**Getting Started Creating Data Dictionaries: Creating Shareable Datasets**

Erin M. Buchanan, Ph.D., Missouri State University

Sarah E. Crain, Missouri State University

Arielle Cunningham, B.S., Missouri State University

Hannah R. Johnson, Missouri State University

Hannah Stash, Missouri State University

\*Note Ari listed as first author for poster presentation (Dr. B last)

[erinbuchanan@missouristate.edu](mailto:erinbuchanan@missouristate.edu), [crain998@live.missouristate.edu](mailto:crain998@live.missouristate.edu), [arielle924@live.missouristate.edu](mailto:arielle924@live.missouristate.edu), [hannah11698@live.missouristate.edu](mailto:hannah11698@live.missouristate.edu), [stash19@live.missouristate.edu](mailto:stash19@live.missouristate.edu)

**Short abstract:** Data dictionaries are metadata documents that allow researchers to make their data more open and easier to interpret. In this presentation, we will demonstrate the data dictionary applications available and explore how researchers might get started with creating open, shareable data.

**Medium abstract (for Great Plains):** Progress in science depends on collaboration, but much of the scientific process happens out of sight in the lab, and often only the final publication is available. Open data is beneficial for both individual researchers and science, because it facilitates the spread of knowledge and improvements in research. As one open data solution, data dictionaries are metadata documents that allow researchers to make their data more public and easier to interpret. We looked into three potential data dictionary creators—Codebook, DataSchema, and DD Creator—each with its own benefits and limitations. Therefore, in this presentation, we will demonstrate the applications available and explore how researchers might get started with creating open, shareable data.

**Full abstract:**

**Problem:**

Progress in science depends on collaboration, but much of the scientific process happens out of sight in the lab, and often only the final publication is available. This process has led to problems such as *p*-hacking, analysis errors, and mass failure to replicate—all without the potential for detection and, ultimately, correction. One potential solution is to make the research process more public. Brian Nosek and Jeffrey Spies began the Open Science movement to aid other scientists by creating an online platform in which they could openly record, report, and share data (Nelson, Simmons, & Simonsohn, 2018). Open data is beneficial for both individual researchers and science, because it facilitates the spread of knowledge and improvements in research. In fields where *p*-hacking and false positives run rampant, such as psychology, open data discourages fraud and makes replication more likely (Piwowar, 2013). Still, many researchers have concerns about opening their data. Concerns include copyright laws, time spent working to make data open, and, perhaps most importantly, participant confidentiality. We believe, nonetheless, that data can safely be made open in most cases.

**Procedure:**

One way to make data more open is to use and share a data dictionary: a document that contains metadata describing a dataset. Metadata documents help researchers understand the dataset more efficiently and can allow for more closely matched replications. We looked into three potential data dictionary creators, each with its own benefits and limitations. Each researcher can choose which works best for his or her data and research goals.

**Results:**

Codebook has the simplest procedure and provides useful information such as histograms, but allows only limited editing of information. DataSpice 2.0 works well in conjunction with Schema.org, which is a standardized set of practices for sharing computer readable information. Finally, DDCreator requires more data entry input but allows for more detailed descriptions, such as levels of a variable. It also outputs a data file with attributes embedded (Rdata).

**Conclusions:**

This project fits in with the broader perspective of the Psychological Data Structure project, which is creating specifications on open data practices. Therefore, in this presentation, we will demonstrate the applications available and explore how researchers might get started with creating open, shareable data.

**Mailing address:**

901 S. National Ave

Springfield, MO 65897

417-836-5592

**Submission type:** Talk/Poster