

# Pointers to more effective software & data in audio research

Mark D Plumbley, Chris Cannam, Steve Welburn  
Centre for Digital Music  
Queen Mary, University of London

---

# Overview

- Introduction & Motivation
- Open Access Publication
- Research Data
- Research Software
- Wrap-up; Questions

---

# Dream: “Ideal” Research Pipeline

Researcher A (“Producer”)

- Read background papers
- Do own research
- Publish paper X

Researcher B (“Consumer-Producer”)

- Read paper X
  - Understand/reproduce results in paper X
  - Do more research building on X
  - Publish paper Y that cites X / produce product that uses X
- ... and so on.

---

# *Real* Research Pipeline

Researcher A (“Producer”)

- Read background papers
- Do own research (including lots of coding)
- Publish paper X (not enough space for all the code)

Researcher B (“Consumer-Producer”)

- Read paper X
- Can't reproduce or use results in paper X
- Tear out hair
- Give up / do something else

NB: A and B may be in same group (or same person later!)

---

# Reproducible Research

(Buckheit & Donoho, 1995; Vandewalle et al, 2009)

Idea: researchers should be able to reproduce the work of others.

Research used to be “reproducible” from the paper alone.

In audio & music research, methods are now too complex.

The paper is not enough: need algorithm, parameters, datasets, ...

So, we need

- The paper (ideally Open Access)
- The data (ideally Open Data)
- The software (ideally Open Source)

Well-known example: WaveLab (Buckheit & Donoho, 1995)

But in audio & music research, few people do this. Why?

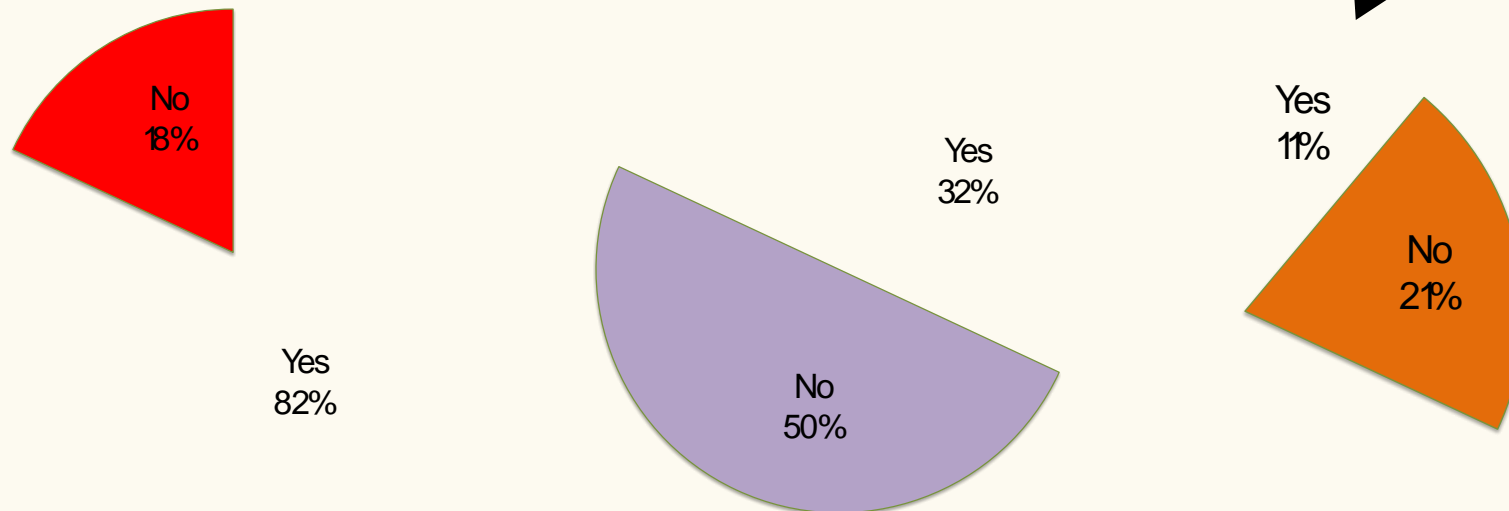
# Research software in practice

We carried out a **Survey of UK audio and music researchers\***.

82% developed software, but only 39% of those took steps to reproducibility,  
and only 35% of *those* published any code

only **11%** tried to be reproducible and published the code.

(Also: 51% said their code never left their own computer)



\* - Oct 2010-Apr 2011, 54 complete + 23 partial responses. For these figures we considered 72 responses.

---

# Why don't we publish code & data?

Our survey suggested:

- Lack of time
- Copyright restrictions
- Potential for future commercial use

Other factors (UK Research Information Network, 2010):

- Lack of evidence of benefits
- Culture of independence or competition
- Quality concerns (self-taught programmers)

Also: it takes effort early in the research cycle;  
hard to find time/motivation after the paper is published

---

# Reasons we don't like to admit?

J M Wicherts, M Bakker and D Molenaar, 2011, *Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results*, PLoS ONE

Does this cut both ways?

Can we improve quality by helping people prepare to share?

<http://bit.ly/vaU435>



---

# Barriers to publication and reuse

- Lack of education and confidence with code
- Lack of facilities and tools
- Lack of incentive for publication
- Platform incompatibilities

These are barriers to publication of *code*.  
Related issues for data.

---

# This Tutorial

- Break through these barriers!
- Improve the accessibility of research:
  - Research papers - open access publication
  - Research data
  - Research software

---

# Research Papers: Open Access

---

# Open Access Publication of Papers

What is “Open Access”

- Publication is free to download – and sharable? (Licence?)

Why make papers open access?

- Helps other researchers
- Increases impact of your research

Funder or institution policies (See <http://roarmap.eprints.org>)

- UK – RCUK - Journal and Conference papers (from Apr 2013)  
<http://www.rcuk.ac.uk/research/Pages/outputs.aspx>
- NIH Public Access Policy – Must submit to PubMed Central  
<http://publicaccess.nih.gov/policy.htm>
- Harvard – Open Access Policy <http://osc.hul.harvard.edu/policies>

---

# Types of Open Access

- Self-archiving
  - Can you upload your paper on your own website?
  - If yes, which version?
    - Preprint: submitted version, before reviewing)
    - Postprint (“Accepted Author Manuscript”):  
updated version, after reviews, before typesetting/proof
    - Publisher’s version: as appears in the journal/conference
- Paid Open Access – “Author Pays” model
  - Article available for free download from publisher
  - May or may not allow self-archiving too

---

# What does my journal allow?

Website: SHERPA/RoMEO: <http://www.sherpa.ac.uk/romeo/>

RoMEO Colours:

- Green - can archive pre-print and post-print or publisher's version/PDF
- Blue - can archive post-print (ie final draft post-refereeing) or publisher's version/PDF
- Yellow - can archive pre-print (ie pre-refereeing)
- White - archiving not formally supported

Also people talk about “Gold” for Paid Open Access

---

# RoMEO Examples: IEEE Trans ASLP

IEEE Transactions on Audio, Speech and Language Processing

- **Green** - pre-print, post-print, not publisher version (since 2011)
- Authors own and employers publicly accessible webpages OK

Special conditions:

- Preprint - Must be removed upon publication
- Set phrase required:  
"(c) 20xx IEEE. Personal use of this material is permitted.  
Permission from IEEE must be obtained ..."
- Etc.

---

# RoMEO Examples: JNMR

Journal of New Music Research

- **Yellow** – Pre-print OK, post-print embargo, not publisher's version
- Can archive pre-print (ie pre-refereeing)
- Can archive post-print after 18 month delay from publishing

Special conditions:

- Set phrase wording required e.g.:  
"This is an Author's Original Manuscript of an article whose final and definitive form, the Version of Record, has been published in the [JOURNAL TITLE] [date of publication] [copyright Taylor & Francis], available online at:  
[http://www.tandfonline.com/\[Article DOI\]](http://www.tandfonline.com/[Article DOI])."
- Paid open access option, fee: \$3,250/£1,725/€1,900



---

# Not all on RoMEO yet

E.g.

- Acta Acustica United with Acustica
  - Policy not on RoMEO
- Journal of the Audio Engineering Society
  - RoMEO colour “White” (“archiving status unclear”)
  - **Update:** New Open Access policy (not in guidelines yet)
  - Paid open access option, fee \$1200
- Could you help?

---

# Conferences? Book chapters?

## Conferences

- Vary widely – check the conference policy
- Read the wording in the author agreement you sign!

## Book chapters

- Not normally open access - read the agreement.
- Don't just sign agreement – Negotiate an “Author Addenda”!
  - E.g. Columbia University:  
<http://copyright.columbia.edu/copyright/copyright-ownership/publication-agreements/>

---

# Things to do tomorrow (or today!)

- Check your journals on RoMEO, read publisher policy
- Self-archive the papers you can (start with most important)
  - NB: Check the conditions, E.g. Set phrases to add
- Ask the publisher for permission to self-archive old papers
- Consider open access policies for your next journal/conference
  - But: Don't choose just because it says "Open Access"
  - Make sure you know about author fees (can be £/€/€ 1000+)
- Negotiate your next book chapter publishing agreement
- Ask your institution library for help – they'll want to help you!