soundsoftware.ac.uk

SoundSoftware.ac.uk: Software sustainability for the audio and music researcher

Chris Cannam, Mark Plumbley, Luís Figueira
Centre for Digital Music
Queen Mary, University of London

EPSRC
Engineering and Physical Sciences Research Council

Who (and why) are we?

EPSRC-funded four-year project, 2010–2014:

Run from the Centre for Digital Music at QMUL

Serving the whole UK audio and music research community

Supporting the sustainable development and use of software and data to enable high quality research

Better science through better handling of software and data

Reproducible research

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

This is no longer possible: too complex, the paper is not enough

Hence “Reproducible Research”:

The paper – ideally Open Access

The software – ideally Open Source

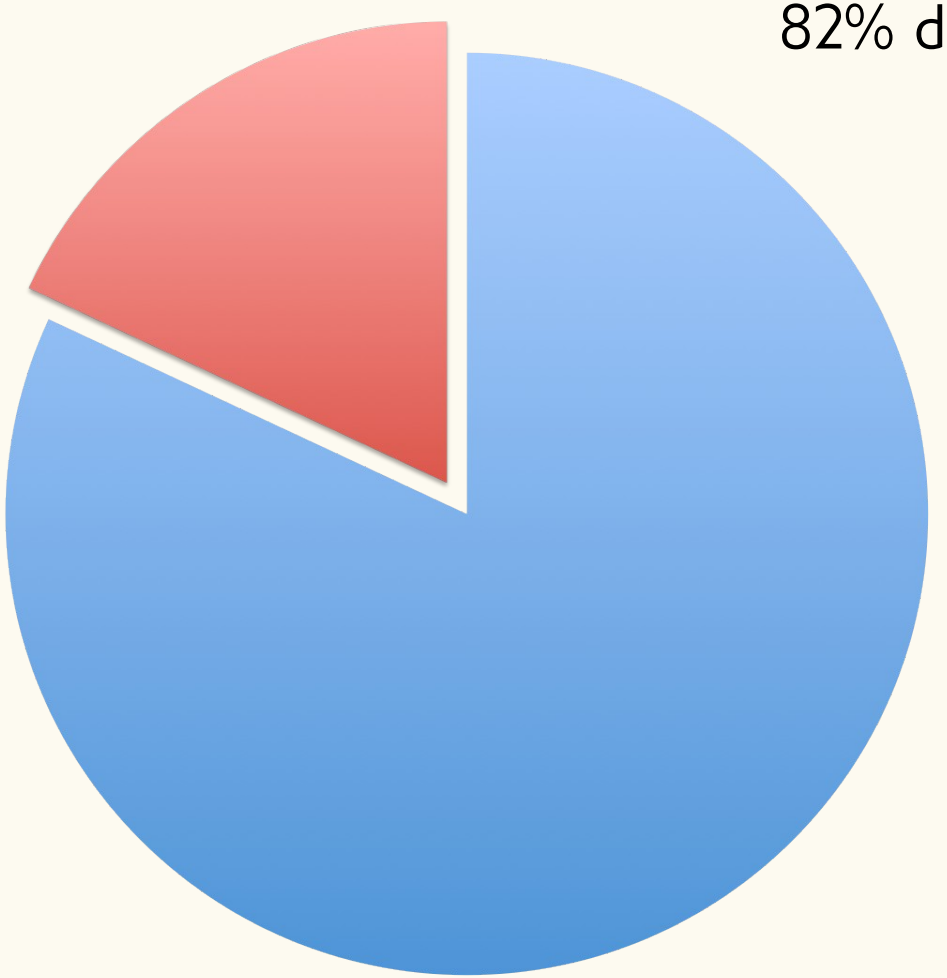
The data – ideally Open Data

all bundled and published together

In audio and music research, few people do this. Why?

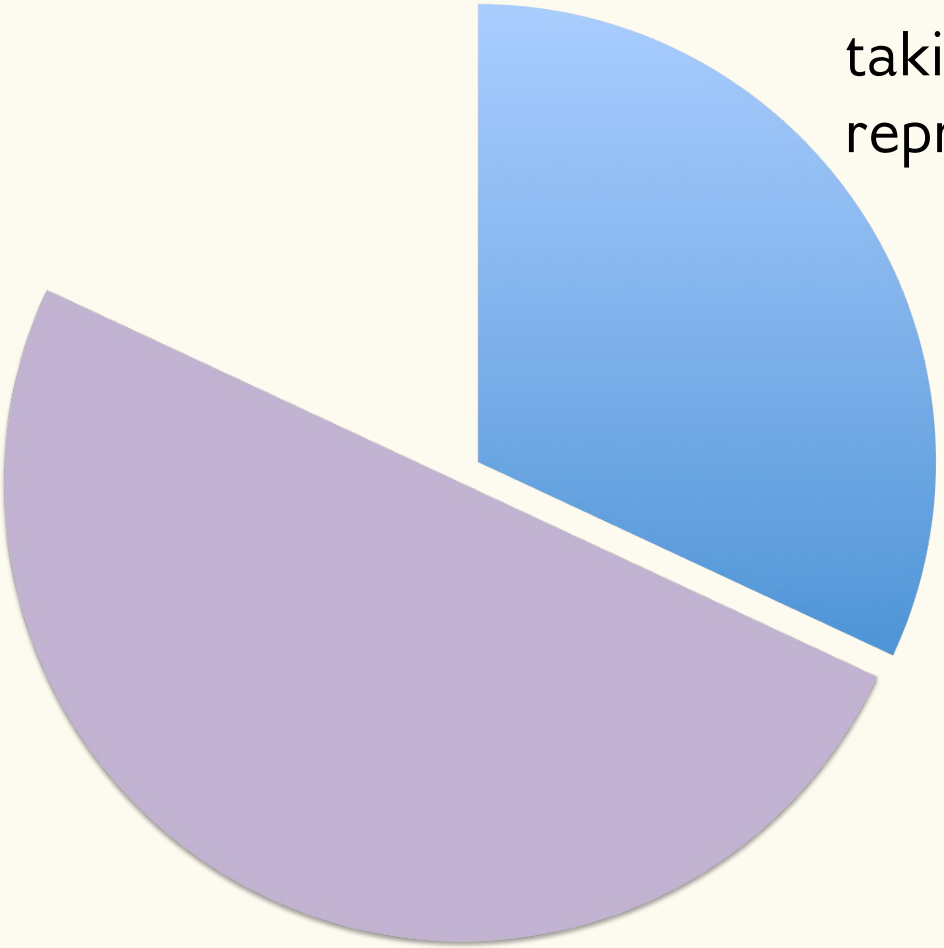
Survey 2010–2011

82% develop code



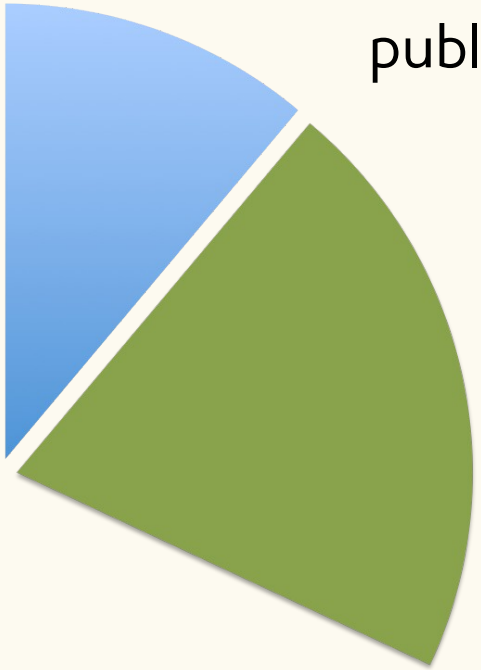
Survey 2010–2011

of whom 39% report taking steps to reproducibility



Survey 2010–2011

of whom 35% report publishing any code



Survey 2010–2011

That's 11% of the whole

It's not just us!

McCullough, 2007

Surveying economics journal **with** a data+code archive policy

9 empirical articles

http://bit.ly/6otJMx

It's not just us!

McCullough, 2007

Surveying economics journal **with** a data+code archive policy

9 empirical articles

7 had empty entries in the journal archive

http://bit.ly/6otJMx

It's not just us!

McCullough, 2007

Surveying economics journal **with** a data+code archive policy

9 empirical articles

- 7 had empty entries in the journal archive
- The other two had code, but it didn't work!
- None of them could be replicated without authors' help

http://bit.ly/6otJMx

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

http://bit.ly/vaU435

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

What can we do to help?

We're taking a **bottom-up approach**:

- aid **incremental improvements** to development practice

by

- identifying **specific barriers** to publication and reuse, that are relatively straightforward to address

So we hope to:

- increase perception among researchers that code is something you can work on together, that can be reused
- prepare the ground for reproducible publication

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*.

Pilot data-management project at C4DM (to be discussed this afternoon) to feed into future work on *data*.

Barrier: Lack of confidence in code

Issue:

Researchers largely self-trained in software development

Our approach:

- Training in software development specifically for researchers
- Relatively small amounts of training can pay off

Autumn School, Software Carpentry

Workshops and tutorials around ISMIR and DAFx this year

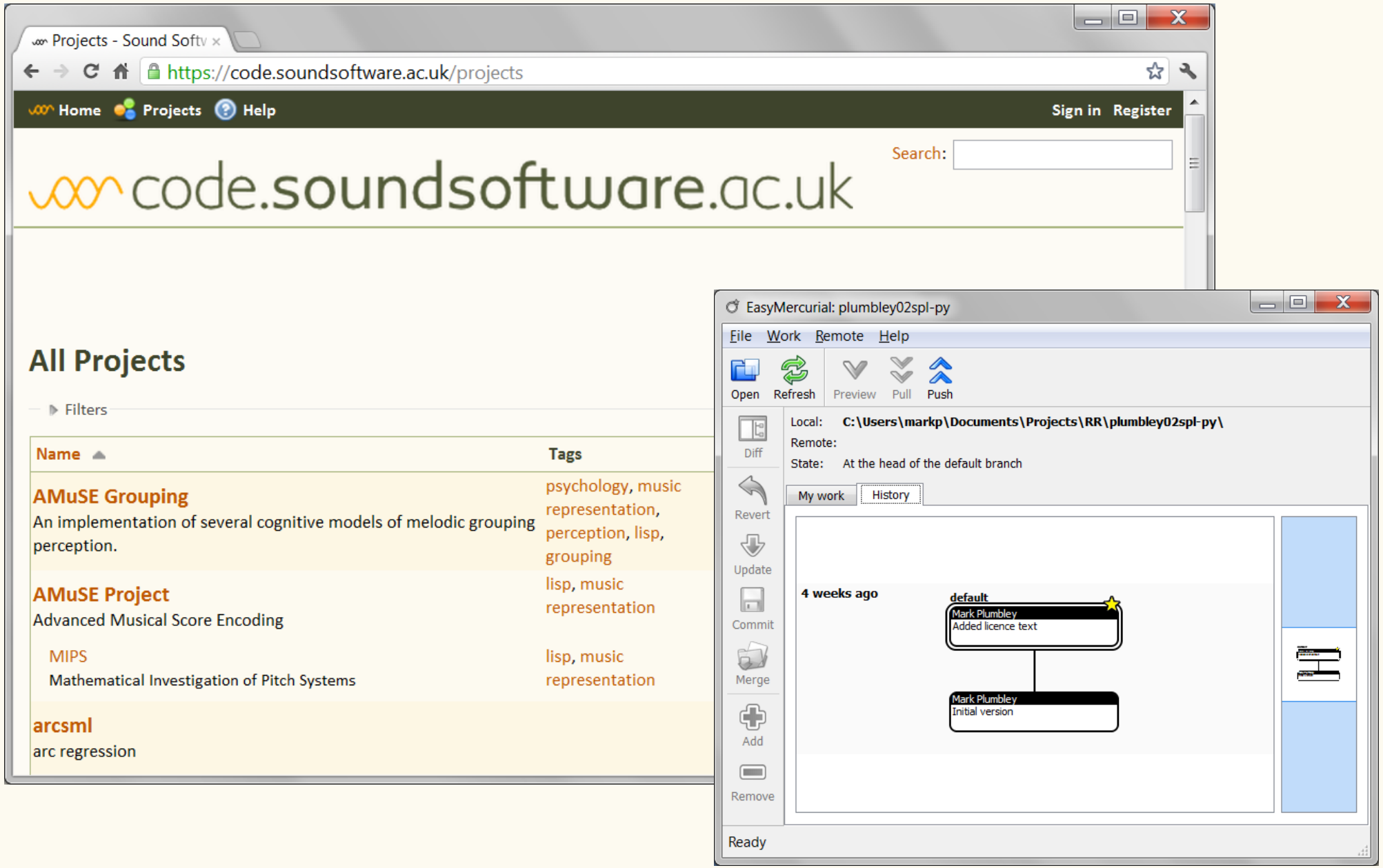
Barrier: Lack of facilities and tools

Issue:

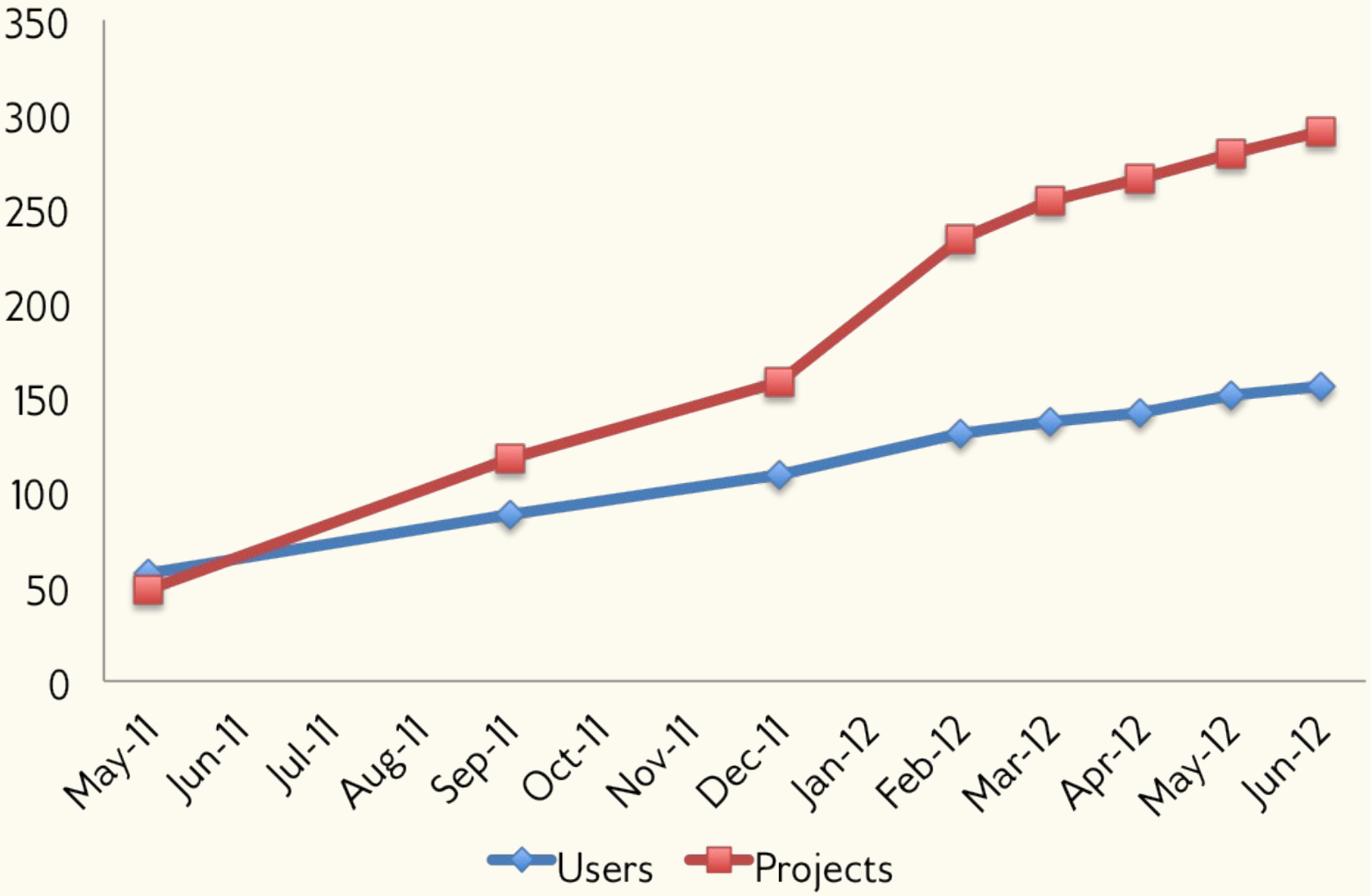
Researchers don't use code hosting / version control

Our approach:

- Code site: http://code.soundsoftware.ac.uk
- Focus on audio and music research
- Public and private projects
- Link publications with code
- Simplified tools, e.g. EasyMercurial, http://easyhg.org


Projects - Sound Softv
https://code.soundsoftware.ac.uk/projects
Home
Projects
Help
Sign in
Register
Search:
code.soundsoftware.ac.uk
All Projects
Filters
Name
Tags
AMuSE Grouping
An implementation of several cognitive models of melodic grouping perception.
psychology, music representation, perception, lisp, grouping
AMuSE Project
Advanced Musical Score Encoding
lisp, music representation
MIPS
Mathematical Investigation of Pitch Systems
lisp, music representation
arcsml
arc regression
EasyMercurial: plumbley02spl-py
File
Work
Remote
Help
Open
Refresh
Preview
Pull
Push
Diff
Revert
Update
Commit
Merge
Add
Remove
Local: C:\Users\markp\Documents\Projects\RR\plumbley02spl-py\
Remote:
State: At the head of the default branch
My work
History
4 weeks ago
default
Mark Plumbley
Added licence text
Mark Plumbley
Initial version
Ready

Projects and users of the code site



Barrier: Lack of incentive

Issue:

Software not well recognised as research output

Our approach:

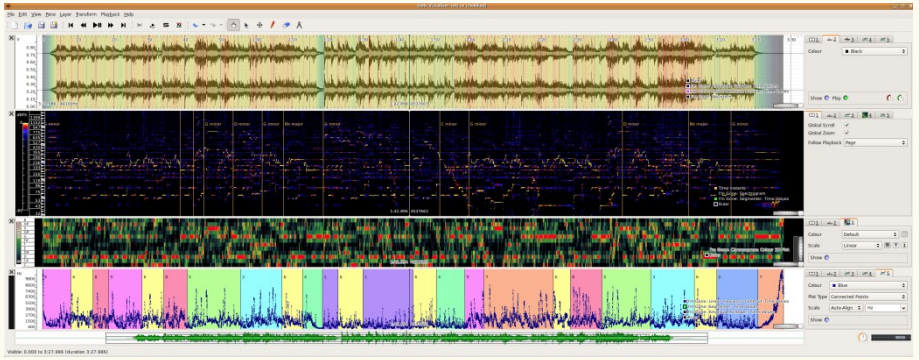
- Link publications to code on the code site
- Increase likelihood of code users discovering your papers
- Ensure users know how to cite your work
- Increase take-up / impact of your research

Barrier: Platform incompatibilities

Issue:

Many different platforms and development tools in use

Some are not available to all possible users (e.g. MATLAB)



Our approach:

- Plugins where possible (e.g. Vamp for SV)
- Take advantage of existing ecosystems

Suggestions for research groups

Aim at easy training targets; researchers may know less about coding than you'd think!

- Program structure, arranging code across files etc

Insist on use of version control

- Use what you have available, or code.soundsoftware.ac.uk

Turn code into plugins or components in modular systems

- Latch onto the existing ecosystems of popular applications

Encourage collaborative development

- Papers often co-authored, why not code?
- Create an environment of confidence about sharing

Links

Our site

- http://soundsoftware.ac.uk

Our code site

- http://code.soundsoftware.ac.uk

These slides, and notes on them

- http://code.soundsoftware.ac.uk/documents/40