# C2Metadata Project

**JARED LYLE**

**Abstract:**
Accurate and complete metadata is essential for data sharing and for interoperability across different data types. However, the process of describing and documenting scientific data has remained a tedious, manual process even when data collection is fully automated. Researchers are often reluctant to share data even with close colleagues, because creating documentation takes so much time.

This presentation will describe a project to greatly reduce the cost and increase the completeness of metadata by creating tools to capture data transformations from general purpose statistical analysis packages. Researchers in many fields use the main statistics packages (SPSS®, SAS®, Stata®, R) for data management as well as analysis, but these packages lack tools for documenting variable transformations in the manner of a workflow system or even a database. At best the operations performed by the statistical package are described in a script, which more often than not is unavailable to future data users.

Our project will develop new tools that will work with common statistical packages to automate the capture of metadata at the granularity of individual data transformations. Software-independent data transformation descriptions will be added to metadata in two internationally accepted standards, the Data Documentation Initiative (DDI) and Ecological MetadataLanguage (EML). These tools will create efficiencies and reduce the costs of data collection, preparation, and re-use. Our project targets research communities with strong metadata standards and heavy reliance on statistical analysis software (social and behavioral sciences and earth observation sciences), but it is generalizable to other domains, such as biomedical research.

**About the Author:**

*Jared Lyle* is an Archivist at the Inter-university Consortium for Political and Social Research (ICPSR), where he directs the Curation Services Unit, which is responsible for Metadata, the Bibliography of Data-Related Literature, and Digital Preservation. He also serves as Director of the Data Documentation Initiative (DDI), an international metadata standard for describing survey and other social science data.