.

Results

Our team has replaced tool-specific metadata files with standard-compliant files to make our pipelines more modular and less fragile. Since the standard is based on JSON, most programming languages will support reading and writing metadata files and we are not restricted to any software. This work has also made our internal tooling interoperable with that of external partners. We find that the standard is well equipped to handle CSV and other delimited files. Our center uses a mix of file types including fixed-width text files which do not have an explicit specification within the standard, but its extensibility has allowed us to define our own file type while staying within the ecosystem. We have also developed tools to convert JSON metadata files to excel workbooks for a more human-friendly format.

Conclusions

Since the initial development of this ongoing crosswalk, two updates have been successfully completed leading to the linkage of over 7,000 study participants between the two platforms. This has led to a more efficient utilization of human and computational resources, compared to earlier data linkage projects completed on a project-by-project basis. Analysis projects in the areas of neurodegeneration, concussion, and neurodevelopmental disorders have already leveraged this crosswalk linkage process.

