Durham University

www.software.ac.uk

Reproducible Research in Archaeology

Alison Clarke, Michelle de Gruchy, Emma Karoune, Nick Syrotiuk

Shared document: https://bit.ly/309U7sU

Find our slides here: https://doi.org/10.5281/zenodo.5564649

Today's workshop

13:00 - 13:10 Introductions and ice breaker

13:10 - 13:40 Presentation

13:40 - 13:55 Case studies

13:55 - 14:00 Comfort break

14:00 - 14:20 Interactive survey and brief discussions

14:20 - 14:45 Breakout discussions

14:45 - 15:00 Q&A and summing up

Q&A on slido

Go to slido.com and enter code #**485809**

Alison Clarke

Durham University

- **Research Software Engineer @ Advanced Research Computing**
  - ARC provide RSE support for researchers
  - Projects in Physics, Education, Psychology, Music
  - Training
- **Fellow of Software Sustainability Institute**
  - Domain-specific paths to software sustainability training: https://bit.ly/3aDAnzI
  - Starting with Archaeology

www.software.ac.uk https://www.software.ac.uk/about/fellows/alison-clarke

Michelle de Gruchy

Durham University

- PDRA on the Climate, Landscape, Settlement and Society (CLaSS) Project
- Responsible for expanding the CLaSS project settlement database
- Conduct spatial analysis and statistics on settlement mainly using the database using tools/buttons in QGIS (ideally this is fully reproducible)

Emma Karoune

- **Archaeobotanist**
  - FAIR Phytoliths project
  - Open reference collections
  - AEA open science skills workshops - Nov 2021
- **Open Researcher**
  - Tools, practices and systems programme
    - Turing Way
    - DECOVID
  - SSI Fellow

Link to Emma's SSI Fellows page

Nick Syrotiuk, Research Data Manager

Durham University

- Based in the Research Support Team in the Library
- Supports researchers with research data management including:
  - interpreting funder requirements
  - writing a data management plan
  - choosing the best storage solution for a project
  - protecting personal data
  - working reproducibly
  - publishing research data in a repository

The Software Sustainability Institute

"A national facility for cultivating better, more sustainable, research software to enable world-class research"

A collaboration between the universities of Edinburgh, Manchester, Oxford and Southampton. Supported by the UK Research Councils through grants EP/H043160/1, EP/N006410/1 and EP/S021779/1, with additional project funding from AHRC, EPSRC, Jisc, NERC and UKRI.

Why are we talking about reproducibility today?

‘Crisis of reproducibility’

The term *Reproducibility crisis* was coined in early 2010’s.

The *Nature* survey (Baker 2016) concerning reproducibility found:

90% of respondents think there is a ‘crisis of reproducibility’

70% of scientists surveyed had tried and failed to reproduce another scientist’s experiments

Is there a reproducibility crisis in Archaeology?

Why?

- Lack of open access
- Lack of open methods
- Lack of open data
- Lack of open analysis (code)

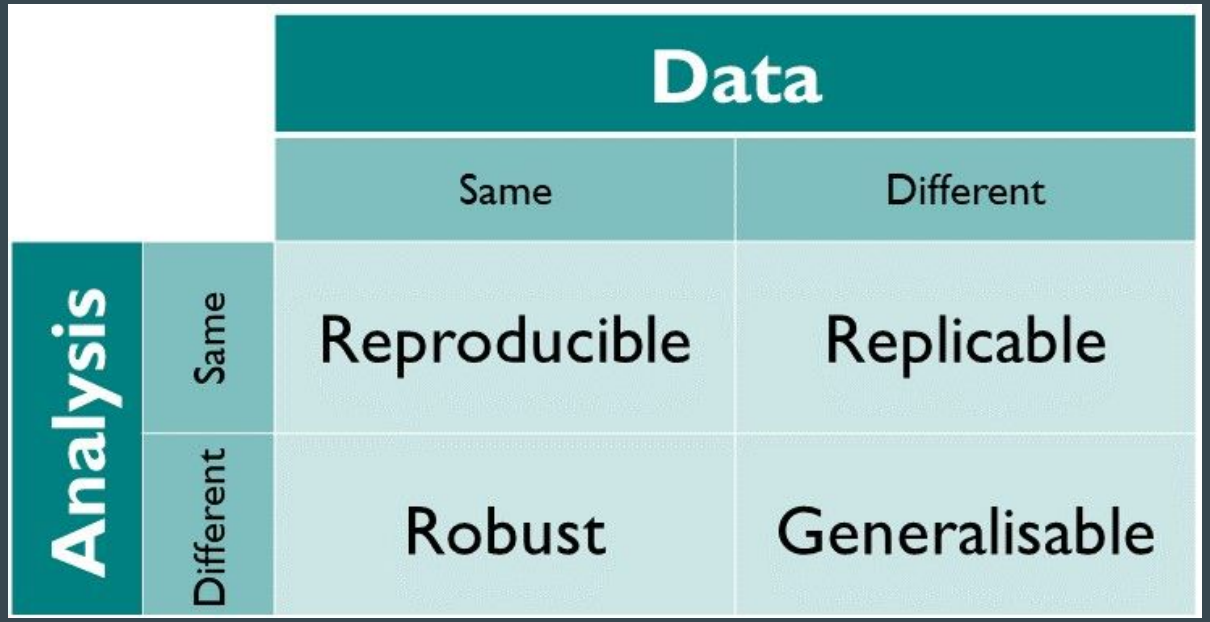
Small steps towards reproducibility

The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: doi.org/10.5281/zenodo.3332807.

What is reproducible research?

Definition from *The Turing Way*:

‘Work that can be independently recreated from the same data and the same code that the original team used’.



Link to The Turing Way Book

Types of reproducibility

Empirical reproducibility: When detailed information is provided about non-computational empirical scientific experiments and observations. In practice, this is enabled by making the data and details of how it was collected freely available.

Computational reproducibility: When detailed information is provided about code, software, hardware and implementation details.

How can archaeological research be reproducible?

Doing archaeology is a destructive process

It is HOW you do your research that can make it reproducible

Why should archaeological research be reproducible?

Overarching reasons

- Research validation
- Sustainability
- Equity, diversity and inclusion

Immediate benefits for you:

- Tracking project history
- Collaborate and review
- Avoid misinformation
- Write papers efficiently
- Get credit fairly
- Ensure continuity

Sensitivities concerning reproducible ways of working

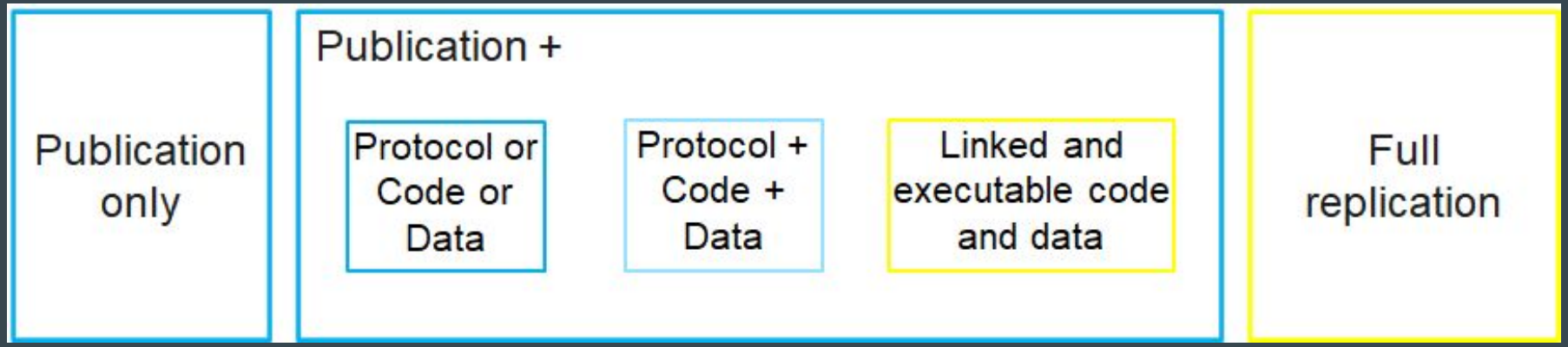
Is my data too sensitive to share?

I'm not sure I want others to see all my work?

Do I have permission to share my work openly?

I don't have time or can't see how to share my work

Reproducible Research Spectrum

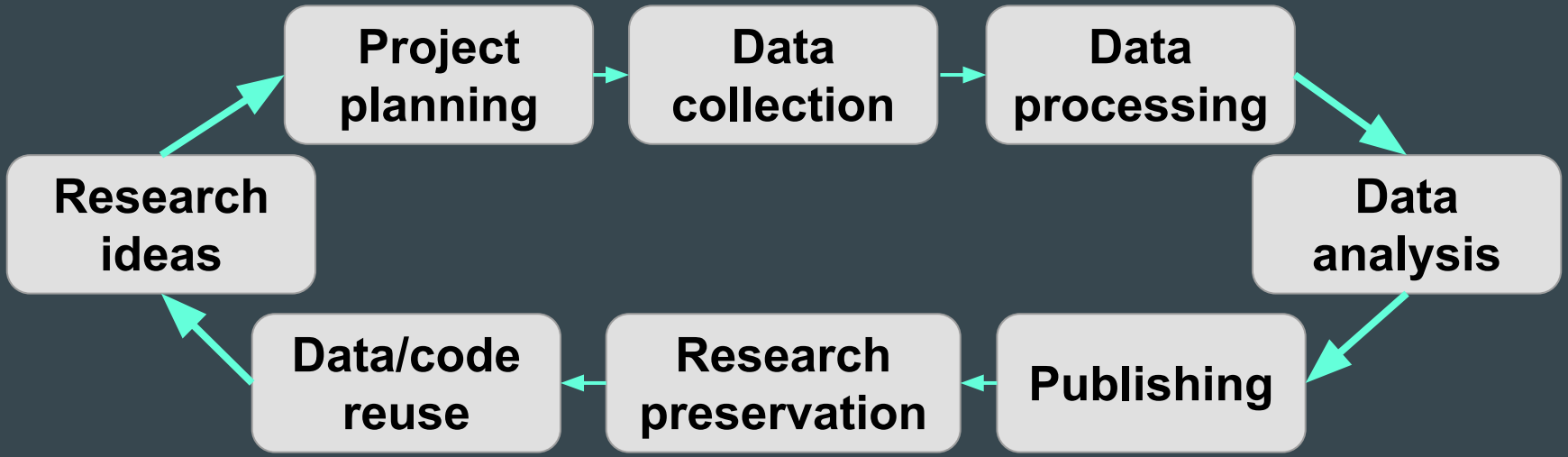


Not reproducible ⟷ **Gold standard**

Adapted from Peng 2011

https://www.science.org/doi/abs/10.1126/science.1213847

Reproducible workflows in archaeology - What?



How do you do reproducible research?

Building a workflow to include:

Transparency

- Clear documentation of your workflow so methods, data, code - a working research compendium.

Collaboration

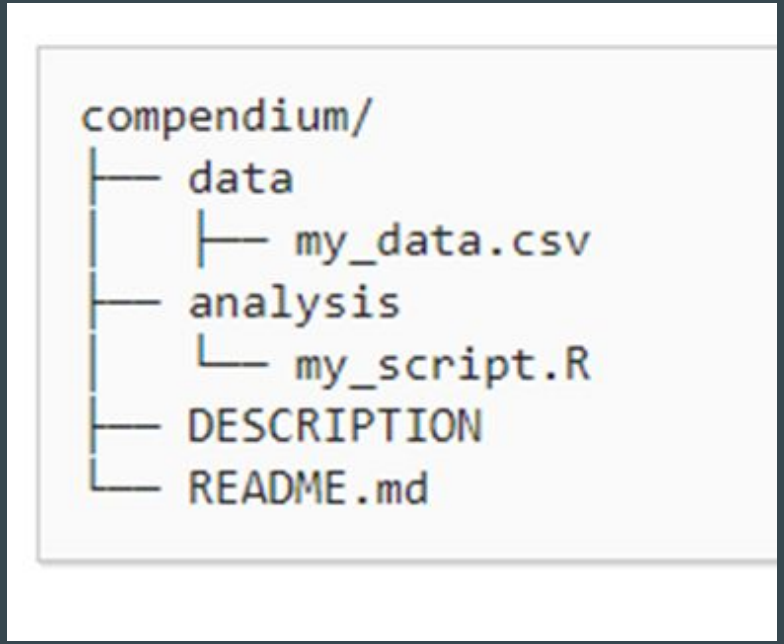
- Open access to work and working with others needs version control.

Research compendia

- Set of files that gives all the information about your reproducible project.

• Data
• Analysis
• Methods

README file



Basic research compendium structure

Version control - records changes made in a file over time

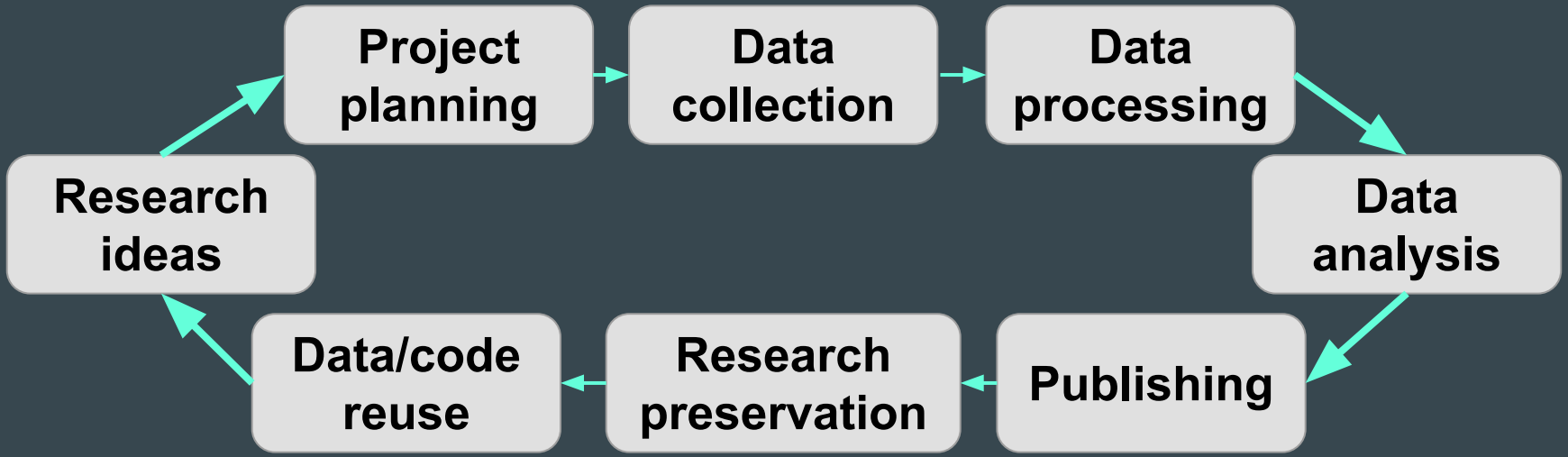
Benefits

- Go back to previous versions.
- Store history of changes.
- Collaborate with other.
- Recording contributions of team - fair credit for work done.

How you can do this:

- Simple file versioning - add v1.0, 1.1, 1.2, etc, to filename.
- Use simple tools - Google drive, Dropbox.
- Advanced tools - Github, subversion.

Reproducible workflows in archaeology - What?



Methods

What you need to know:

- All stages need to be described fully
- Data collection on site
- Sampling during excavation
- Sampling for lab work
- Samples used in lab

What you need to do:

- Open repository - Simple document - README
- Reference for a method - make sure it is the actual method you use.
- Protocols.io

All these need to be full protocols so others could replicate the method or know that their data could be combined with yours.

Data

"As open as possible, as closed as necessary"

What you need to know:

- How open? CARE
- How to deposit? FAIR
- Where? - open repositories - free!
- What data? - Raw data
- Metadata- Research compendia

What you need to do:

- Choose open repository - DOI, licenses, space, feature add-ons.

C - Collective benefit
A - Authority to control
R - Responsibility
E - Ethics

+

F - Findable
A - Accessible
I - Interoperable
R - Reusable

Publishing data: Three types of repository

Subject-specific

- E.g., Archaeology Data Service.

  Slightly expensive but your funder might pay.

Find a subject-specific repository: **re3data.org**

Multi-disciplinary

- Zenodo
- Figshare
- Open Science Framework
- Mendeley
- Dryad

Institutional

- Durham research data repository

- Managed by Research Support Team in the Library.
- Single deposits up to 50 Gb.
- Available to staff and students at no charge.

Analysis

What you need to know:

- Open source
- R/Python
- Version control - Git/Github

What you need to do:

- Document steps and parameters used in GUI tools
- Write scripts to do your analyses
- Save scripts to version control
- Create live papers with Jupyter or RMarkdown

What is Github?

GitHub

- **Git is an open source, version control tool**
  - Stores history of what changed, when, who changed it and why
- **Github is an online interface that uses Git (+ extras!)**

Why it is useful for reproducible projects

- Storage of all project documentation, data, code and you can create a webpage
- Web interface for version control
- Working collaboratively - kanban project pages, issues, pull requests.

Computational environments

What you need to know:

- Results can change between different versions of software and operating systems

(Optionally):

- Docker
- mybinder.org

What you need to do:

- Document software versions
- Document operating system

(Optionally):

- Create a docker container to reproduce your environment

Any questions?

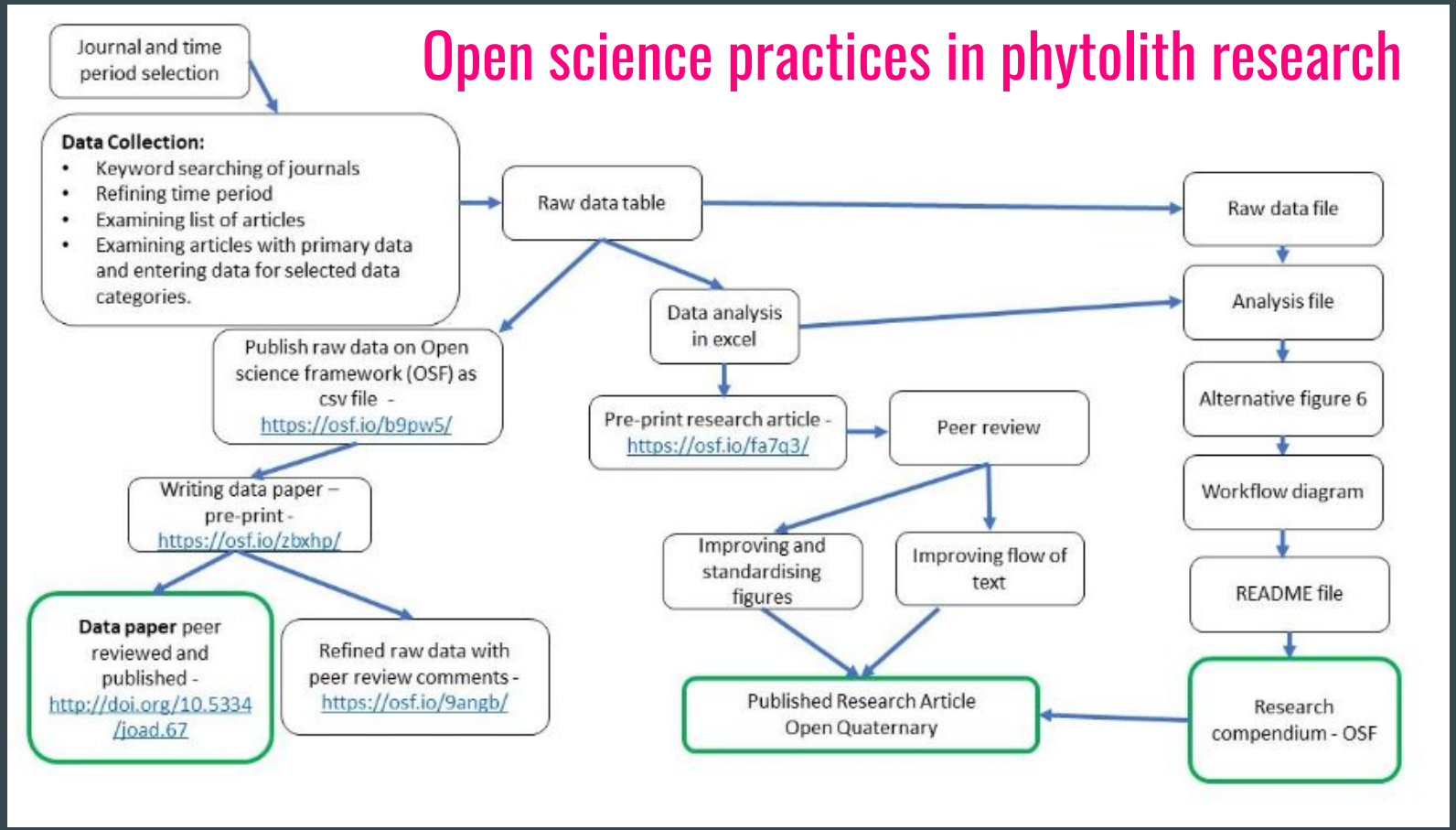
slido

Audience Q&A Session

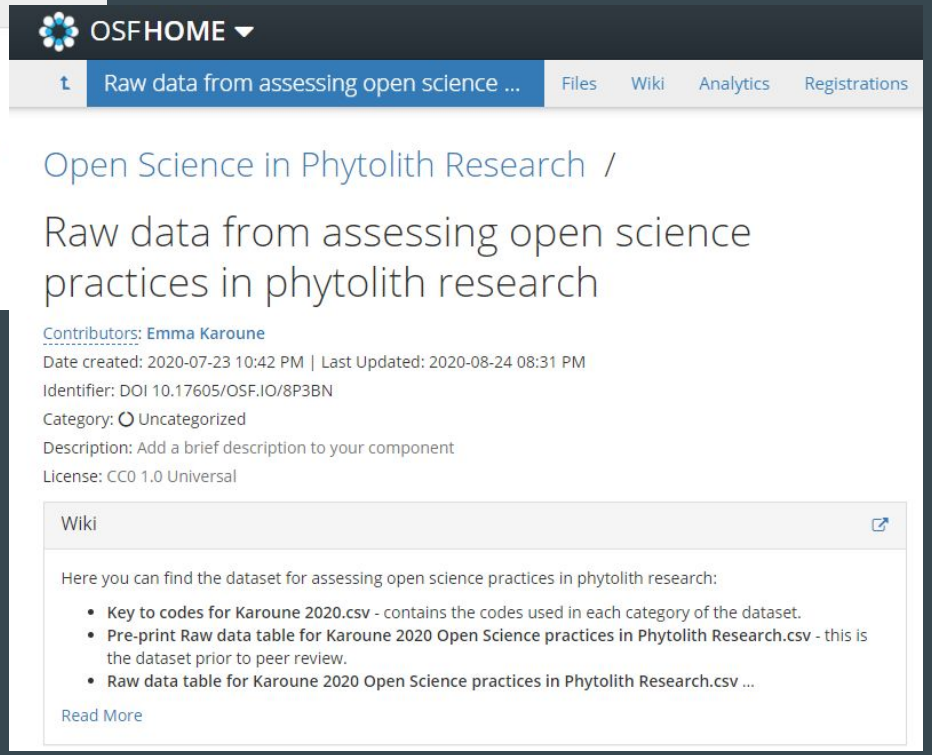
ⓘ Start presenting to display the audience questions on this slide.

Case studies

Case studies - Reproducible workflow with publications



Data paper + Open data



Research article + Research compendia

OPENQUATERNARY

Examples of Open source projects

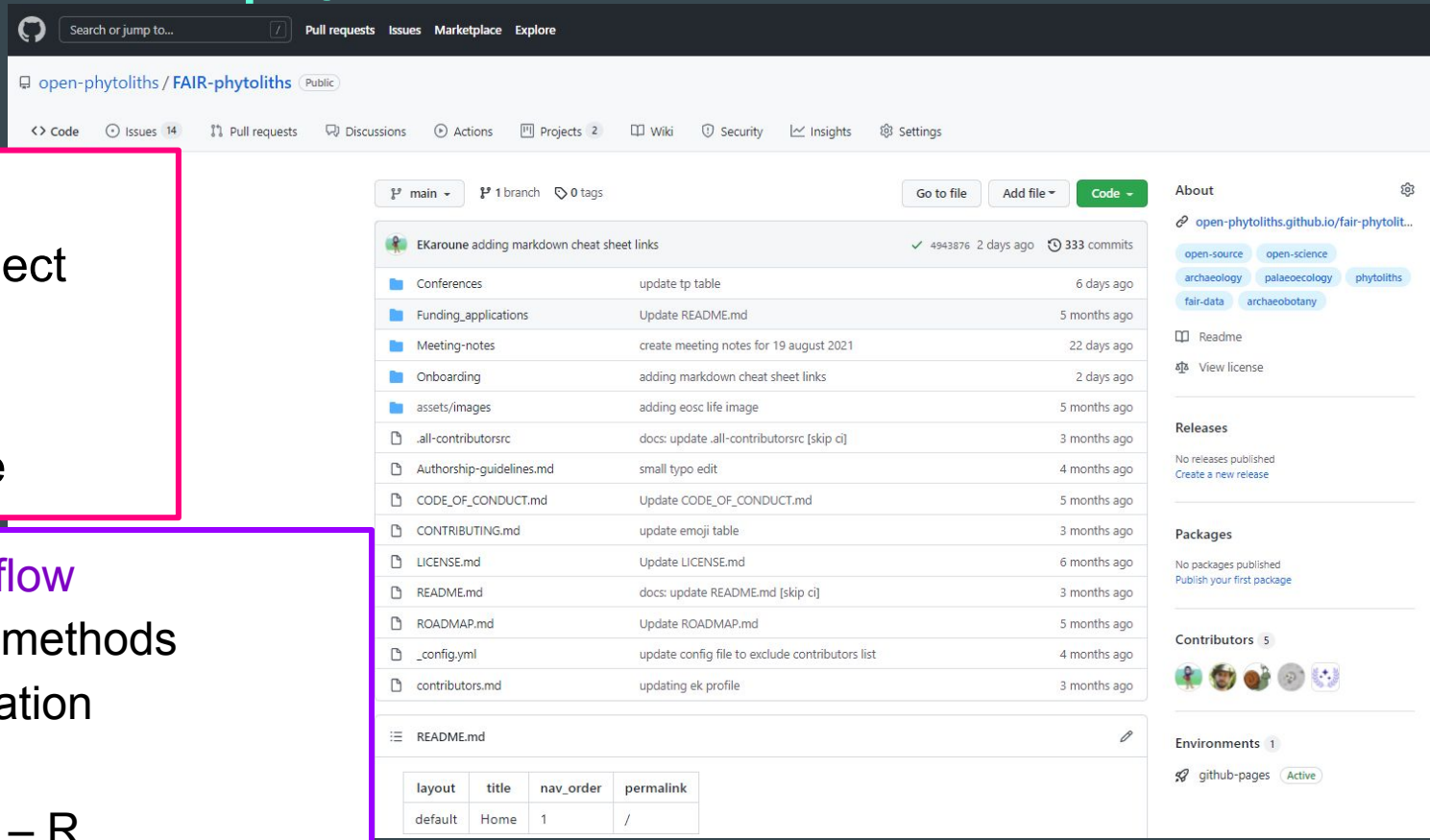
FAIR Phytoliths

- Open source project
- Github/Zenodo
- License

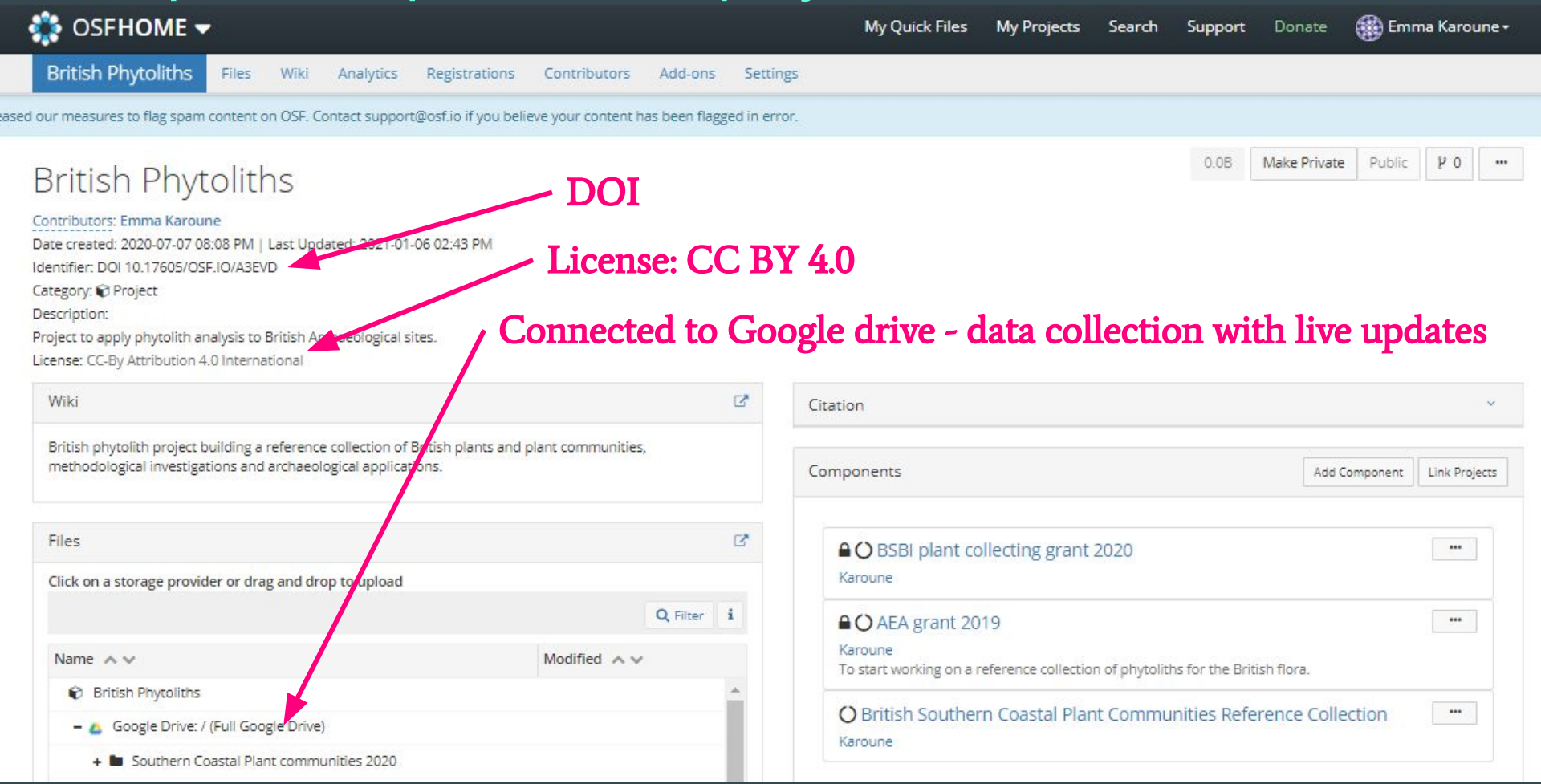
Contributing guide

Reproducible workflow

- Readme file – methods
- Full documentation
- Open data
- Open analysis – R
- Binder - computational env



Examples of Open source projects



Case Study: Michelle de Gruchy & Reproducible GIS

Main Challenges

1. Reproducibility in a button pressing programme e.g. QGIS
2. Avoiding digital clutter
3. Creating an open and accessible database
4. Full replicability so others can genuinely repeat the work

QGIS Demos

Demo is in **QGISDemos.mov**

Tool is at: https://bit.ly/3brcWtK

Break

Slido questions about reproducible research

Slido link:

https://slido.com

Slido code:

#485809

slido

What is your career stage?

ⓘ Start presenting to display the poll results on this slide.

slido

What is your subdiscipline?

ⓘ Start presenting to display the poll results on this slide.

slido

What software do you use for analysing your data?

ⓘ Start presenting to display the poll results on this slide.

slido

Where do you think you are on your reproducibility journey? (1*=Complete beginner; 5*=Fully reproducible)

ⓘ Start presenting to display the poll results on this slide.

slido

Do you think your group would be supportive of moving towards a more reproducible approach?

ⓘ Start presenting to display the poll results on this slide.

slido

What do you find most daunting about reproducibility?

ⓘ Start presenting to display the poll results on this slide.

slido

What could your first (or next) step be?

ⓘ Start presenting to display the poll results on this slide.

Q & A with speakers - Breakout rooms

Please choose a breakout room:

- Room 1: Research data - Nick
- Room 2: Methods - Emma
- Room 3: Analysis - Alison

Shared document: https://bit.ly/309U7sU

What next?

Association *for* Environmental Archaeology

Software Sustainability Institute

Open Science Skills Workshop

1 day workshop

- 19th (European time zone) or
- 20th November (Americas time zone)

On all aspects of open science and related to archaeology

Find out more here: https://envarch.net/news

Thank you!