Status of reproducibility and open science in *hep-lat* in 2021

Ed Bennett

UKLFT Annual Meeting, 2022-05-27

Data: doi:10.5281/zenodo.6584001

Slides: https://edbennett.github.io/uklft-talk-20220527 •

Swansea University
Prifysgol Abertawe

SAC

SUPERCOMPUTING WALES
UWCHGYFRIFIADURA CYMRU

UNDEB EWROPEAIDD
EUROPEAN UNION
Llywodraeth Cymru
Welsh Government
Cronfa Datblygu
Rhanbarthol Ewrop
European Regional
Development Fund

UKRI
Science and
Technology
Facilities Council

Outline

- Definitions
- Motivation
- Survey of hep-lat in 2021:
  - Survey scope
  - Reproducibility and openness of:
    - Configuration production
    - Observable measurement
    - Analysis and post-processing
- Case studies
- Conclusions and next steps

Definitions

Reproducibility

Same data + same analysis → Same results

- Related concepts:
  - Replicability: New data + same analysis → same results
  - Robustness: Same data + new analysis → same results

Open science

The movement to make all research accessible to all levels of society.

Including, but not limited to:

- Publications
- Physical samples
- Data
- Software

FAIR

Research data (and software) should be:

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

Orthogonal to data being open

Motivation

Why open science?

- The ideal scientific process
- Public funding ⇒ Public results
- Our funders say so

"Data resulting from publicly funded research should be made publicly available…
unless there are specific reasons (e.g. legislation, ethical, privacy and security)
why this should not happen" —STFC Scientific Data Policy

Open science accelerates progress

The war over supercooled water (DOI:10.1063/PT.6.1.20180822a)

- Discrepancy discovered in 2011
- Student hired to revisit computations in 2012
- Code requested in 2013
- Code promised "available on request" in 2016
- Code eventually provided after involvement of Nature editors
- Discrepancy caused by a poor choice of initialisation function
- Findings published in 2017

Was that the best use of 6 years of arguments?

Why automated reproducibility?

- Communication with words is imprecise
- Papers have limited space
- Human error is inevitable
- Computers are pretty good at doing the same thing every time

Survey of *hep-lat* in 2021

Survey scope

- Every hep-lat arXiv submission from 2021
  - Including cross-lists
  - Skim-read plus keyword searches
- Series of questions: yes/no, categorisation, and free text
  - Answers based solely on text of paper
- Data and analysis code are available on Zenodo

What computations does an LFT paper do?

A very reductive view:

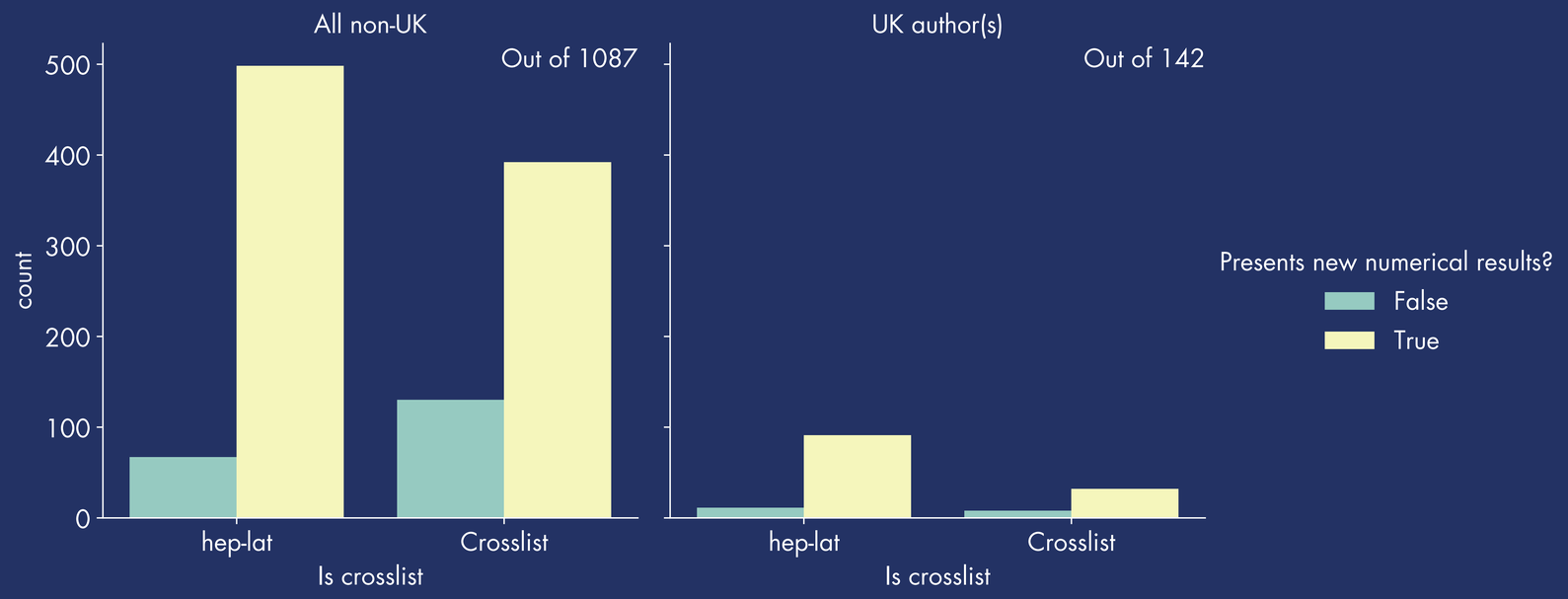
1. Generate field configurations
2. Measures observables on configurations
3. Analyses, plots, tabulates measured observables

Less focus on emerging techniques, e.g.

- tensor networks
- quantum simulation

High level numbers

Out of 1,229 arXiv submissions in 2021:



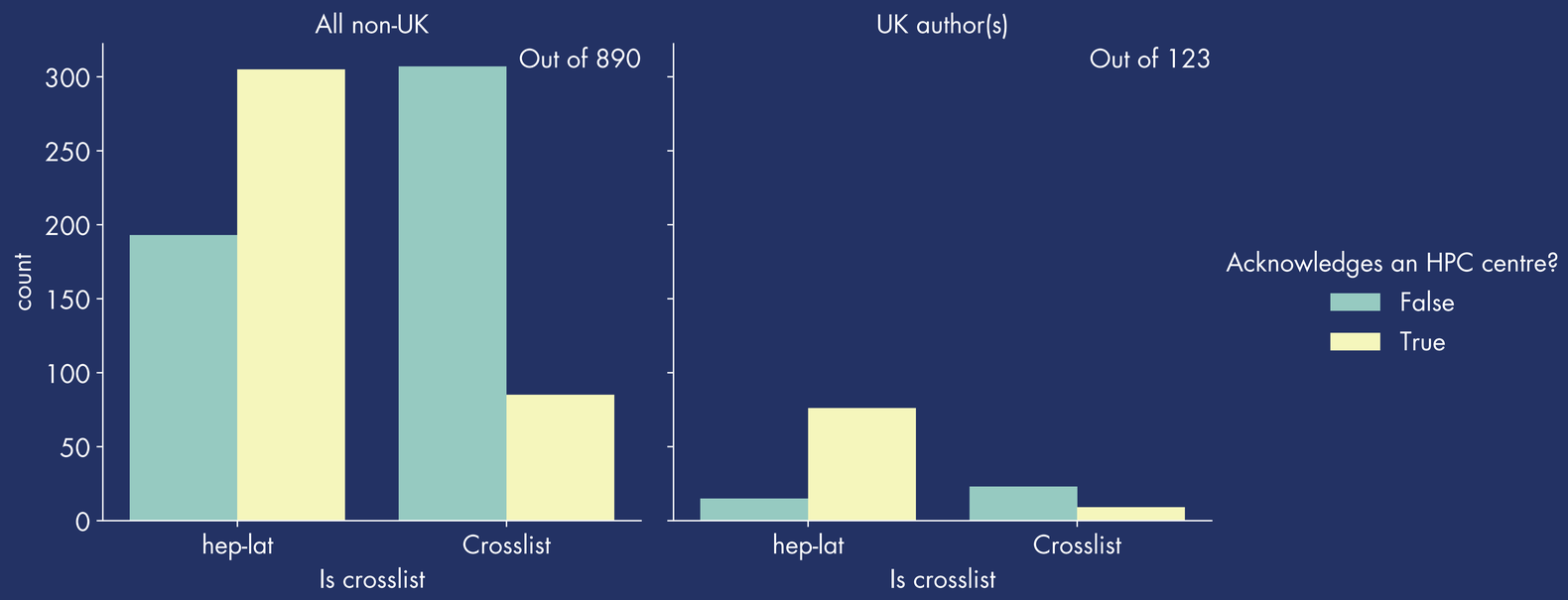
- Use of preprints is already decades ahead of many disciplines!

Why cite software?

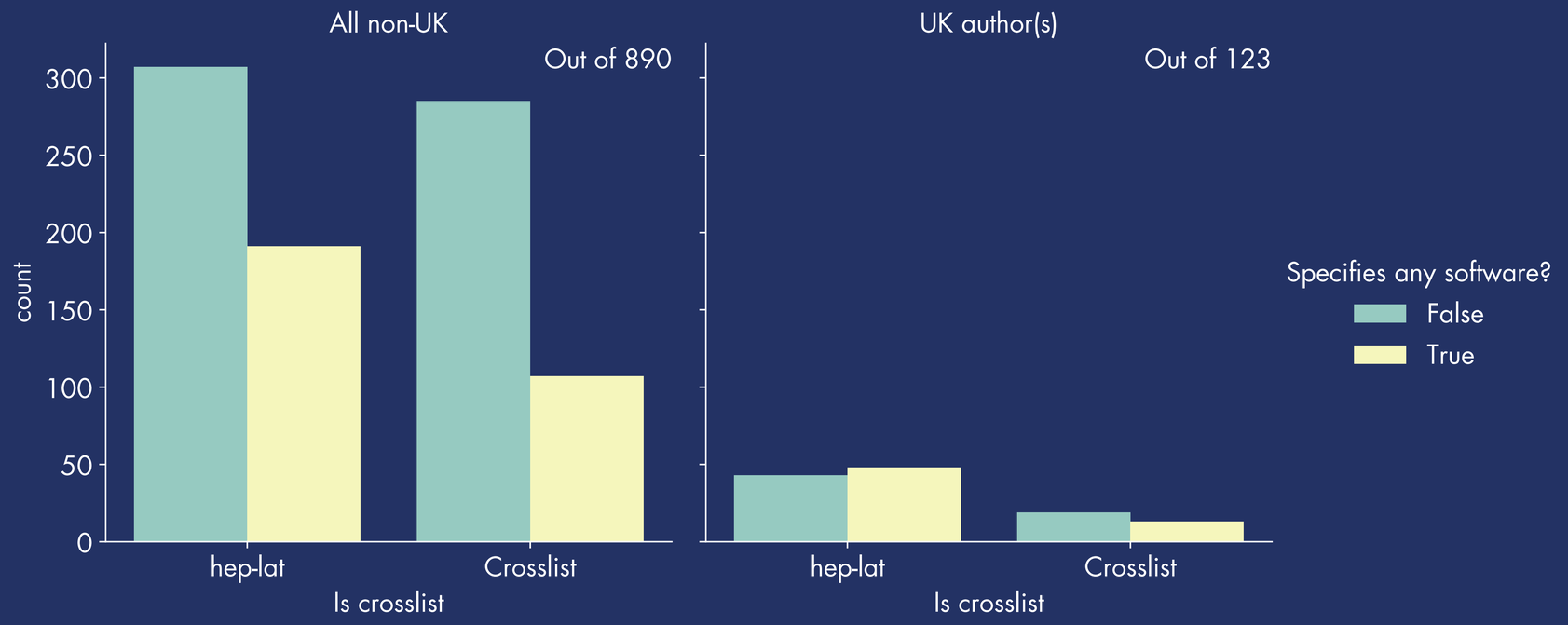
Citing software:

- Gives credit to those who built it
  - Avoids paper-centric metrics
  - Justifies funding maintenance
- More precisely specifies what was done
  - Implementations vary in subtle details
  - Referring to an algorithm is not sufficient

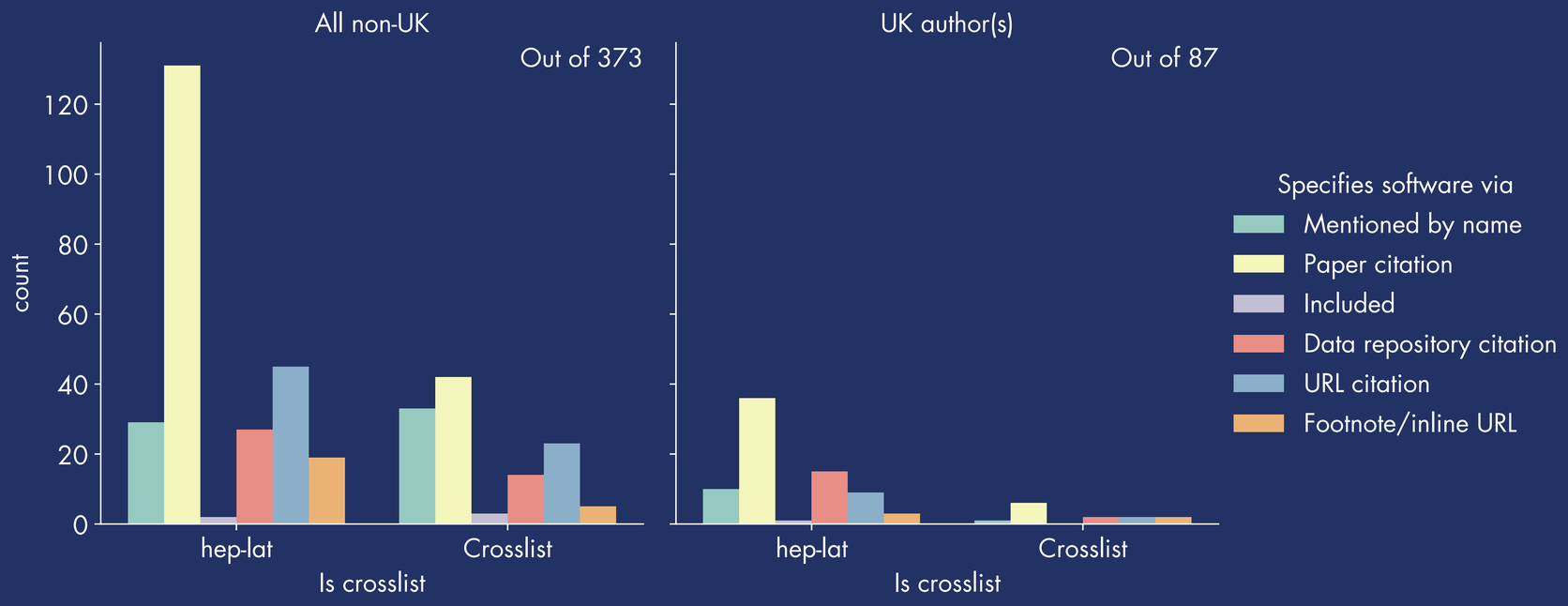
Setting the scene: Acknowledging HPC



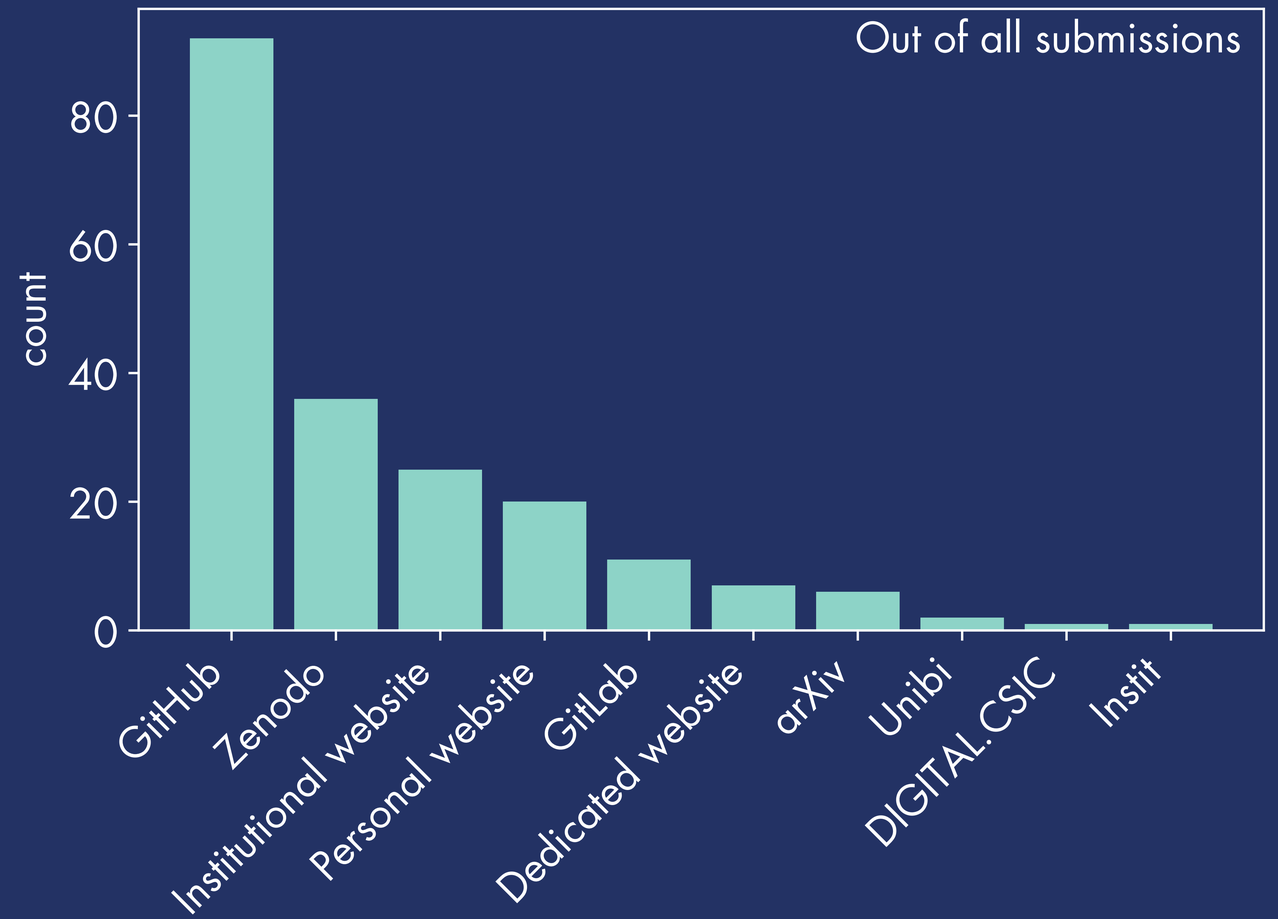
How many submissions specify *any* software?



How is software acknowledged?



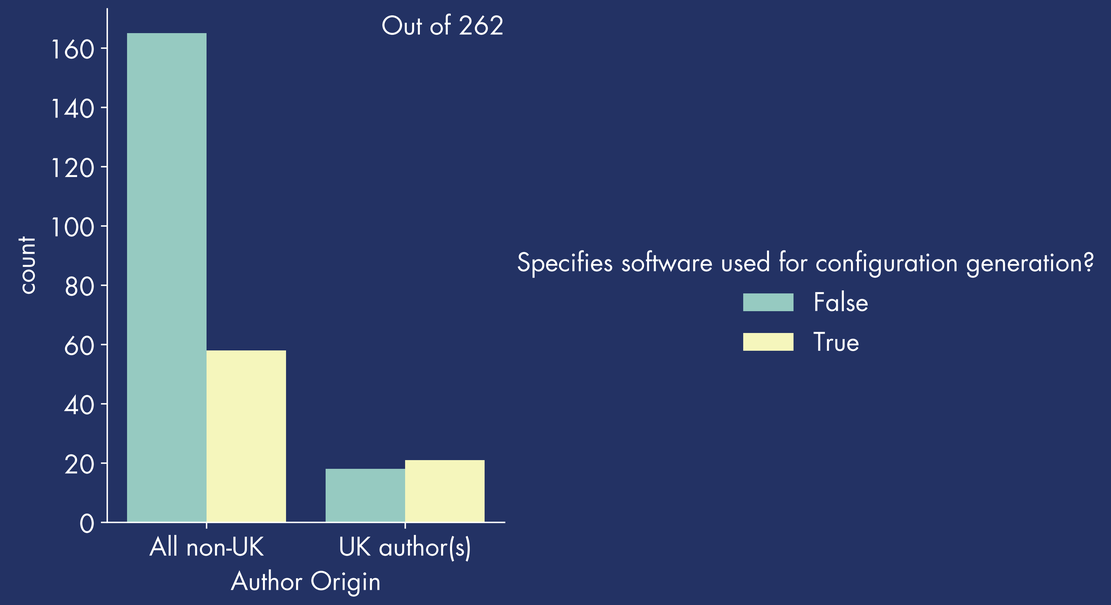
Where does software live?



Generating field configurations

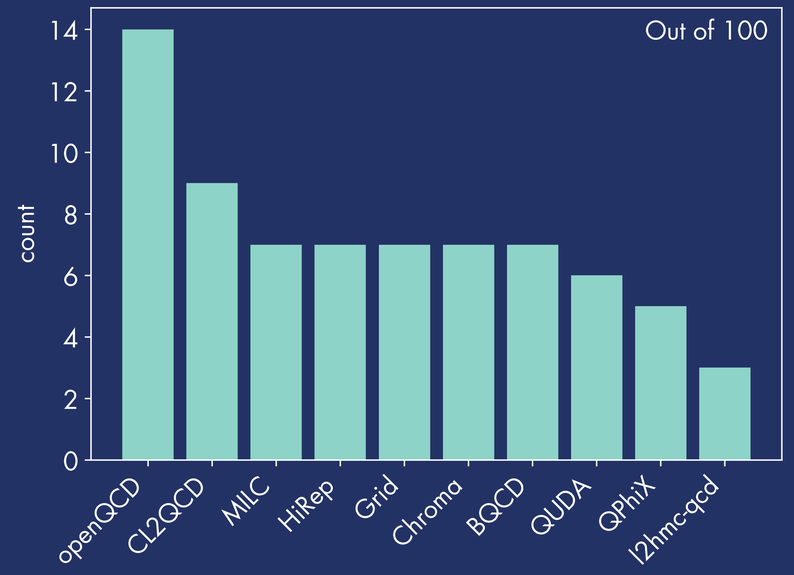
- Usually extremely expensive
  - Hard to test automated workflows end-to-end
  - Hard for others to reproduce (wait for Moore's law?)
  - Open sharing of configurations is good
    - Needs infrastructure (more later)
- Reproducibility efforts include:
  - Seedable RNG, RNG checkpoints
  - Include run parameters in output, configurations files
  - Include code version/commit ID within output
- Around 44% of publications do this

Do authors specify how configurations are generated?



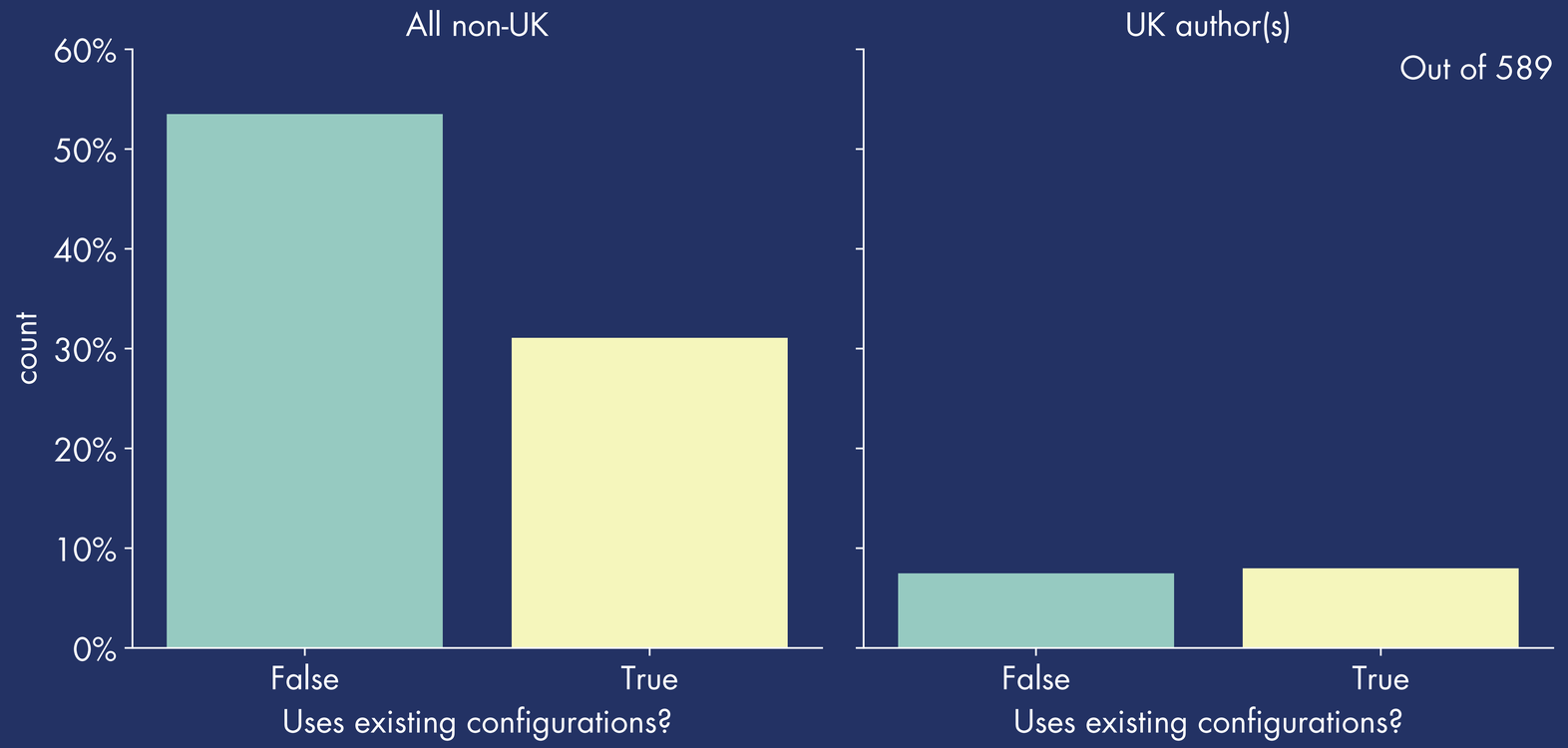
The UK is significantly better than average here.

What software is used to generate configurations?

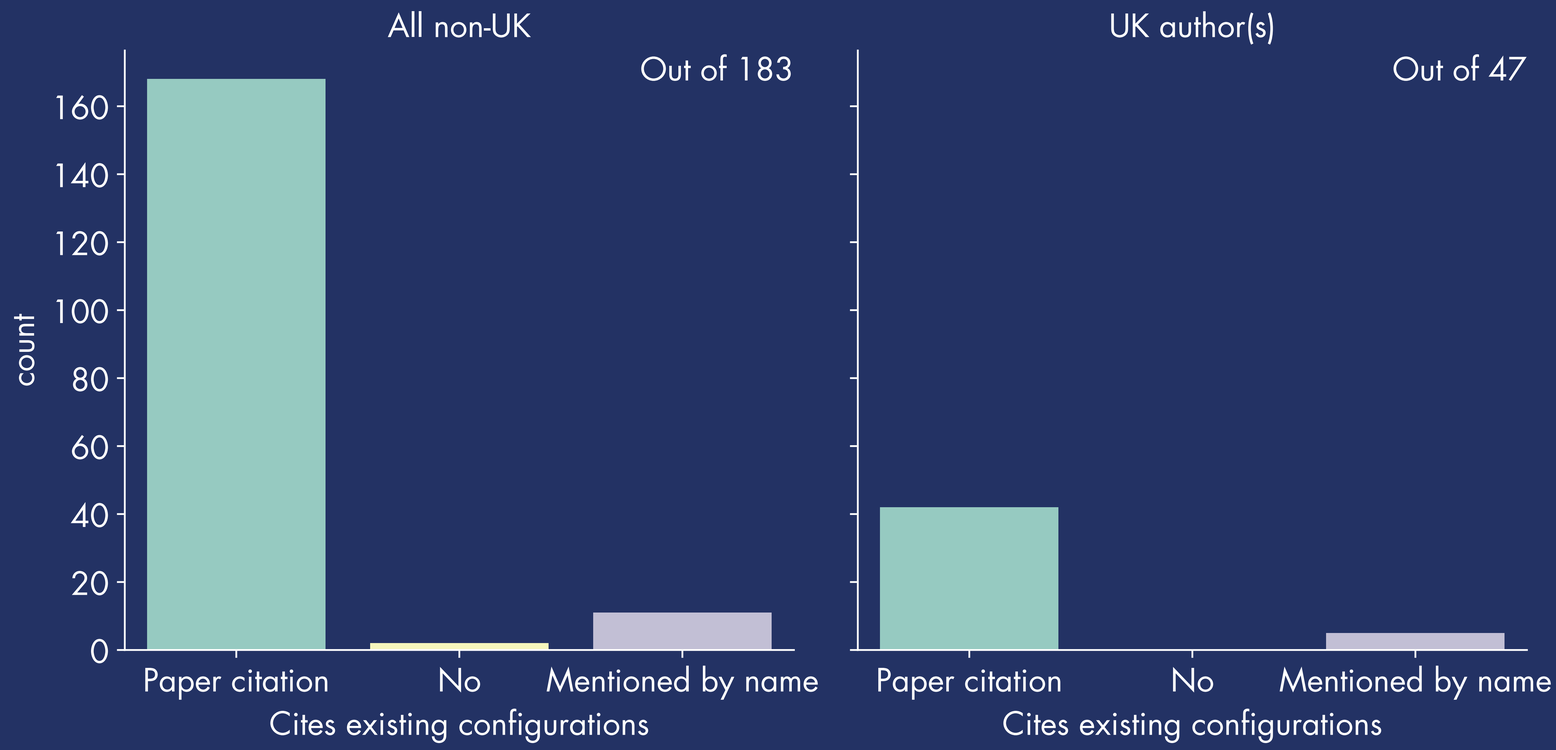


- 11 indicate unreleased modifications
- More only name toolkits (e.g. Grid, Chroma)

What about work that doesn't generate configurations?



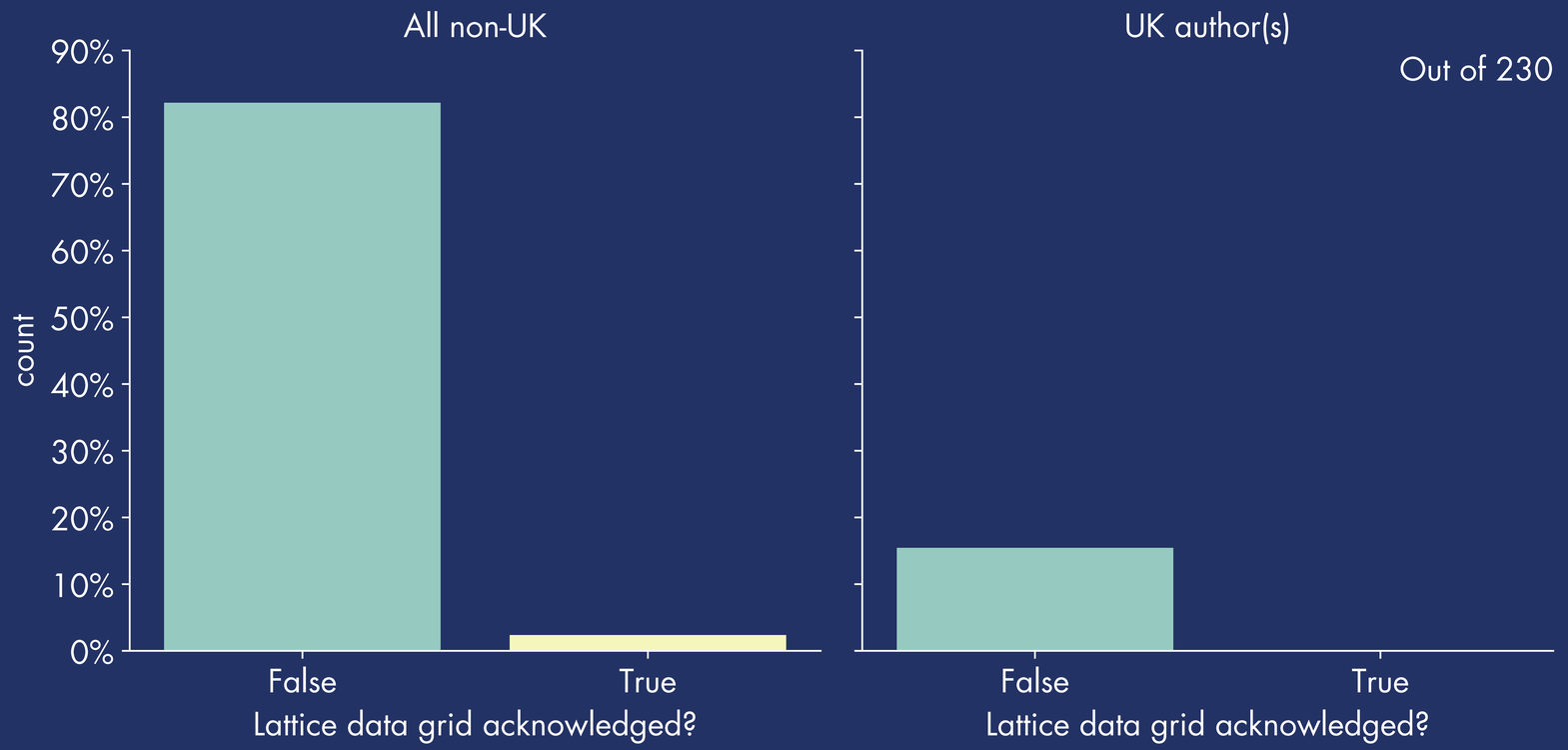
How are existing configurations acknowledged?



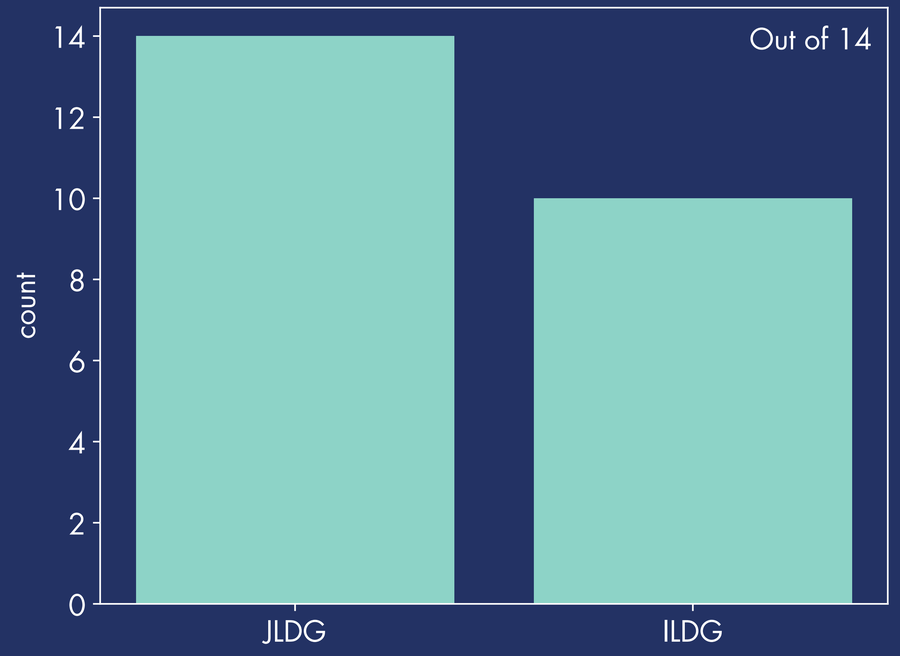
Lattice Data Grids

- International Lattice Data Grid
  - Defines protocols and standards
  - Local deployments in US, UK, Europe, Japan, Australia
- FAIR before FAIR
- Early-ish example of open science

How many papers acknowledge an LDG?



Which LDGs are acknowledged?

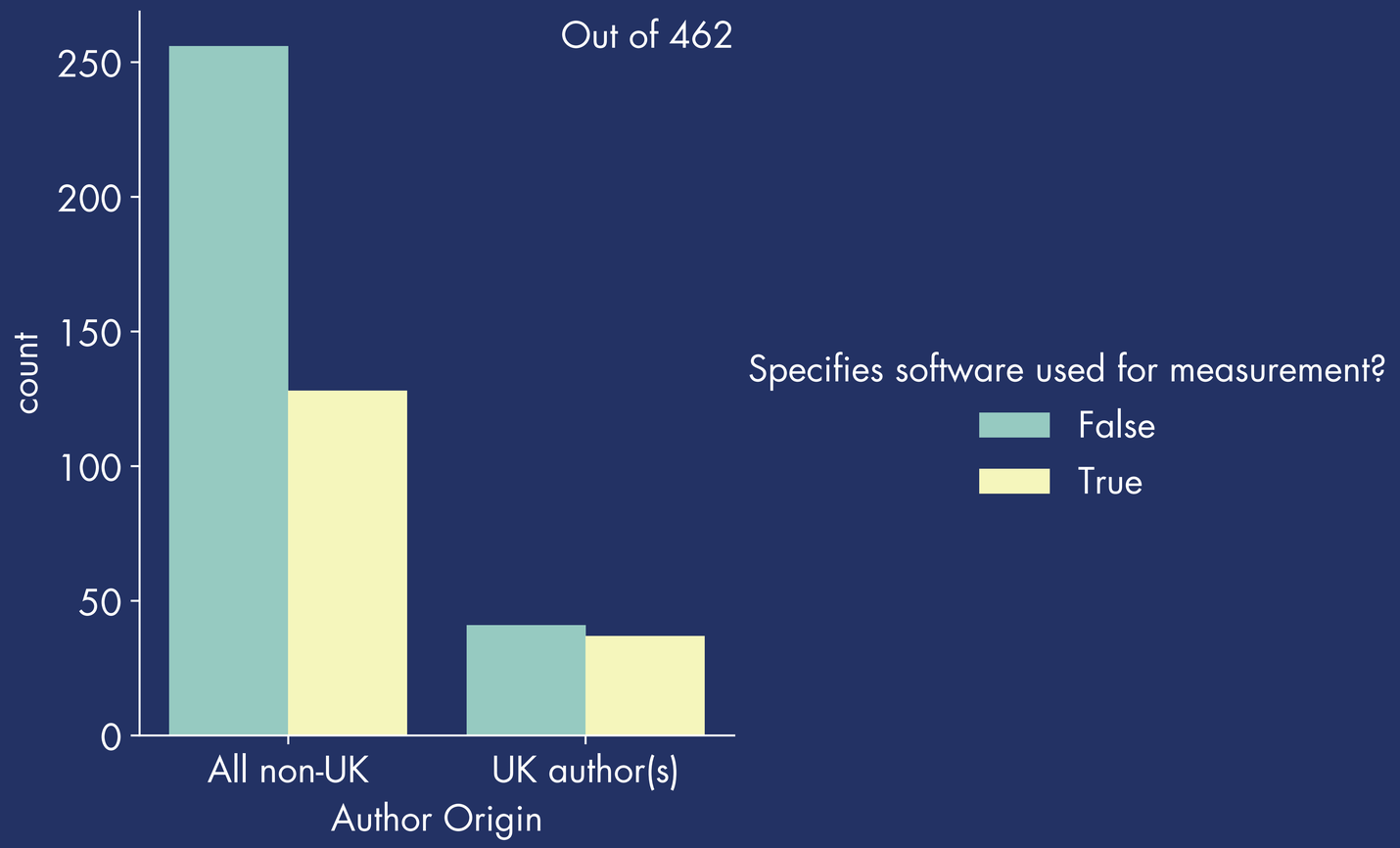


- Japan has the most active(ly cited) LDG
- Either the others aren't used, or aren't cited

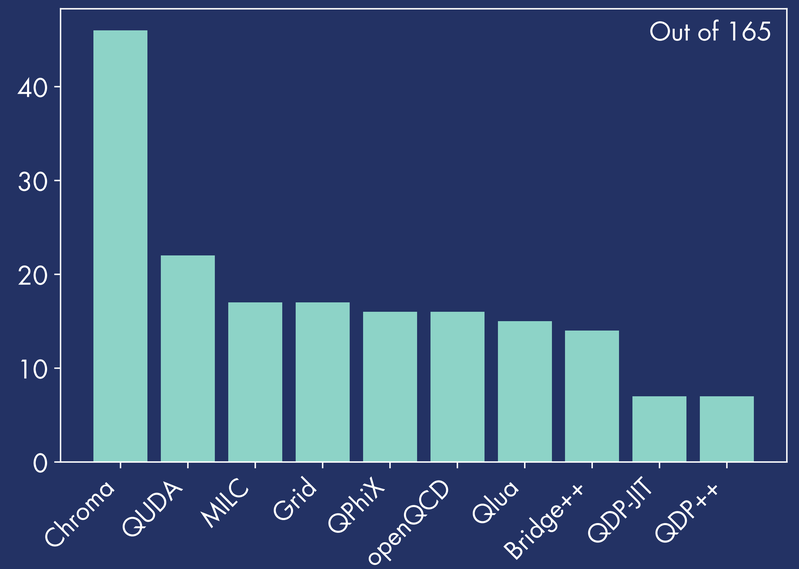
Ongoing work on ILDG

- Perceived issues with ILDG:
  - DOIs, citability
  - Grid certificates
  - Rigidity of metadata
- ILDG committees recently resumed activity
  - Significant German government funding
  - Dedicated staff to address these problems

Performing measurements



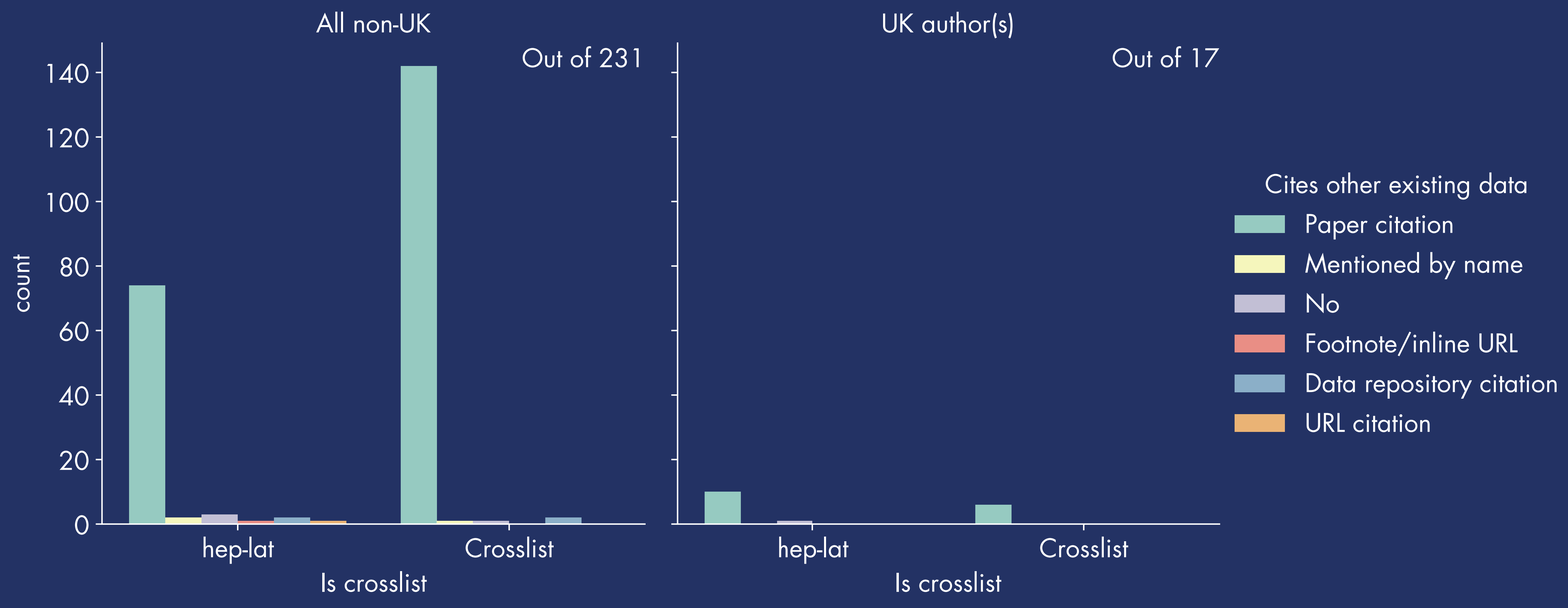
What measurement codes are in use?



- 27 indicate unreleased modifications
- More only name toolkits (e.g. Grid, Chroma)

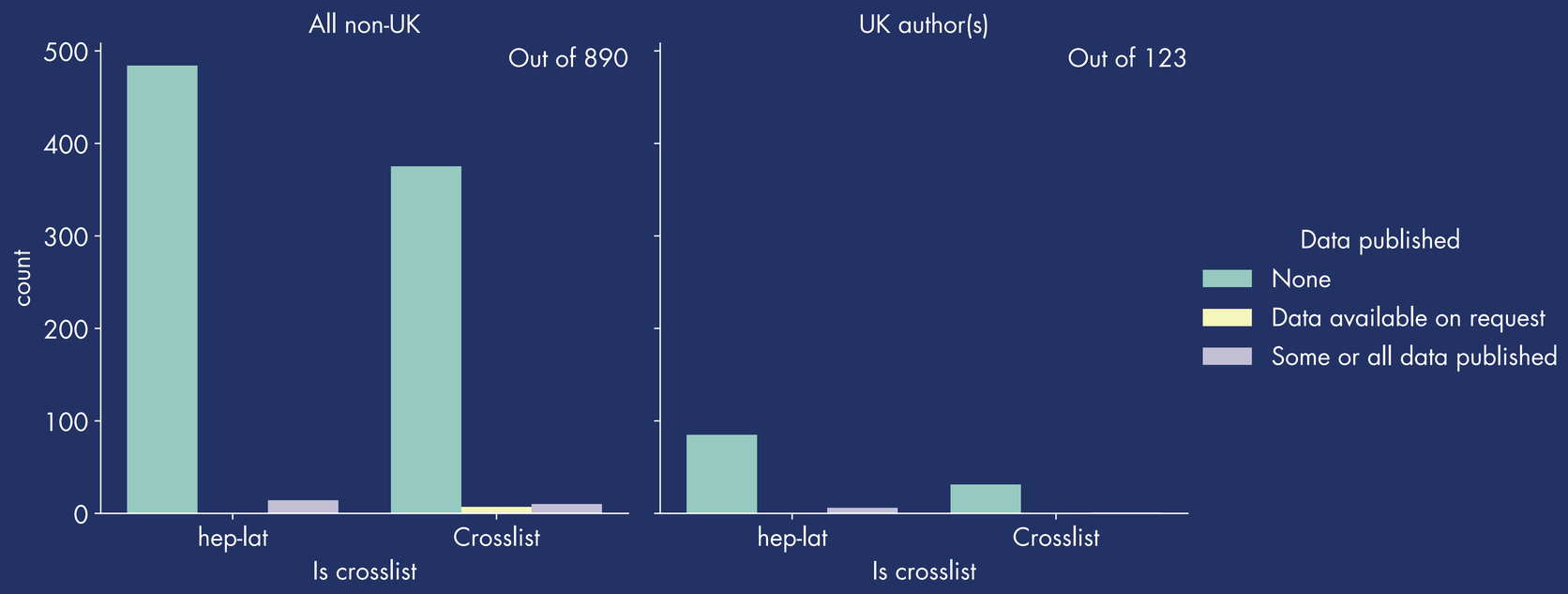
Use of open data

(Excluding field configurations)

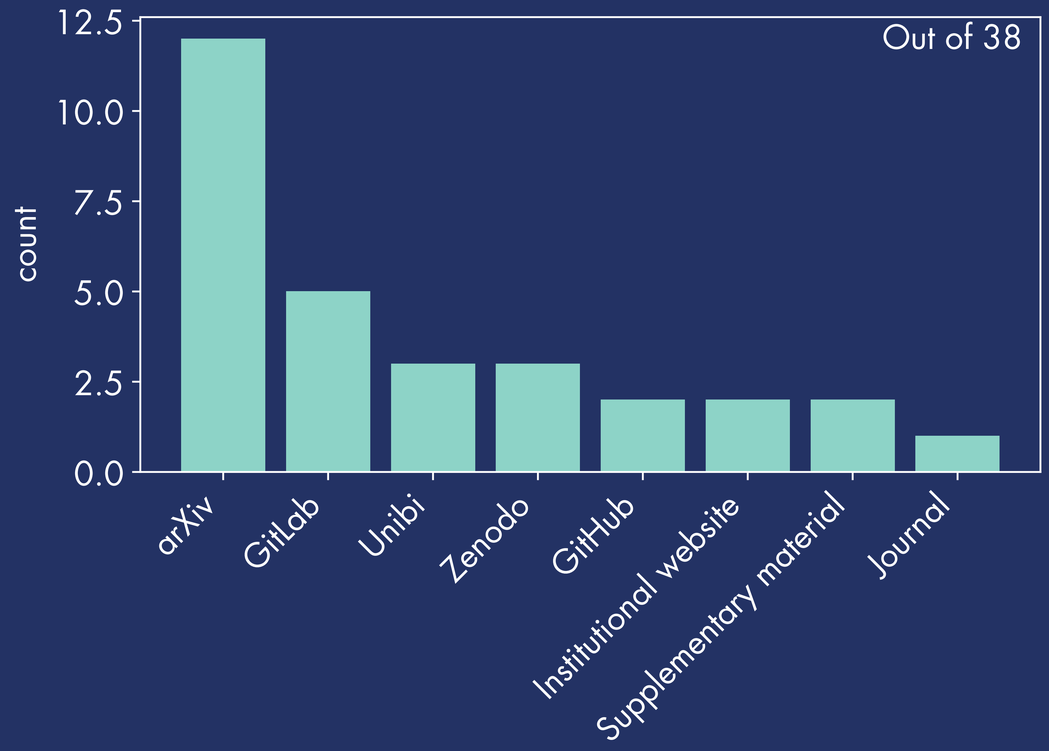


- Acknowledgements to individuals
- Not FAIR

Do authors publish data?



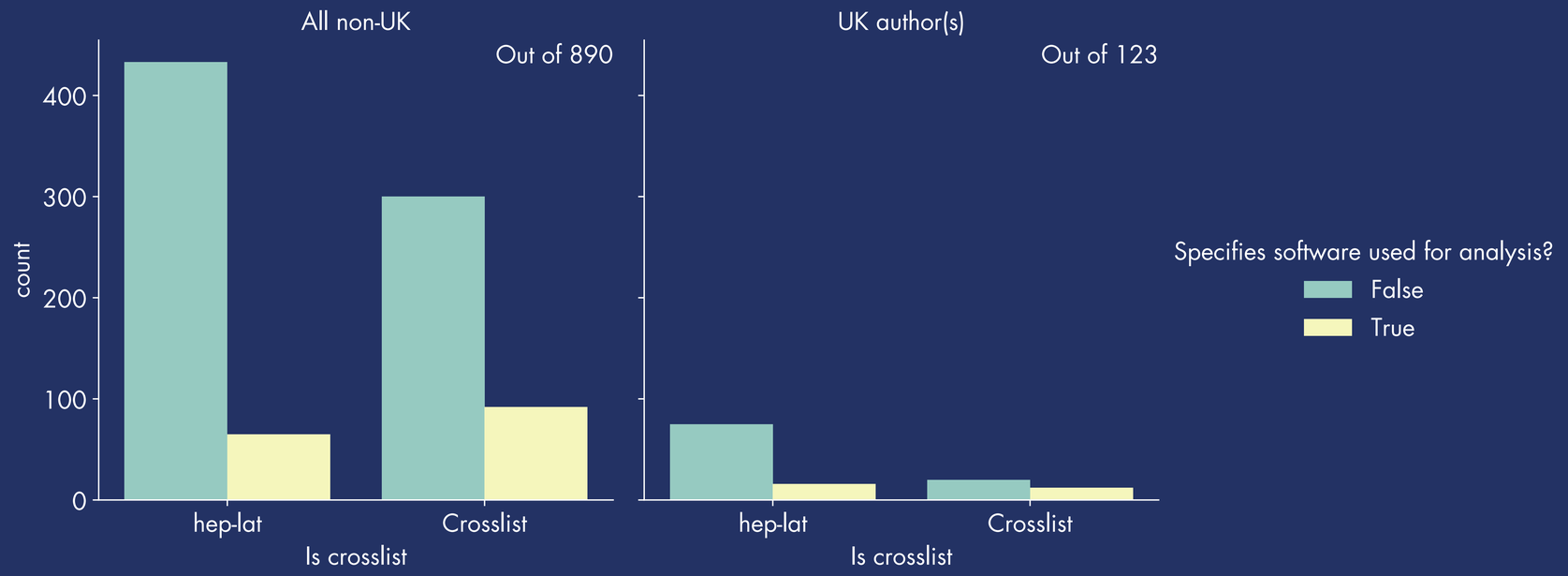
Where are data published?



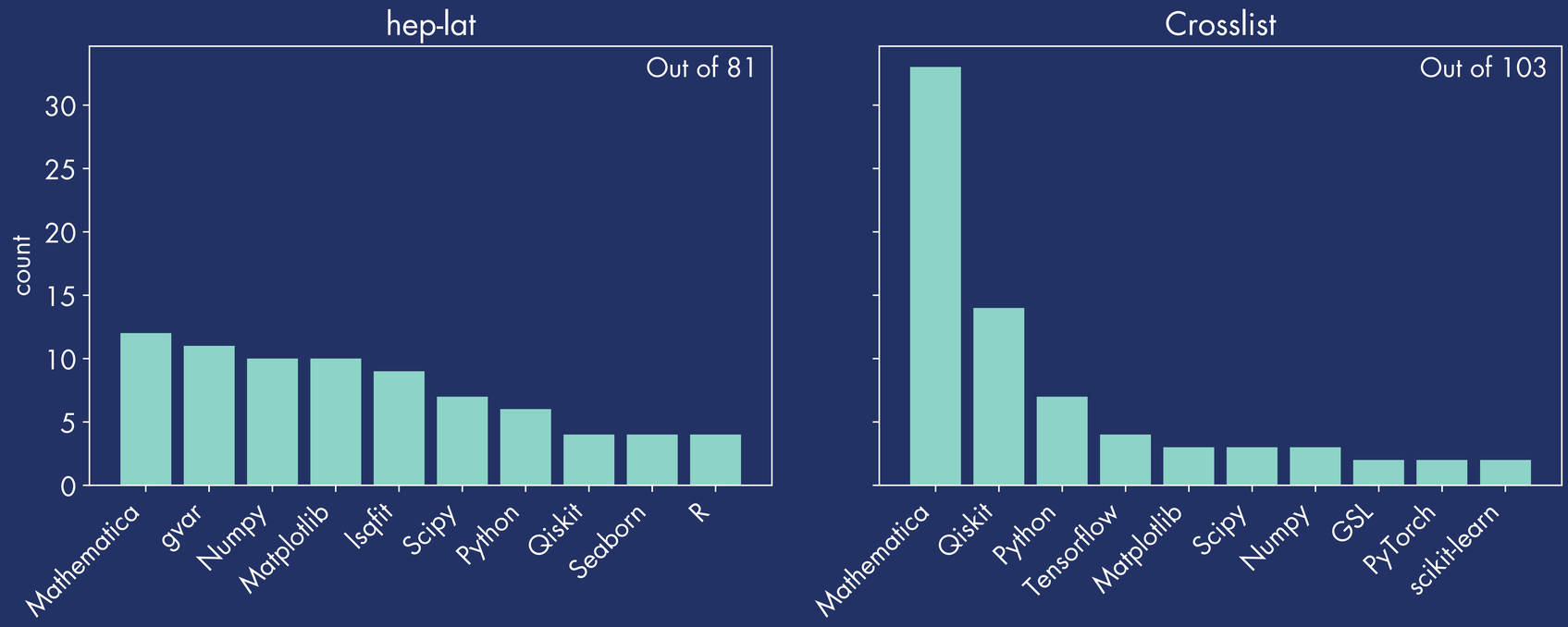
Data analysis

- Experimental research:
  - Does not generate configurations
  - Does not perform computationally reproducible "measurements"
  - Still has a substantial reproducibility effort
  - ⇒ Data analysis of measurement results is the key reproducibility question

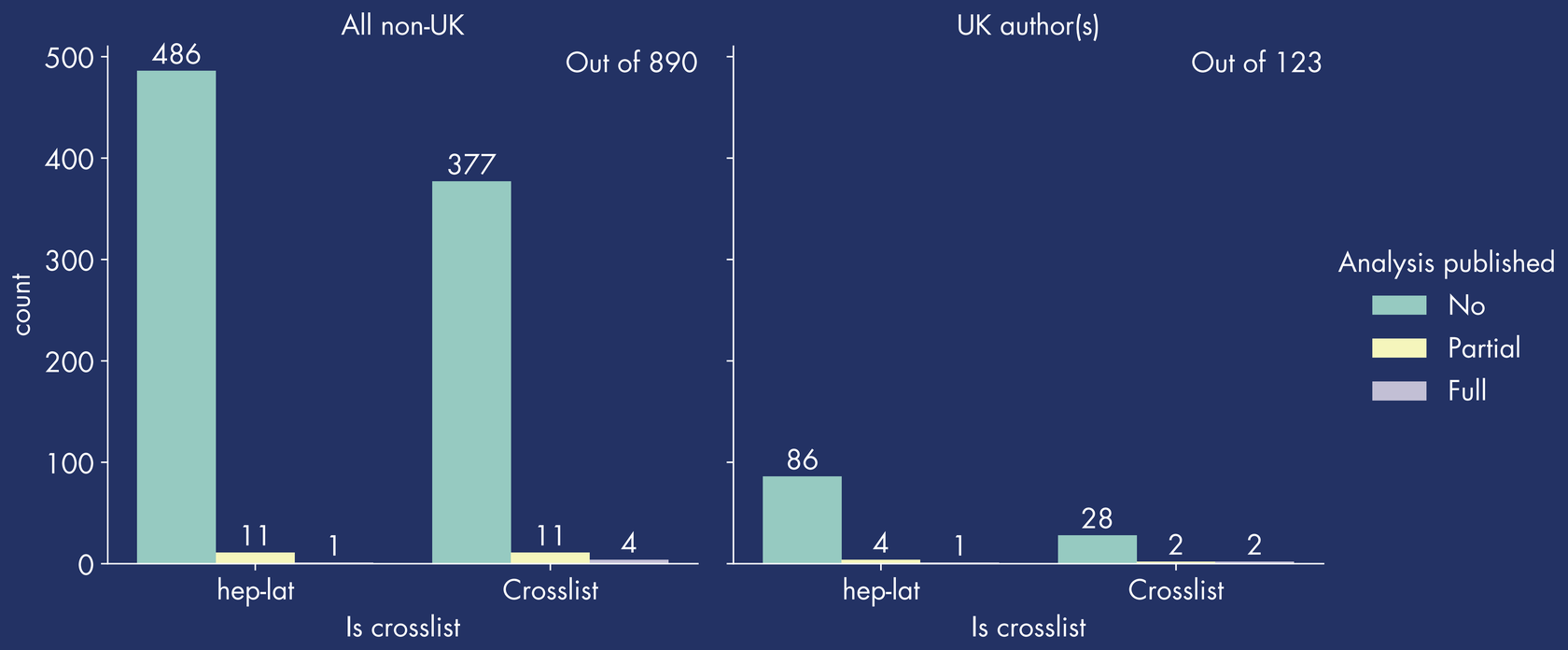
Do authors specify *any* software is used for analysis?



What software is specified?



Do authors publish a full analysis workflow?



Examples/Case studies

CalLat Collaboration, 2104.05226

- Performs measurements on configurations
  - Specifies software used
- Data and analysis workflow both on GitHub
  - Not tagged; not obvious which commit generated paper
  - Pure Python
- README indicates how each plot in paper generated
- All 20 figures automatically generated
- Tables not obviously generated

Scott Lawrence, 2111.13007

- Performs a conformal bootstrap analysis
  - Does not perform measurements on field configurations
- All 5 plots generated programmatically
- Code available on GitLab
  - Primarily reusable components
  - Majority Python
- Full set of plots can be generated from one `Makefile`

EB et al, 2202.05516

- Gauge configurations not shared
  - Modified HiRep code and parameters shared
- Measurement outputs available on Zenodo
- Almost all 20 plots and 6 tables generated programmatically
  - Remainder are schematic, not numerical
  - Table contents available in Zenodo data release
- Code available on GitHub and Zenodo
  - Mix of Python and Mathematica

Conclusions and next steps

Conclusions

- LFT has been at the forefront of many aspects of open science
  - Some areas suffer from first-mover disadvantage
- Opportunities remain to do more
  - Some low-hanging fruit
    - Specify software
    - Share existing code
  - Some require more effort
    - e.g. Automating analyses and presentation of data

Next steps

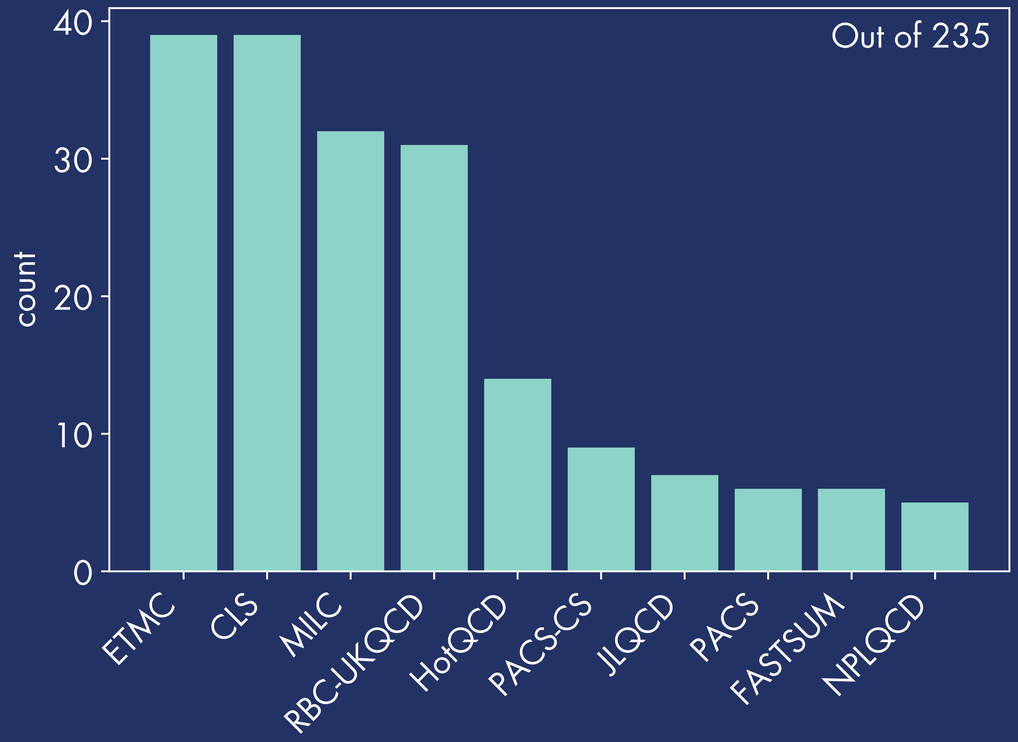
- Produce a manifesto of good practice in open science in lattice
- Develop tooling to better enable automated analysis and presentation
  - Aspiration: "easier to use than not to"
- Survey of reproducible and open science practices
  - Watch your inboxes

Thank you

Backup slides

Aside: The importance of a compute ecosystem

Who is generating configurations?



How is data analysis software cited?