soundsoftware.ac.uk

Reusable Software and Reproducibility in Music Research

Mark D Plumbley, Chris Cannam
Centre for Digital Music
Queen Mary, University of London
(with contributions from Steve Welburn, Luis Figueira, Simon Dixon)

EPSRC
Engineering and Physical Sciences Research Council

Overview

- Introduction & Motivation
- Research Software
- Research Data
- Open Access Publication
- 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?

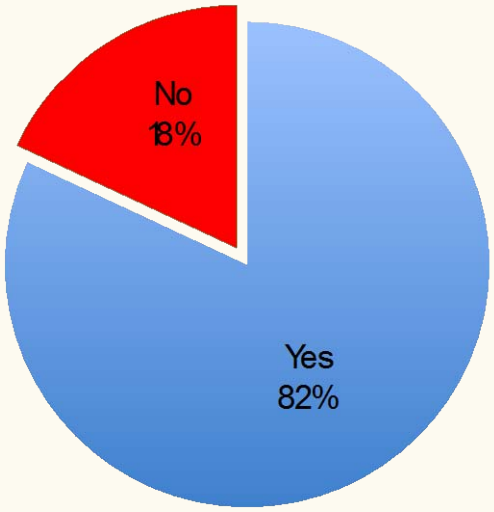
Research software in practice

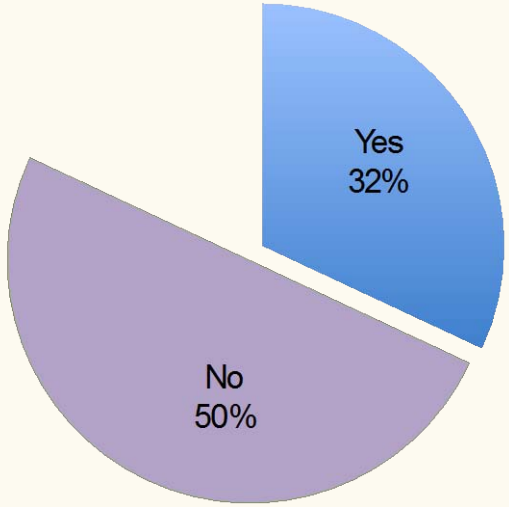
We carried out a **Survey of UK audio and music researchers***.

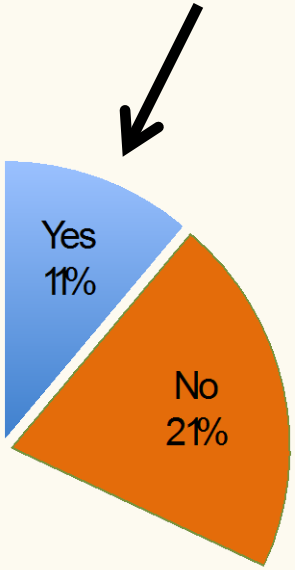
82% developed software, but only 39% of those took steps to reproducibility, and only 35% of *those* published any code

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 software
  - Research data
  - Research papers - open access publication