Crowdsourcing Codebook Enhancements: A DDI-based approach

Benjamin Perry, Venkata Kambhampaty, Kyle Brumsted, Lars Vilhuber, & William C. Block; Cornell Institute for Social & Economic Research

Recent years have shown the power of user-sourced information evidenced by the success of Wikipedia and its many emulators. Agencies on the other hand publish metadata on their data products, but have little opportunity to get structured feedback on that metadata (codebooks) from their users. Creating and augmenting metadata is a labor-intensive endeavor. Harnessing collective knowledge from actual data users can supplement officially generated metadata. As part of our Comprehensive Extensible Data Documentation and Access Repository (CED²AR) infrastructure, we demonstrate a prototype of crowdsourced metadata (codebooks), using DDI-C and supplemental XML. The system allows for any number of network connected instances (web or desktop deployments) of the CED²AR DDI editor to concurrently create and modify metadata. The backend transparently handles changes, and frontend has the ability to separate official edits (by designated curators of the data and the metadata) from crowd-sourced content. Features of the software allow official curators to merge in select enhancements into revisions of official documentation. CED²AR ingests and publishes widely-used DDI metadata standards, and can thus easily augment existing metadata publishing systems. Using standards-compliant ingest and publication methods, our proposed crowd-sourced metadata system can greatly enhance user acceptance of documentation, user-engagement in the documentation process, and reduce agency costs while improving the quality of data documentation. This development is part of the NSF-Census Research Network node at Cornell University.