

Open research practices: Value vs sustainability (part II)

Emma Marsden & Cylcia Bolibaugh, University of York
Open Scholarship in Applied Linguistics June 2022

cylcia.bolibaugh@york.ac.uk | [twitter: @CBolibaugh](https://twitter.com/CBolibaugh) | CC - BY 4.0 | [doi 10.5281/zenodo.6625659](https://doi.org/10.5281/zenodo.6625659)

Zooming in on open data

Why open?

Moving to open research 2.0

Where are we now in Applied
Linguistics?

Why open?

See Gilbert & Corker (2017). [Research Transparency: 5 questions about open science answered.](#)

Corker (2018) osf.io/5rvc/

Scientific Norms & Values (Merton, 1942)

Communalism

Scholarship is for everyone



Transparency

Nullius in verba

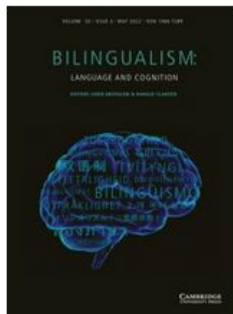


See Lupia (2012). [What's the value of social science?](https://doi.org/10.5281/zenodo.3572600) <https://doi.org/10.5281/zenodo.3572600>

Corker (2018) osf.io/5ravic/

Why share data?

- Computational reproducibility (verify)
- Analytic robustness (reconceptualise)
- Research synthesis (evaluate and build)



Bilingualism:
Language and
Cognition

Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research

Published online by Cambridge University Press: 27 August 2021

[Cylcia Bolibaugh](#) , [Norbert Vanek](#) and [Emma J. Marsden](#)

[Show author details](#) ▼

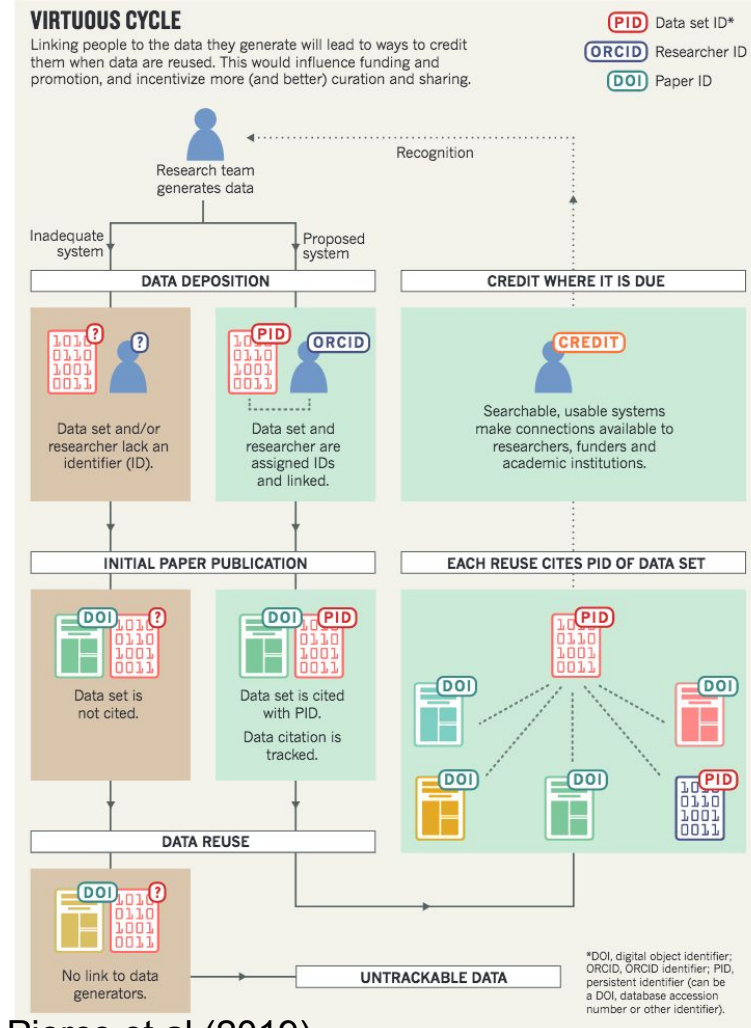
Article

Metrics

Why share data?

Papers with publicly available datasets receive a higher number of citations than similar studies without available data.

- Piwowar et al 2007 (> 69%)
- Piwowar et al 2013 (>9%)
- Colavizza et al 2020 (>25%)



Why share data?

- Provide the URL, DOI, or other permanent path for accessing the data in a public, open access repository.
- Is there sufficient information for an independent researcher to reproduce the reported results? If no, explain.

The Open Data badge recognizes researchers who make their data publicly available, providing sufficient description of the data to allow researchers to reproduce research findings of published research studies.

[Badges to Acknowledge Open Practices](#)



Open Research 1.0

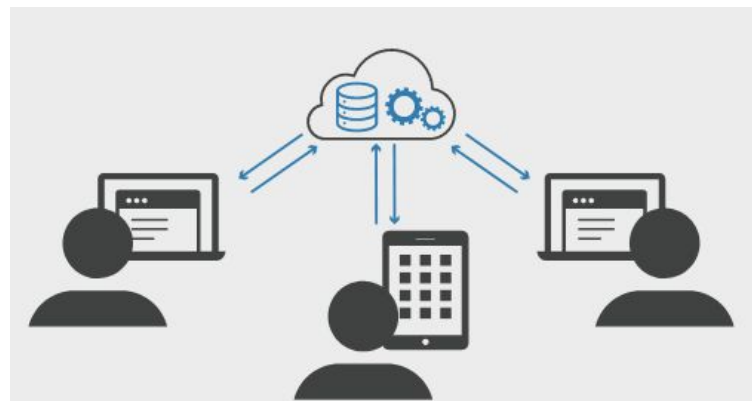
- Largely driven by reproducibility crisis
- Focus on transparency and accessibility of individual outputs related to publications
- Development of 1.0 infrastructure (internet, repositories, dois)
- Characterised by a “just do it” and “it’s good for you” approach



vs

Open Research 2.0

- A move from a focus on verification to a more synthetic way of working (reuse, reconceptualise, ...)
- Tools and support for decentralised, networked research (multi-site replications, manylabs type approaches)
- Recognition (for individuals) for a wider variety of research outputs and contributions

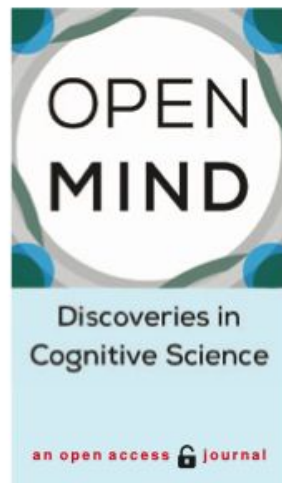


Moving to Open Research 2.0

Data reuse case 1: Synthesise

“We examined the replicability of a seminal study that showed monolingual–bilingual differences in infancy (Kovács & Mehler, 2009a) by collecting new data from 7-month-olds and 20-month-olds and reanalyzing three open datasets from 7–9 month-olds (D’Souza et al., 2020; Kalashikova et al., 2020; 2021).” (Dal Ben et al, 2022)

“We found all aspects of the multisite registered replication approach to be useful although the registration component itself appeared to be an especially feasible and valuable first step toward increasing the robustness and generalizability of findings in our field.” (M.- S. et al, 2018)



REPORT

Bilingualism Affects Infant Cognition: Insights From New and Open Data

Rodrigo Dal Ben^{*ID}, Hilary Killam^{ID}, Sadaf Pour Eliaei^{*ID}, and Krista Byers-Heinlein^{ID}

Concordia University

*Rodrigo Dal Ben is now at Ambrose University, Calgary. Sadaf Pour Eliaei is now at the Ontario Institute for Studies in Education, University of Toronto.

Keywords: infancy, bilingualism, cognitive control, inhibitory control, anticipatory looking

Downloaded from



EMPIRICAL STUDY |  Open Access |  

Multisite Replication in Second Language Acquisition Research: Attention to Form During Listening and Reading Comprehension

 **Correction(s) for this article** ▼

Kara Morgan-Short[✉], Emma Marsden, Jeanne Heil, Bernard I. Issa II, Ronald P. Leow, Anna Mikhaylova, Sylwia Mikołajczak, Nina Moreno, Roumyana Slabakova, Paweł Szudarski

Data reuse case 2: Reconceptualise

This repo contains the data reported in Hartshorne, Tenenbaum, & Pinker. A Critical Period for Second Language Acquisition: Evidence from 2/3 Million English Speakers: <https://osf.io/pyb8s/>

Codes for: van der Slik, Schepens, Bongaerts, & van Hout. Critical Period Claim Revisited: Re-analysis of Hartshorne, Tenenbaum, and Pinker (2018) Suggests Steady Decline and Learner-Type Differences. <https://osf.io/gqm87/>



Original Articles

A critical period for second language acquisition: Evidence from 2/3 million English speakers

Joshua K. Hartshorne ^{a, b}, Joshua B. Tenenbaum ^a, Steven Pinker ^c

[Show more](#) ▾

[+](#) Add to Mendeley [🔗](#) Share [🗣️](#) Cite

<https://doi.org/10.1016/i.cognition.2018.04.007>

[Get rights and content](#)

LANGUAGE LEARNING

A Journal of Research in Language Studies

EMPIRICAL STUDY | [Open Access](#) |

Critical Period Claim Revisited: Reanalysis of Hartshorne, Tenenbaum, and Pinker (2018) Suggests Steady Decline and Learner-Type Differences

Frans van der Slik , Job Schepens, Theo Bongaerts, Roeland van Hout,

First published: 07 September 2021 | <https://doi.org/10.1111/lang.12470>

Data reuse case 3: Build

About MetaLab

Interactive, community-augmented
meta-analysis tools for cognitive
development research

The [MetaLab database](#) contains **2,497 effect sizes** from **30 meta-analyses** across two domains of cognitive development, based on data from **688 papers** and **45,260 subjects**.

Applications

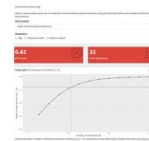
Visualization

Explore a variety of interactive charts driven by the MetaLab database by your datasets and moderators



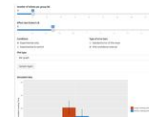
Power Analysis

Analyzes power of your experiment under a variety of conditions



Power Simulation

Simulate power of meta-analyses under a variety of conditions



Domains



Early Language

How do children learn their native language?

**24 meta-
analyses**

551 papers

**2,135
effect sizes**

**38,977
subjects**



Cognitive Development

What is the nature of children's understanding?

**6 meta-
analyses**

137 papers

**362 effect
sizes**

**6,283
subjects**

Where are we now in Applied
Linguistics?

Data sharing: A primer from UKRN

What is data sharing?

Data sharing is the process of making [research data](#) available for wider dissemination. The **FAIR principles** state data and metadata should be: **findable** (easy to find for both humans and computers); **accessible** (accompanied by clear instructions for access and authorisation); **interoperable** (compatible with other data and/or tools); and **reusable** (suitably described to allow further use).

The data itself

To enable shared data to be used effectively, they should be [understandable to humans and machines](#), which means conforming to standards. [Meta-data standards](#) facilitate long-term use and integration; some also apply to [specific kinds of data](#). The DCC keeps a [list of standards](#) you can check. [Roche et al \(2015\)](#) provide several generally-applicable recommendations for making data complete and reusable, including:

Two questions

→ Prevalence of data sharing?

→ FAIR-ness of data sharing?

An open question, but ...

- ◆ Roche et al 2015
- ◆ Mons et al 2017
- ◆ Towse et al 2021

Datasets at IRIS: a peek under the hood

- Disciplinary scope (37 Journals, AAAL, BAAL)
- Datasets:
 - ◆ 74 in press/2022
 - ◆ 44 from 2021
 - ◆ 24 from 2020
 - ◆ 17 from 2019
 - ◆ 16 from 2018
 - ◆ 14 from 2017
- Findable: PID (url / DOI)
- Accessible: cc-by-sa



iris-database.org

BUT WHAT ABOUT Interoperable and Reusable components? (e.g. are the datasets complete, interpretable and reusable?)

MA project
Spring
2022

Datasets at IRIS: a peek under the hood



Criteria	Levels	Criteria_explanation	
PID	1 - present; 0 - not present	Stored in a 3rd party repository with a permanent link (of which a DOI is one well known example of a permanent identifier)	
License	1 - present; 0 - not present	clear instructions for access and authorisation - IRIS states license	
openFormat	1 - present; 0 - not present	non-proprietary format alongside original?	
minData	1 - present; 0 - not present	minimal: sufficient data to verify reported analysis	
maxData	1 - present; 0 - not present	Raw data files present	
missing	1 - present; 0 - not present	Explanation of missing participants, missing conditions, missing values (in paper or read-me)	
dictionary	1 - present; 0 - not present	well-described (variables names and levels within data file or with separate data dictionary read-me)	
code	1 - present; 0 - not present	is the analysis code to reproduce the analyses provided?	

Datasets at IRIS: a peek under the hood

- 9 datasets (2017 - 2022)
- Journals: BLC, CALICO, JML, LL, LAB, MLJ, SLR, SSLA
- Total scores (summed out of 8 possible criteria) by dataset
 - ◆ 3; n = 4
 - ◆ 4; n = 2
 - ◆ 5; n = 1
 - ◆ 6; n = 1
 - ◆ 8; n = 1



Criteria	total	n
License	9	9
minData	9	9
PID	9	9
dictionary	3	9
maxData	3	9
openFormat	3	9
code	2	9
missing	1	9

Moving to Open Research 2.0 (the data edition)

What do we need to move forward?

- Consensus - what constitutes “the data required to reproduce the findings in a published report”?
- Individuals - recognition at the institutional and award levels (e.g. UKRI “resume for researchers”)
- Support (& training) - researchers are not data archivists
- Funding - infrastructure (for OR 1.0 and 2.0!)



References

- 1-6. Bolibaugh, C., Vanek, N., & Marsden, E. J. (2021). Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research. *Bilingualism: Language and Cognition*, 15(4), e0230416.
- Colavizza, G., Hrynaskiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS one*, 15(4), e0230416.
- Dal Ben, R., Killam, H., Pour Iliaei, S., & Byers-Heinlein, K. (2021). Bilingualism affects infant cognition: Insights from new and open data. *Open Mind*, 1-30.
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263-277.
- Merton, R. K. (1942). A note on science and democracy. *J. Legal & Pol. Soc.*, 1, 115.
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49-56.
- Morgan-Short, K., Marsden, E., Heil, J., Issa Ii, B. I., Leow, R. P., Mikhaylova, A., ... & Szudarski, P. (2018). Multisite replication in second language acquisition research: Attention to form during listening and reading comprehension. *Language Learning*, 68(2), 392-437.
- Pierce, H. H., Dev, A., Statham, E., & Bierer, B. E. (2019). Credit data generators for data reuse. *Nature*, 570(7759), 30-33.
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS one*, 2(3), e308.
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175.
- Roche DG, Kruuk LEB, Lanfear R, Binning SA (2015) Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLoS Biol* 13(11): e1002295. <https://doi.org/10.1371/journal.pbio.1002295>
- Towse, J. N., Rumsey, S., Owen, N., Langford, P., Jaquiere, M., & Bolibaugh, C. (2020) Data Sharing: a Primer from UKRN. <https://doi.org/10.31219/osf.io/wp4zu>
- Towse, J.N., Ellis, D.A. & Towse, A.S. (2021). Opening Pandora's Box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. *Behav Res* 53, 1455–1468 <https://doi.org/10.3758/s13428-020-01486-1>
- van der Slik, F., Schepens, J., Bongaerts, T., & van Hout, R. (2022). Critical Period Claim Revisited: Reanalysis of Hartshorne, Tenenbaum, and Pinker (2018) Suggests Steady Decline and Learner-Type Differences. *Language Learning*, 72(1), 87-112.