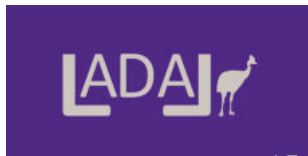# Establishing the
Language Technology and Data Analysis Laboratory
experiences, challenges, and future directions

Dr. Martin Schweinberger
www.martinschweinberger.de
m.schweinberger@uq.edu.au

Language Technology and Data Analysis Laboratory

What is LADAL?

- ▶ school-based, collaborative support infrastructure for digital humanities
- ▶ established and maintained by the School of Languages and Cultures at the University of Queensland
- ▶ assists with data processing, visualization, and analysis
- ▶ offers guidance on matters relating to language technology and digital research tools.

`https://slcladal.github.io/index.html`
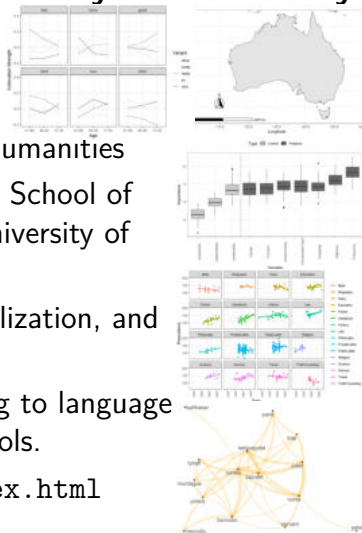
LADAL

Background and Motivation

Challenges

Experiences

Future Directions

Language Technology and Data Analysis Laboratory

Development of computational skills by providing information and practical, hands-on tutorials on data and text analytics

- ▶ Introductions to research designs and computational methods relevant to research involving natural language
- ▶ Aimed at people with different degrees of prior knowledge
- ▶ F-2-F consultations about methodology
- ▶ Provision of self-guided study materials
  - ▶ quantitative reasoning and basic concepts in empirical language studies
  - ▶ R as programming environment
  - ▶ basic of (language) data science
  - ▶ data visualization and data analytics
  - ▶ text analysis/mining, distant reading, corpus linguistics
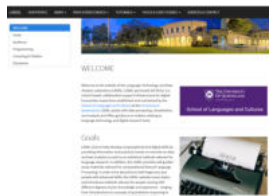
# Background and Motivation

A Computational Economy in a Digital Era

Computational approaches to processing and transforming, analyzing and visualizing data are becoming increasingly prevalent in the economy and in the humanities
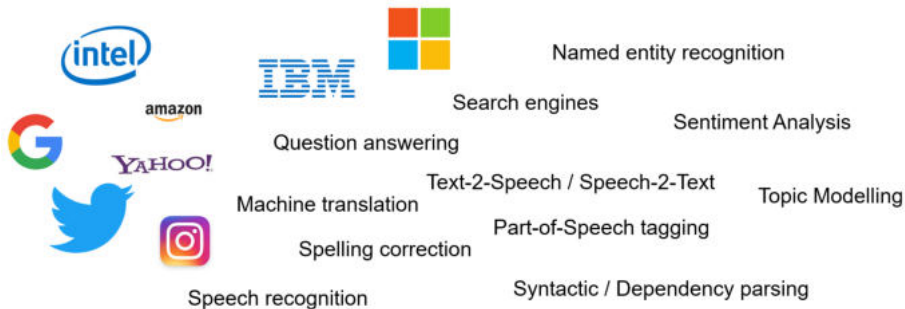
Humanities and Computing

Vast potential for the humanities by extending computational methods to humanties research (spreading the tentacles)

▶ Big Data (e.g. mega corpora)

▶ Multimodal / multipurpose data sets and digital archiving

▶ Mapping (combining data and geographical space)

▶ Real world applications (applying corpus-based research in classrooms)
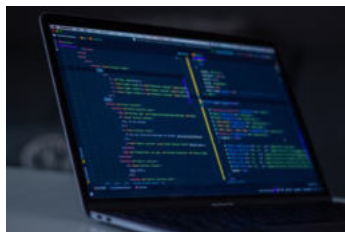
Creating a computationally capable and philosophically informed workforce that is competitive on the private and public sector market

Replication Crisis

Linguistics
An Interdisciplinary Journal of the Language Sciences
Editor-in-Chief: Gast, Volker

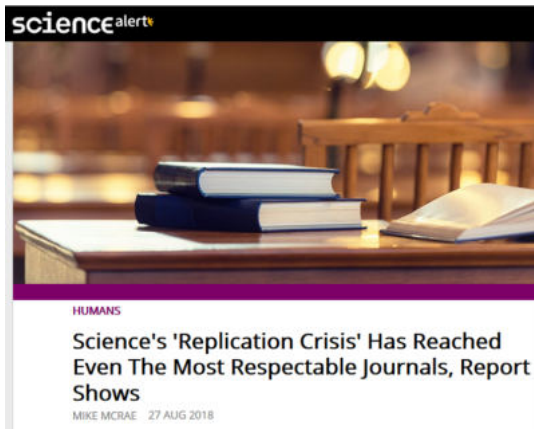**Band 56, Heft 1** · Vorheriger Artikel Nächster Artikel ›

Reproducible research in linguistics: A position statement on data citation and attribution in our field

Andrea L. Berez-Kroeker ☒ / Lauren Gawne / Susan Smythe Kung / Barbara F. Kelly / Tyler Heston / Gary Holton / Peter Pulsifer / David I. Beaver / Shobhana Chelliah / Stanley Dubinsky / Richard P. Meier / Nick Thieberger / Keren Rice / Anthony C. Woodbury

Online erschienen: 08.12.2017 | DOI: https://doi.org/10.1515/ling-2017-0032

🔲 Flyer drucken
📖 · Alert: eTOC ›

Citations 6

Replication Crisis

HUMANS

Science's 'Replication Crisis' Has Reached Even The Most Respectable Journals, Report Shows

MIKE MCRAE 27 AUG 2018

Linguistics
An Interdisciplinary Journal of the Language Sciences
Editor-in-Chief: Gast, Volker

≣ Band 56, Heft 1 › Vorheriger Artikel Nächster Artikel ›
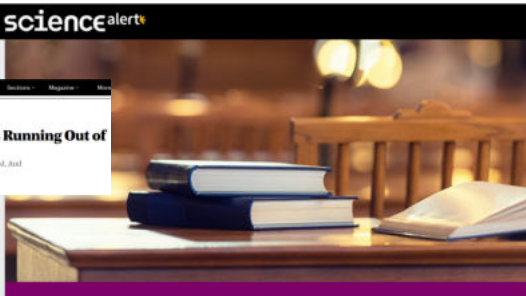
Reproducible research in linguistics: A position statement on data citation and attribution in our field

Andrea L. Berez-Kroeker / Lauren Gawne / Susan Smythe Kung / Barbara F. Kelly / Tyler Heston / Gary Holton / Peter Pulsifer / David I. Beaver / Stanley Dubinsky / Richard P. Meier / Nick Thieberger / Keren Rice / Anthony C. Woodbury

Online erschienen: 08.12.2017 | DOI: https://doi.org/10.1515/ling-2017-0032

Replication Crisis

sciencealert*

SCIENCE

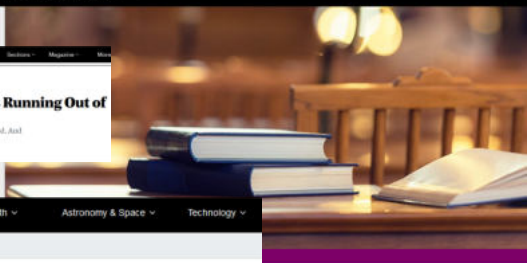**Psychology's Replication Crisis Is Running Out of Excuses**

Another big project has found that only half of studies can be repeated. And this time, the usual explanations fall flat.

HUMANS

Science's 'Replication Crisis' Has Reached Even The Most Respectable Journals, Report Shows

MIKE MCRAE 27 AUG 2018

Linguistics
An Interdisciplinary Journal of the Language Sciences
Editor-in-Chief: Gast, Volker

Flyer drucken

Alert: eTOC

≡ Band 56, Heft 1 › Vorheriger Artikel Nächster Artikel ›

Reproducible research in linguistics: A position statement on data citation and attribution in our field

Andrea L. Berez-Kroeker / Lauren Gawne / Susan Smythe Kung / Barbara F. Kelly / Tyler Heston / Gary Holton / Peter Pulsifer / David I. Beaver / Stanley Dubinsky / Richard P. Meier / Nick Thieberger / Keren Rice / Anthony C. Woodbury

Online erschienen: 08.12.2017 | DOI: https://doi.org/10.1515/ling-2017-0032

Citations 6

Replication Crisis

Replication Crisis

Replication Crisis

Solutions

Restoring Public Trust through substantive upskilling efforts in computational literacy in the humanities

Code

- ▶ Coding rather than tool-based analyses
  ($\rightarrow$ transparency and replicability)

Data

- ▶ FAIR principals: Findable,
  Accessible, Interoperable, Re-usable

Open Science

- ▶ Adoption of Best Practices
  (Documentation)
- ▶ Open access repositories
- ▶ Pre-registration

Challenges

Challenges

- ▶ Best practices and transparency in research (access to/sharing of data, naming conventions, version control, folder structure, collaborative writing, open access repos)

- ▶ (Over-)reliance on existing (commercial) tools (hinders both reproducibility and efficiency)

- ▶ Unwillingness to give up on accustomed practices (SPSS, licences, SketchEngine) or lack of time to learn/improve

- ▶ Vastly different needs within and across disciplines (differences in experience and expectations)

Lack of programs, materials and professional training (both for general and specialized audiences)

Experiences

Experiences

Positive

- ► Hugh uptake (demand is there!)
  → mostly for easy solutions...
- ► Tremendous support from the school!
- ► Development of resources and workshop sessions
- ► International collaborations and contributors
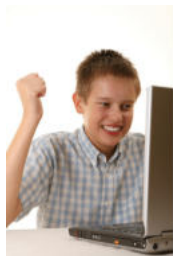
Negative

- ► Lack of long-term funding
- ► Restricting policies and unsupportive admin/professors
- ► Development of programs difficult due to drastically different and diverging interests/expectations
- ► Tendency to go with easy, cheap but non-sustainable solutions (Potemkin villages)

Future Directions

Future Directions

- Experts and centers/programs for language data science
    - Professors/lecturers/postdocs in social and cultural data analytics/language data science/DH
    - Programs in social, cultural, linguistic data science
    - Software engineers to set up/improve platforms
- Systematic collaboration is needed between centers and researchers within and across universities and faculties to address the existing skills shortage
- LADAL
    - Building international collabs
    - Adding/Extending content
    - Professionalize design
    - Add functionality

Thank you very much!

Acknowledgements

I would like to thank. . .

the School of Languages and Cultures at the University of Queensland, Australia, and,

in particular, Michael Haugh, without their substantive support establishing LADAL

would not have been possible!

Aschwanden, C. (2018). Psychology's replication crisis has made the field better.
https://fivethirtyeight.com/features/psychologys-replication-crisis-has-made-the-field-better/.

Berez-Kroeker, A. L., L. Gawne, S. S. Kung, B. F. Kelly, T. Heston, G. Holton, P. Pulsifer, D. I. Beaver,
S. Chelliah, S. Dubinsky, et al. (2018). Reproducible research in linguistics: A position statement on data
citation and attribution in our field. *Linguistics 56*(1), 1–18.

Diener, E. and R. Biswas-Diener (2019). The replication crisis in psychology.
https://nobaproject.com/modules/the-replication-crisis-in-psychology.

McRae, M. (2018). Science's 'replication crisis' has reached even the most respectable journals, report shows.
https://www.sciencealert.com/replication-results-reproducibility-crisis-science-nature-journals.

Resnick, B. (2018). More social science studies just failed to replicate. here's why this is good.what scientists learn
from failed replications: how to do better science.

Velasco, E. (2019). Researcher discusses the the science replication crisis.
https://phys.org/news/2018-11-discusses-science-replication-crisis.html.

Weir, K. (2015). A reproducibility crisis? the headlines were hard to miss: Psychology, they proclaimed, is in crisis.
Monitor on Psychology 46, 39.

Yong, E. (2018). Psychology's replication crisis is running out of excuses. another big project has found that only
half of studies can be repeated. and this time, the usual explanations fall flat.
https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/.

Establishing the
Language Technology and Data Analysis Laboratory
experiences, challenges, and future directions

Dr. Martin Schweinberger

www.martinschweinberger.de

m.schweinberger@uq.edu.au