Collaborative Editing and Versioning of DDI Metadata: The Latest from Cornell's NCRN CED2AR Software

Benjamin Perry, Cornell University (bap63@cornell.edu)

Venkata Kambhampaty, Cornell University (vrk4@cornell.edu)

Kyle Brumsted, McGill University (kyle.brumsted@mail.mcgill.ca)

William Block, Cornell University (wcb67@cornell.edu)

This presentation reports on the latest innovations of the Comprehensive Extensible Data Documentation and Access Repository (CED2AR), part of Cornell’s NSF-Census Research Network Project. CED2AR is an online repository for metadata on surveys, administrative microdata, and other statistical information. CED2AR uses DDI 2.5 and supplemental XML to document these studies without the use of any relational databases, making the entire repository is extremely portable and lightweight. The user interface not only provides the means to navigate and search across multiple codebooks, but the ability to edit DDI through simple and intuitive web forms. This allows us to collaboratively curate DDI metadata.

However, in addition to the study it describes, the DDI document itself evolves over time. Tracking and documenting these changes is an essential part of a researcher’s job. Rather than modifying the existing DDI Codebook schema, the CED2AR project uses Git, a distributed version control system traditionally built for software development. Git allows CED2AR to show how DDI metadata changes over time, and can retrieve old versions or parallel branches of the metadata. Git runs passively in the background, so CED2AR begins versioning its codebooks without any user intervention, however the more advanced features of Git are still accessible.