

# REPRODUCIBLE RESEARCH

MSL, DURHAM, 2018

---

Tuomas Eerola

Updated: 23/03/2018

Reproducible\_Research

Durham University



1. What is Reproducible Research?
2. How to Achieve Reproducible Research?
3. Analysis
4. Sharing Data

References

# WHAT IS REPRODUCIBLE RE- SEARCH?

---

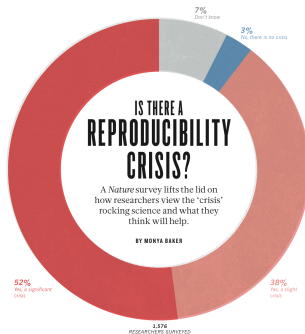
# What is Reproducible Research?

"A study is reproducible if there is a specific set of computational functions/analyses that exactly reproduce all of the numbers and data visualizations in a published paper from raw data. Reproducibility does not require independent data collection and instead uses the methods and data collected by the original investigator." (Marwick, 2016, p. 4)

# Why Reproducible Research?

## REPLICATION CRISIS

Crisis on replication and transparency in empirical research



Nature (26 May 2016), [doi:10.1038/533452a](https://doi.org/10.1038/533452a)

# Why Abundance of Irreproducible Research?

- We don't know better
- We have pressure to publish
- There is no incentive to produce reproducible research
- We do selective reporting (data subsets, p-value fishing, pos. results only)
- It keeps us artificially in the business (we are the only ones who know how to make sense of the data, etc.)

# Benefits of Reproducibility (1)

- Comply with the demands for transparency
  - first in computer sciences and biosciences
  - recently in social sciences (Asendorpf et al., 2013)
  - **Note: Open Access Data is required by all RCUK funding**
  - some journals promote reproducible research and data release (PLOS One, Nature, PNAS, a growing host of others)

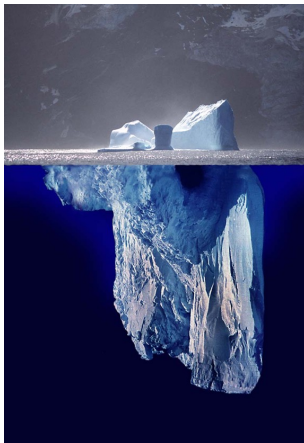
# Benefits of Reproducibility (2)

- To collaborate more easily and effectively
  - spot mistakes
  - encourage learning & exploration (Sandve, Nekrutenko, Taylor, & Hovig, 2013)
- Communicate research more clearly
  - workflows that integrate analysis and reporting
  - share data, figures, analysis and gain visibility (Piwowar, Day, & Fridsma, 2007)



# Reproducible Research

Article – is just a tip of the iceberg. Reproducible Research makes the whole workflow accessible.



# Types of Reproducible Research

3 kinds of reproducibility (Stodden, Leisch, & Peng, 2014):


- Computational reproducibility: code, software, hardware and implementation details.
- Empirical reproducibility: detailed information about non-computational empirical scientific experiments and observations. Basically the data.
- Statistical reproducibility: detailed information is provided about the choice of statistical tests.

# Reproducible Music Research?

Music Research is interdisciplinary, and some disciplines are closer to demands of reproducibility than others

- Music Information Retrieval (MIR) has started this (e.g., [soundsoftware.ac.uk](http://soundsoftware.ac.uk))
- Music Psychology follows the path (interest in **replication**)
- Music Intervention research (has to register study protocols)
- Music analysis using corpus studies (rare but increasing: Jazzomat project, Huron's Humdrum, JAMS proposal by Humphrey et al., 2014)

# HOW TO ACHIEVE REPRODUCIBLE RESEARCH?



# How to Achieve Reproducible Research?

- Workflows that are reproducible and transparent
  - Designs (sometimes requiring pre-registration, see OSF)
  - Sharing data (data, analysis pipelines, etc.)
  - Analysis (using tools that allow reproduction)
  - Reporting (linking data and analysis, + additional information)
  - See our shared data at <https://musicscience.net/resources/collections/>

# ANALYSIS



Analysis tools that facilitate reproducibility

- Statistics: R

## ASK YOURSELF

Can you easily replicate a complex analysis in SPSS? No!

# R Is Excellent Because It...

- is accessible, free, open source, available for all OS
- is completely programming driven (ie fully transparent)
- has excellent coverage of statistical modelling tools
- is pedagogical in many of its functionalities
- has good support for producing reports (R Markdown, knitr, Sweave) or interactive websites (Shiny).

See my [Reproducible Research using R - A template for analysing behavioural experiments](#) (separate pdf document)



## SHARING DATA

---

Requirement for UK Research Council Funded Projects.

- UK Data Reshare (<http://reshare.ukdataservice.ac.uk>)
- Harvard Dataverse (<https://dataverse.harvard.edu>)
- FigShare (<http://figshare.com/>)
- OpenScience (<https://osf.io>)
- Zenodo (<https://zenodo.org>)
- Researchcompendia  
(<http://researchcompendia.org/compendia/>)

Better than institutional repositories, keep track of the usage,  
ensure long-term storage

# EXAMPLE: Data Share

## DATVERSE



Harvard Dataverse (/dataverse/harvard)

A collaboration with Harvard Library, Harvard University

Harvard Dataverse (/dataverse/harvard) > **Data related to "Mild dissonance preferred over consonance in single chord perception"**

Metrics

7 Downloads

### **Data related to "Mild dissonance preferred over consonance in single chord perception"**

Eerola, Tuomas; Lahdelma, Imre, 2016, "Data related to "Mild dissonance preferred over consonance in single chord perception", doi:10.7910/DVN/GE5PPL (<http://dx.doi.org/10.7910/DVN/GE5PPL>), Harvard Dataverse, V1

Learn about  
p

#### Description

**Data related to Mild dissonance preferred over  
in single chord perception**

Eerola & Lahdelma, 2016, [doi:10.1038/533452a](https://doi.org/10.1038/533452a)

# EXAMPLE: UK Datashare

## UK DATA SHARE

UK Data Service  
Discover



Discover

Variable and question  
bank

QualiBank

About us

Get data

Use data

Manage data

Deposit data

News and events

Discover > Catalogue

### Catalogue

SHARE

UK Data Service data catalogue record for:

### Moods and activities in music

[Download](#) | [DDI XML](#)

#### TITLE DETAILS

**SN:** 852024  
**Title:** Moods and activities in music  
**Persistent identifier:** [10.5255/UKDA-SN-852024](#)  
**Depositor:** Tuomas Eerola, Durham University, UK  
**Principal investigator(s):** Tuomas Eerola, Durham University, UK  
Pasi Saari, Durham University, UK  
**Sponsor(s):** ESRC  
**Grant number:** ES/K00753X/1

#### SUBJECT CATEGORIES

Psychology  
Society and culture

# Good Ways to Promote Reproducibility

- Require reproducibility from PhD students
- Run replication studies in UG teaching (Music and Science)
- Make it one of your themes in lab meetings
- Take up the challenge of reproducibility in collaborations

“Reproducibility is like brushing your teeth. It is good for you, but it takes time and effort. Once you learn it, it becomes a habit.” (Baker, 2016)

# References

- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., ... others (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27(2), 108–119.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454.
- Marwick, B. (2016). Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory*, 1–27.
- 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.