Embodying Research Methods into Fields and Tables:
A Process of Informed Database Design

L. Wynholds	David Fearon	Christine L. Borgman	Sharon Traweek
wynholds@ucla.edu	dfearon@ucla.edu	borgman@gseis.ucla.edu	traweek@history.ucla.edu

Department of Information Studies
University of California, Los Angeles

ABSTRACT

One of the invisible aspects of large research projects in the social sciences is the method by which observations and other collected data are managed. In sufficiently large projects, it may be effective to address the data management problem at the outset by creating a database architecture and data processing workflow. Research methods, assumptions and technical limitations often drive the structure of the data to be collected, but this is rarely discussed within the framework of the research. This design process represents a complex selection and trade-off matrix of predictive approximation, given that aspects of the analysis are not performed until the data is collected, and the design is done before the data collection is started. An elegant design can afford an equally elegant analysis of the data, but also creates a cycle where the data structure dictates the focus and granularity of the analysis.

We were faced with the problem of creating a system to support the projected data collection projects for a major, multi-method, 5-year research project on data curation practices. Our research focuses on specific techno-social practices of astronomers and will rely on a large volume of complex and heterogeneous source materials, such as email archives, scholarly publications, websites, reports, metadata headers, as well as in-person interviews.

The research questions focus on the data management, curation, and sharing practices of astronomers, how these practices evolved, and mapping who shares what, when, with whom, and why, with specific interest in what data they generate, use, keep and discard. We also ask what is most important to curate, and how do they do so, what do they expect to use and decide will be of future use to others, and who do they envision as future users?

The database structure will act as the connective tissue for the full term of the project while embodying the research methods, facilitating analysis, enabling data sharing, and minimizing effort. However, the process also represents a complex selection and trade-off matrix of predictive approximation of the intended analysis as the design defines the data set and the data set drives the analysis. This process-oriented poster documents the matrix we followed, the challenges and the solutions developed while operationalizing a data system for a large research project with major relational and descriptive aspects. Our resulting system utilizes existing competencies and departmental resources while meeting basic prerequisites for data security, sharing, interoperability, best practices and extensibility.