Recap of Workshop 1

September 18-20, 2015, Boulder
Workshop aim

1. To introduce the idea of *reproducible research* in linguistics
2. To identify the needs, skills, and values of several intra-linguistic communities
   a. OWLs (ordinary/outspoken/official working linguists)
   b. Journal editors
   c. Language archivists
   d. IT/Big Data specialists
3. To understand how other fields value and handle data citation
   a. Tracking metrics
   b. Models: proxy article, published data sets, coauthorship
   c. The progress of discussion in other fields (such as polar research and earth sciences)
4. To identify areas of need for more action
5. To identify homework tasks before workshop 2
Workshop structure - Friday morning
Featured Presentations

● Ruth Duerr: *State of the art in ethno-scientific data management*
  ○ Pan-scientific discussion of layers of data management, the data lifecycle, data levels, and the data citation landscape.
  ○ Many science fields are grappling with these same issues.

● David Carlson: *Data publication opportunities and challenges*
  ○ Data publishing in journals - *Earth System Science Data*
  ○ Publishing short articles about archived data sets
  ○ This becomes the citation of the data set

● Lauren Gawne and Andrea Berez-Kroeker: *Data citation: State of the art in linguistics*
  ○ Overview of citation practices of linguistic examples in 9 journals
  ○ Little data citation happening regularly so far
Workshop structure - Friday morning
Featured Presentations

● Peter Pulsifer: *Simple integration of data citation into research practice*
  ○ Discussed similar efforts in the Polar Research community
  ○ Including bringing together multiple bodies, multipronged education efforts, and promoting a culture of data publication

● Gary Holton: *Issues in data attribution*
  ○ Surveyed several models for academic attribution, including co-authorship, proxy publications, review articles, and peer review of data sets
  ○ Examined *verifiability* versus *reproducibility*
Workshop structure - Friday afternoon/Saturday Mini-presentations & Working Groups

● Current citation practices in journals and subfields
  ○ Felix Ameka: *JALL*
  ○ Keren Rice: *IJAL*
  ○ Nick Thieberger: *LD&C*
  ○ Stan Dubinsky: *Language*
  ○ Julie Legate: *NLLT*

● Evaluation
  ○ Tony Woodbury: *Reviewing and evaluating archived collections*
  ○ Richard and Sally: *Evaluating datasets in tenure review*
  ○ Meagan Daily, Ryan Henke, Jaime Perez Gonzalez, Nick Williams: *Evaluating dissertations based on primary data collections*
  ○ Susan Smythe Kung: *Altmetrics and archives*
Workshop structure - Friday afternoon/Saturday Mini-presentations & Working Groups

● Data Packaging and Archives
  ○ Gary Holton: Granularity
  ○ Paul Trilsbeek, Susan Kung, Laura Welcher and Mark Turin: The roles of archives in promoting data citation and attribution
  ○ Mandana Seyfeddinipur: Issues of access to primary data

● Digital Humanities
  ○ Gary Simons: Co-authorship as a means of crediting data creators
  ○ Tanya Clement: Data citation & attribution in the Digital Humanities

● Educating the Linguistics Community
  ○ Peter Pulsifer: Education and training the discipline
  ○ Geoff Nathan: The role of LSA/CoSCIL in standardizing citation and attribution practices in linguistics
Workshop structure - Sunday morning
Homework Assignments

- [View list](#)
Working Group Final Reports

- **Archiving Community Working Group**
  - Archives are the permanent hosts for data sets
  - Archives can provide the DOI for data sets
  - Language Documentation is the subfield most versed in this kind of archiving
    - We can educate other linguists to evaluate IRs, etc.
  - Archives can collaborate with Journal Editors

- **Journal Editors Community Working Group**
  - We support citing data sources in the best possible way
  - We need a unified stylesheet for citations
  - We need some guidance from the field re what constitutes the data set
  - We would appreciate citation templates (e.g. Zotero, Endnote, Mendeley…)
  - We support citation tracking when possible
Working Group Final Reports

- **Ordinary Working Linguist Community**
  - We need improved training of graduate students and early career researchers in data management and archiving
  - We need a distributed culture that encourages proper data management at all stages
  - We need to adopt metrics for evaluating data sets in tenure & promotion
  - We need help evaluating ('certifying') archives/IRs as homes for our data
  - We need solutions for managing the costs of archiving

- **IT/Big Data Community Working Group**
  - Our role is to make known to linguists what kind of technology is available
  - Funding and sustainability are our biggest challenges
  - DOIs are key to citation tracking, which is new to linguistics
  - OLAC needs continued support and development