The Austin Principles of Data Citation in Linguistics

Helene N. Andreassen
UiT The Arctic University of Norway

FiNo, 8-9 February 2018
Lund University
Why publish research data?

• For others
  • Verification
  • Reuse

• For yourself
  • Visibility of the complete output of your research
  • Allow readers to understand your research in even more depth
  • Secure proper reuse of your data

• Growing requirement by funders and journal editors
TROLLing
The Tromsø Repository of Language and Linguistics

https://sites.google.com/a/hawaii.edu/data-citation/welcome
https://dataverse.no/dataverse/trolling
Linguistics Data Interest Group (LDIG)

Co-chairs

Helene N. Andreassen
UiT The Arctic University of Norway

Andrea L. Berez-Kroeker
University of Hawai‘i at Manoa

Lauren Gawne
La Trobe University

https://www.rd-alliance.org/groups/linguistics-data-ig
Linguistics Data Interest Group (LDIG)

Main objectives

1. Development and adoption of common principles and guidelines for data citation and attribution by professional organisations, such as the Linguistic Society of America and the Societas Linguistica Europaea, academic publishers, and archives for linguistic and language data.

2. Education and outreach efforts to make linguists more aware of the principles of reproducible research and the value of data creation methodology, curation, management, sharing, citation and attribution.

3. Efforts to ensure greater attribution of linguistic data set preparation within the linguistics profession.

https://www.rd-alliance.org/groups/linguistics-data-ig
The Austin Principles

Set of guidelines that enable linguists to make informed decisions regarding the accessibility and transparency of their research data.

Based on the FORCE11 Joint Declaration of Data Citation Principles.
The Austin Principles

1. Importance of data
2. Credit and attribution
3. Evidence
4. Unique identification
5. Access
6. Persistence
7. Specificity and verifiability
8. Interoperability and flexibility

Linguistics Data Citation

Preamble

Data is central to empirical linguistic research. Linguistic data comes in many different forms, and is collected and processed with a wide range of methods. Data citation recognizes the centrality of data to research. Furthermore, it facilitates verification of claims and repurposing of data for other studies.
The Austin Principles

For all types of data, all subfields of linguistics.

All comments, questions and reflections are HIGHLY WELCOME.
Linguistics Data Citation

Endorse

Data citation will only become standard practice in linguistics if, as individual researchers, we commit to improving data citation in our own work, and in our specific research communities.

By endorsing the Austin Principles you show other linguists that you support reproducible research. You can join the list of linguists who endorse the principles by adding your name to this Google Form.

If you would like to be more actively involved in shaping the future of linguistic data, we encourage you to join the Linguistics Data Interest Group as part of the Research Data Alliance.

We, the undersigned, endorse the Austin Principles of Data Citation in Linguistics. We support the idea that the data on which linguistic analyses are based are of fundamental importance to the field, and should be treated as such.

List of Endorsers (last updated January 17 2018)
Reproducible research in linguistics:
A position statement on data citation and attribution in our field.