

International Journal of Population Data Science

Journal Website: www.ijpds.org



The Enriched Data Dictionary

Cassia Smith¹

¹University of Edinburgh, Edinburgh, United Kingdom

As the research data community turns its attention to federated systems, data dictionaries and variable-level metadata become increasingly important. Not only do variables need to be identified for their value to various linkage methods, but the breadth and depth of available data needs to be easily navigable by characteristics ranging from collection timeframe to coding matrix. In a moment when the data dictionary is poised to engage systems far more complex than a spreadsheet or a codebook, it's worth considering the ways this data documentation format can be expanded to assist both researchers and data providers in their navigation of the wider federated landscape. Three areas in particular are worth considering. First, domain-appropriate metadata systems are one obvious way to promote federation by presenting data in a consistent

and meaningful format. Complications arise when considering which format to adopt, but they are not insurmountable. Second, as DOI assignment becomes standard for research datasets, data dictionaries and metadata catalogues in general provide a possible route into data citation concepts that can help data providers track usage and researchers clearly identify the data they need to request. Third, creating linkages between data dictionaries and research outputs can both demonstrate the value of the data to funders and spark creativity among researchers. With over ten years of experience in metadata management and a current project developing a small federated data dictionary for a suite of longitudinal population studies, the presenter aims to spark conversation around this overlooked form of data documentation.

