

Research Data and AI

Solving the Survival Bias of Publish or Perish



UNIVERSITY
of York

Dr Penn Rainford
Research IT, UoY

Dr Penn Rainford

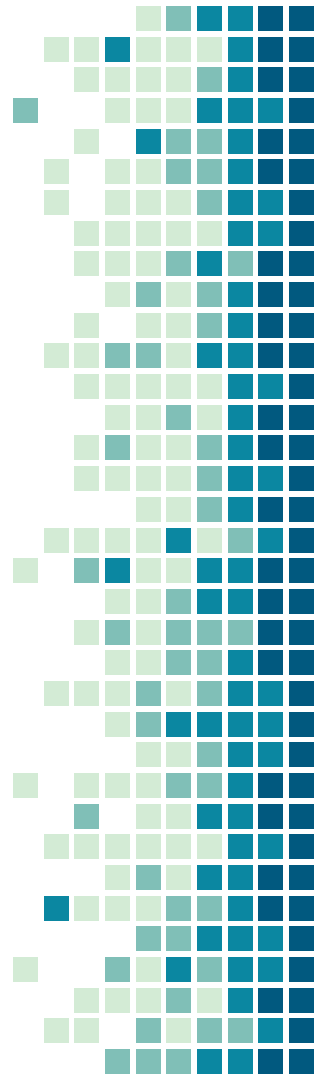
Research Software Engineer
University of York

- Interdisciplinary Research
- 6 years RA
- 7+ departments (5+ subjects)



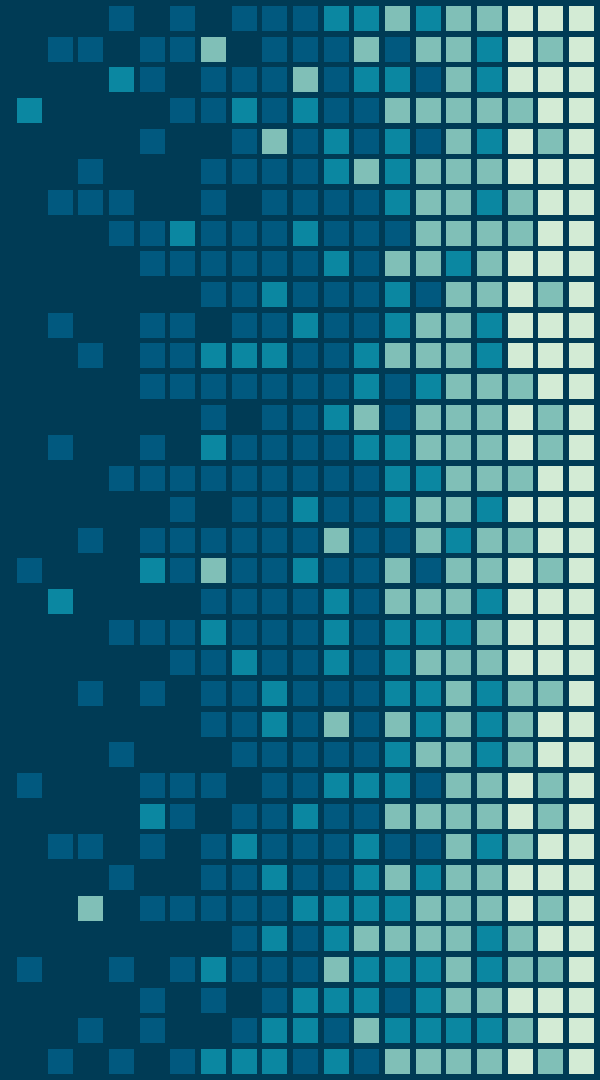
Research Data and AI

- (Current) Research Data Culture
- Publication Bias
- HPC and AI
- Research Culture Change
- Achieving Change



Research Data Culture

Current norms, tools and standards



Comment | [Open access](#) | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

[Mark D. Wilkinson](#), [Michel Dumontier](#), [Jesbrand Jan Aalbersberg](#), [Gabrielle Appleton](#), [Myles Axton](#), [Arie Baak](#), [Niklas Blomberg](#), [Jan-Willem Boiten](#), [Luiz Bonino da Silva Santos](#), [Philip E. Bourne](#), [Jildau Bouwman](#), [Anthony J. Brookes](#), [Tim Clark](#), [Mercè Crosas](#), [Ingrid Dillo](#), [Olivier Dumon](#), [Scott Edmunds](#), [Chris T. Evelo](#), [Richard Finkers](#), [Alejandra Gonzalez-Beltran](#), [Alasdair J.G. Gray](#), [Paul Groth](#), [Carole Goble](#), [Jeffrey S. Grethe](#), ... [Barend Mons](#) [✉](#)

[+ Show authors](#)*Scientific Data* **3**, Article number: 160018 (2016) | [Cite this article](#)[Save article](#)1.29m Accesses | 17k Citations | 2476 Altmetric | [Metrics](#)[i](#) An [Addendum](#) to this article was published on 19 March 2019

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

data

2634 | [VIEW METRICS](#)[PDF](#) [CITE](#) [TOOLS](#) [SHARE](#)

at data are science, and thus
ort of, much-improved data

”

ge of data by using new methods and
ata for new discoveries and
y available as sharing is common in
Data sharing practices in other fields
ms, are handled in many ways, using
et once removed from their initial
tionales for sharing data are
cial sciences, and humanities: (1) to
of publicly funded research available
ons of extant data, and (4) to advance
les differ by the arguments for



Journal Information
Journal TOC

Research Data Pipeline



Open Science and Open Data

Repositories

- Addgene
- BioBricks
- UK Biobank
- NCBI (GenBank, GEO and SRA)
- AiiDA
- GitHub
- JoVE
- JoSS
- Zenodo

Communities

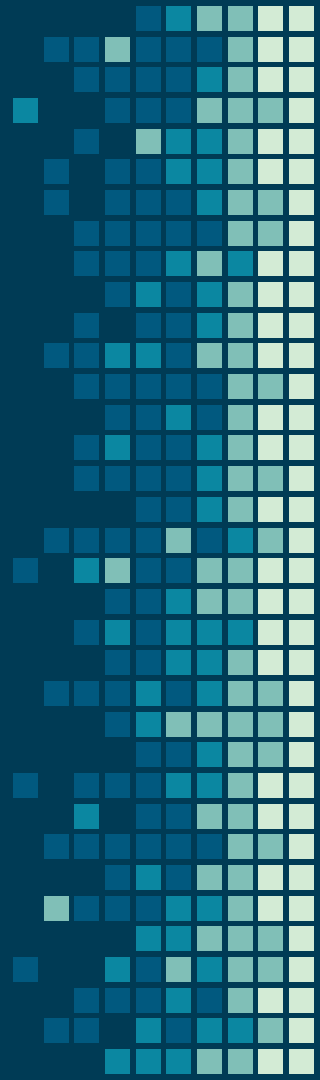
- GoFair
- SBOL
- Benchling
- Data Hazards
- FAIRsharing
- COMBINE
- ResearchGate
- reproducibility network
- SSI

Barriers

- Fixed funding
- Incentives
- Publication
- Time pressure
- Skills
- Standards
- Calibration
- Infrastructure
- Storage

Publication Bias

Novelty, certainty and quantity



Abstract

When studying analytical because the thing. Our independently immigration: Does researchers engaged from of pro-imm on public more negative different mechanism estimates.

► R Soc Open Sci. 2025 Apr 30;12(4):240688. doi: [10.1098/rsos.240688](https://doi.org/10.1098/rsos.240688)

Publication bias is bad for science if not necessarily scientists

[Remco Heesen](#)^{1,✉}, [Liam Kofi Bright](#)¹

► [Author information](#) ► [Article notes](#) ► [Copyright and License information](#)

PMCID: PMC12040460 PMID: [40309174](#)

Abstract

It might seem obvious that the scientific process should not be biased. We strive for reliable inference, and systematically skewing the results of inquiry apparently conflicts with this. Publication bias—which involves only publishing certain types of results—seems particularly troubling and has been blamed for the replication crisis. While we ultimately agree, there are considerable nuances to take into account. Using a Bayesian model of scientific reasoning we show that a scientist who is aware of publication bias can (theoretically) interpret the published literature so as to avoid acquiring biased beliefs. Moreover, in some highly specific circumstances she might prefer not to bother with policies designed to mitigate or reduce the presence of publication bias—it would impose a cost in time or effort that she would not see any benefit in paying. However, we also argue that science as a social endeavour is made worse off by publication bias. This is because the social benefits of science are largely secured via go-between agents, various non-experts who nonetheless need to make use of or convey the results of scientific inquiry if its fruits are to be enjoyed by society at large. These are unlikely to be well-informed enough to account for publication bias appropriately. As such, we conclude, the costs of having to implement policies like mandatory pre-registration are worth imposing on scientists, even if they would perhaps not view these costs as worth paying for their own sake. The benefits are reaped by the go-between agents, and we argue that their perspective is quite properly favoured when deciding how to govern scientific institutions.

Keywords: publication bias, filedrawer effect, preregistration, philosophy of statistics, replication crisis

[b.2024.2846](#)

nd familiar bird species

[ald Miles](#)⁴, [Henry Streby](#)¹

formation

familiar stimuli are well documented and
ans are drawn to charismatic, familiar
s permeate research choices among
research effort, such as aesthetics,
report that, from 1965 to 2020, nearly half
American male passerine and near-
to human bias: aesthetic salience (visual
versities within ranges (accessibility). We
red on journal covers had higher
ere studied about half as much as those
lge, and decisions based thereon, is heavily
disparity feeds a cycle of public interest,
ties and scientific narratives, shrouding
image of drab, distant, disregarded species.
y exacerbate organismal inequalities amid
ific inquiry.

harisma, honorific common name,
ornithology, avian ecology, species names

iped by evolution, culture and experience.
use humans are typically drawn to colours

researcher" bias: How our assumptions technology affect research of older ults

Catherine A. Clair^{1,2}, [Tonisha J. Melvin](#)², [Janiece L. Taylor](#)²

Martha Abshire Saylor²

partment of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health,
sore, MD, United States
ns Hopkins School of Nursing, Baltimore, MD, United States

roduction

earchers, we are comfortable with different aspects of the scientific method: Designing studies,
menting interventions, collecting and analyzing data, and synthesizing results. We progress in our
rs, solidifying our skills and potentially, our assumptions.

ptions are human. Assumptions are heuristics that allow us to more efficiently operate within
atural world (1). However, if we do not actively engage with and question our assumptions, we
re comfortable with our own biases.

ose with careers in the field of gerontology, we have undoubtedly heard comments like, "You
get good data by doing interviews with older adults over the phone," or "Older adults will not
gate in research if it involves technology." These assumptions—that older adults cannot operate
ology or cannot answer questions over the phone or on Zoom—reflect not only ageism but a
ific bias (2).

omfort with technology exists on a spectrum (3, 4). We know that many older adults are
ortable with technology, such as smartphones and tablets (5). Adoption of technology has
eased, particularly due to the COVID-19 pandemic. One study found using data from the
nal Health and Aging Trends Study reported that older adults were engaging in more technology-
activity and technology-based health communication (6). At the same time, we also know that

Confirmatory
Data

Unexplained
Data

Novelty



Expected
Data

Outlier
Data

Unnecessary
Data



Certainty



Quantity

Disagreeable
Data

Findable

Metadata inconsistencies

M

F

Accessible

Infrastructure unreliability

Lost repos

Dead links

I

A

I

S

Data standards

Code standards

APIs

Documentation

Interoperable

R

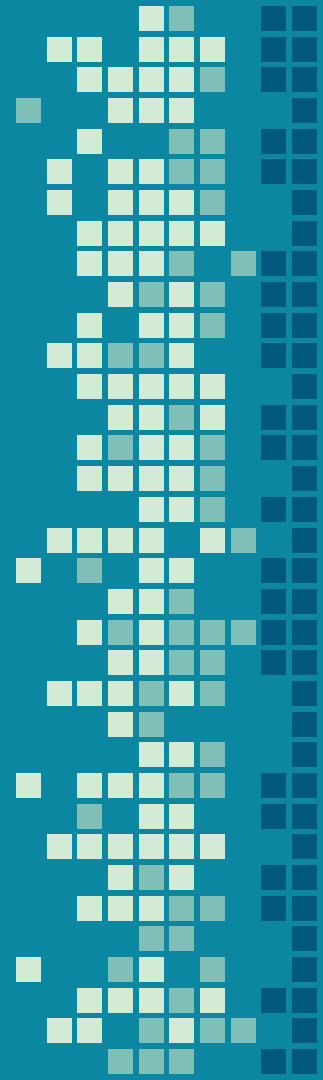
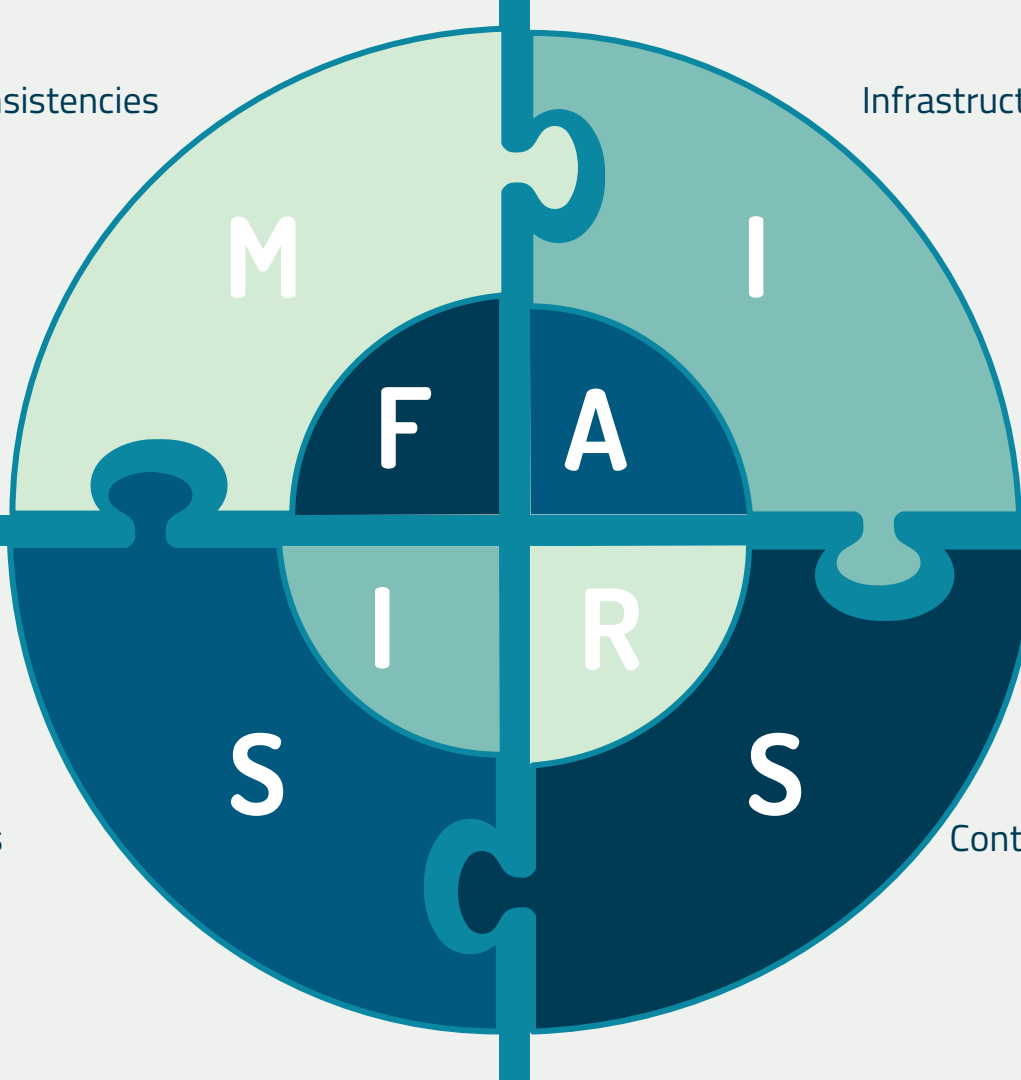
S

Context dependence

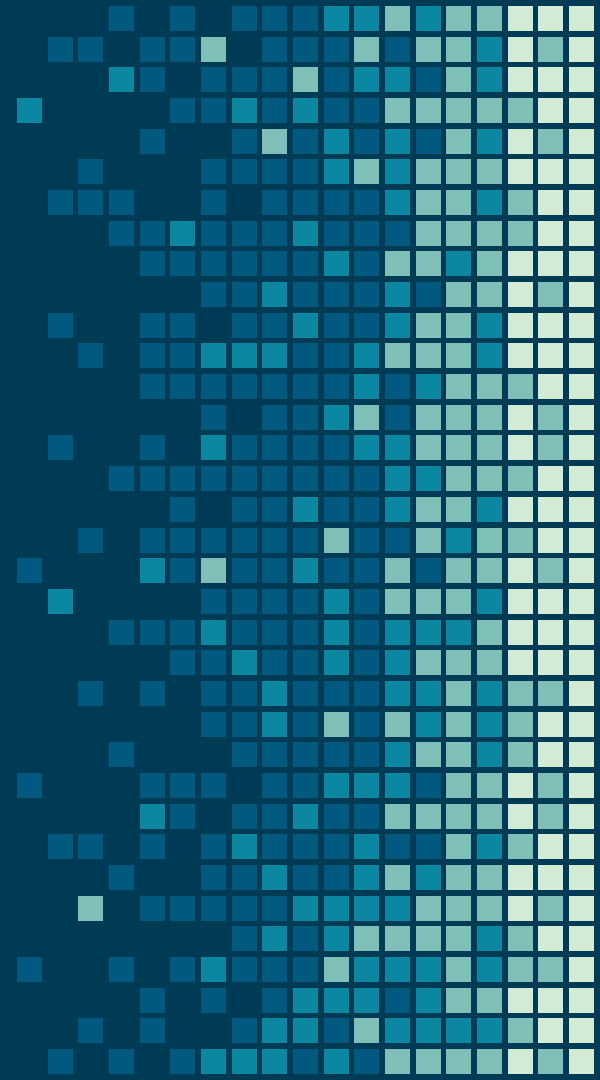
Specificity

Sensitive data

Reusable



Research Culture Change



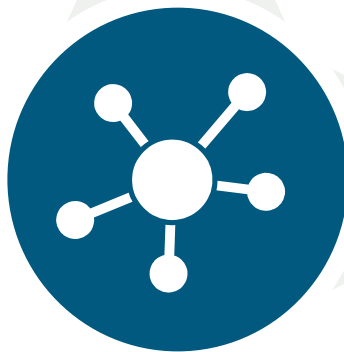
Time

- Data
- Research
- Reuse



Funding

- Data
- Infrastructure
- Training

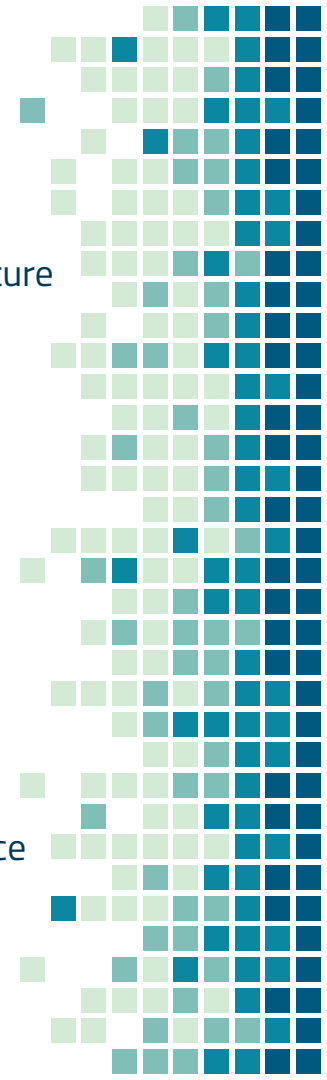


People

- Data Science
- Skills
- Continuity

Support

- Institutional
- Industry
- Policy



Time

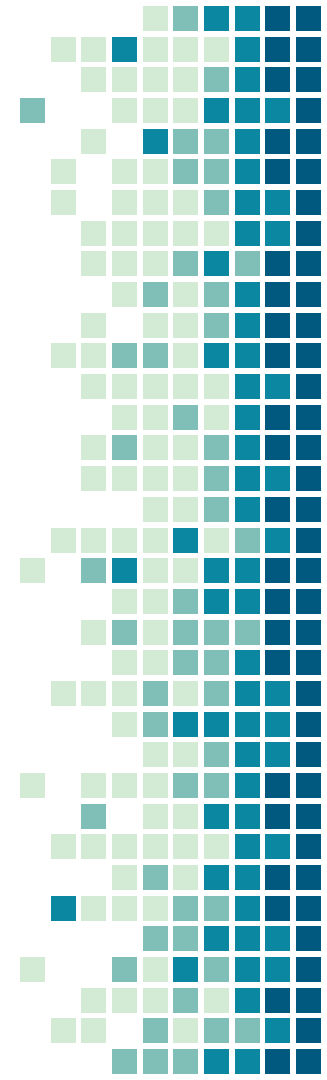
Data Review



Quality Assurance



Findability



People



Research
Software
Engineers

- Fixed term contracts
- Broad skill base requirement
- Minimal Continuity
- Publish or perish
- Moving and orphaned artifacts



Accessible
Infrastructure



Easy
Use
Tools

Funding

Time limited

Allocated



Novelty focused

Competitive



UK academics spend only a third of their time doing research

By Rachel Magee



Figure even lower for teaching staff as much time is spent on teaching. UK academics on research-only contracts spend just over a third of their time on research, a new survey has revealed.

The figure is much lower for academics who teach alongside their research, with the working hours being spent conducting research activities, according to a commissioned review into research bureaucracy.

For research-only academics, 7 per cent of their time is spent on administrative tasks and 8 per cent on external meetings.

Other tasks taking up their time included supervision of postgraduate students (6 per cent), peer review (3 per cent) and engaging with funders (3 per cent).

Magee, Rachel. 2022. "UK Academics Spend Only a Third of Their Time on Research - Research Professional News." Research Professional News, 8, 2022. <https://www.researchprofessionalnews.com/2022-8-uk-academics-spend-only-third-of-their-time-on-research/>



Rachel James · Sep 26, 2025 · 8 min read

Who gets to overwork in academia? Time and inequality in research

I was talking with a friend about work. She kept mentioning 'OT'. 'What's that?' I said. 'Overtime' she replied. I found it amusing that overworking had an acronym. 'Can't you work overtime if you need to?' she asked. 'No', I replied 'We are expected to work overtime but for no extra pay'. She found this very strange. And if you think about it, it doesn't really make sense. I am not saying that putting in an extra hour or two, here and there, need always be financially compensated. However, in academic research, the expectation is not that we work a couple of hours over our contracted hours. Reports suggest that some researchers are working nearly as much as having another full-time job. All for free. The Wellcome Trust Research Culture survey (2020) found that 37% of full-time researchers are working an average of 41-50 hours per week, with 11% working more than 60 hours (the standard contracted hours for a UK university is 35 hours per week). So even if you feel adequately compensated by your salary, if you are overworking, your hourly rate may not be much more than someone on a lower salary. This makes it challenging when you consider that people work to financially support themselves. With an ongoing cost-of-living crisis and the rise of income-boosting 'side hustles', the expectation of overwork for career progression is not an option that is equally available to all. Indeed Nature's 2023 post-doc survey reported that only 20% of postdocs can save the money they want (need) from their salary (Nordling, 2023). With a stubbornly opaque career path, this means researchers are constantly gambling the promise of job security against risk of future financial poverty.

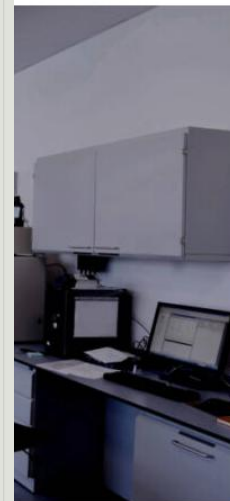
So why do we value overwork so much in academic research?

In a sector that has innovation somewhat assumed within its job title, overwork is counterintuitive. As neuroscientists, we know that a stressed, tired, and overworked brain is far from an ideal substrate for the creative ideation needed to innovate. This disconnect between optimal and actual behaviour can be partly explained by the structural and cultural shaping of the research sector. For example, the highly time-intensive performance metrics

James, Rachel. 2025. "Who Gets to Overwork in Academia? Time and Inequality in Research." Women in Neuroscience UK. September 26, 2025. <https://www.womeninneuroscienceuk.org/post/who-gets-to-overwork-in-academia-time-and-inequality-in-research/>

enduring cultural image of the 'ideal academic': someone who can work long hours, travel,

s UK jobs



g on fewer shoulders"

crificed' as UK Universities Cut Costs
Research Professional News. March
onalnews.com/rr-news-uk-
ced-as-universities-cut-costs-and-

rding to figures obtained

Most UK university professional staff badly hit by sector upheaval

By Anna McKie

Share

Image: [Vitaly Gariev](#), via Pexels

Restructuring and cuts are leaving support staff stressed and overstretched, especially in Scotland, report warns

Around two-thirds of professional and support staff at UK universities say change and uncertainty over the past five years have had a negative impact on them, the Higher Education Policy Institute has found.

A Hefi report—The Invisible Majority: Professional and Support Staff and the Future of UK Higher Education—says 65 per cent of respondents reported negative effects from change and uncertainty, rising to 81 per cent in Scotland.

Hefi and AHEP Consulting surveyed more than 400 professional and support staff between December 2025 and January 2026, as well as analysing 50 institutional annual reports and interviewing senior leaders.

Loss of academic staff and “cold spots” for humanities and social science degrees threaten strategically vital skills for UK, says British Academy

15 JUL 2026



the soci “UK HE Shrinking.” 2024. March 23, 2024. <https://qmucu.org/qmul-transformation/uk-he->

o rely shrinking/. Please
s in the branch.

How to apply for the Strike
Fund

'Super-university' created amid funding squeeze



The two universities, which legally merged last month, will retain their current names

7 September 2026

The universities of Kent and Greenwich have merged to form the UK's first "super-university" amid sector-wide financial pressures.

The new London and South East University Group (LASE) becomes the UK's third largest, with more than 50,000 students across four campuses in south-east London and Kent.

Leaders say the institutions are "much stronger" together as higher education providers face mounting economic strain, but the University and College Union previously raised concerns that the merger signalled a "broader problem for the sector".

The two universities, which legally merged last month, will still operate as distinct academic divisions within the group, retaining their current names.

Students will still apply to the institution they choose and degrees will be awarded in the name of Kent or Greenwich.

But the group will now be able to offer subjects that they could not before

Referring to this as the "quality-time-cost triangle", he said he believes the "competitive grant system only delivers two out of the three: it protects the efficiency of public money and assures quality, but it doesn't

'Excessive' financial risks threaten survival of many English universities, report warns

High levels of borrowing and rapid expansion among dangers identified by Higher Education Policy Institute

Sally Weale Education correspondent

Thu 9 Apr 2026 00:01 BST

Share

Prefer the Guardian on Google



The report suggests limiting each university's annual growth in student numbers to 5%. Photograph: aberCPC/Alamy

Many English universities are taking excessive financial risks that threaten not only their own survival but that of others in the sector, a thinktank has warned.

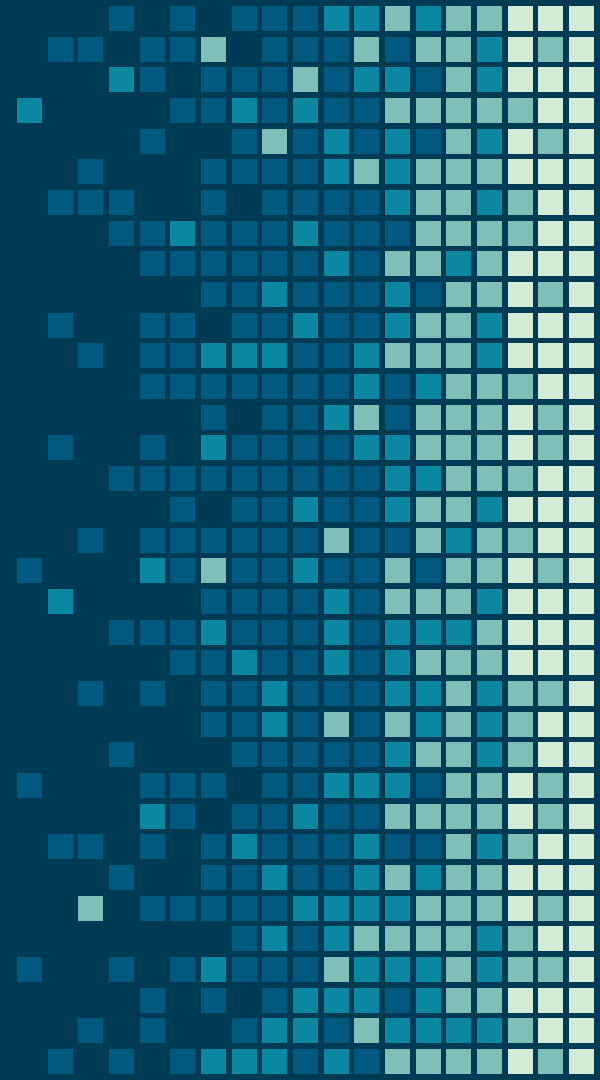
High levels of borrowing at some institutions and rapid expansion of student numbers are among the dangers identified in a report by the Higher Education Policy Institute (HePi).

It also flags over-reliance on international students and the growth of franchised provision – where a degree-awarding body authorises another to deliver all or part of a course – which it says are putting the sector at risk.

On rapid expansion, the report says Canterbury Christ Church University in Kent has almost tripled in size over the last decade, while Arden University

HPC and AI

No free lunch



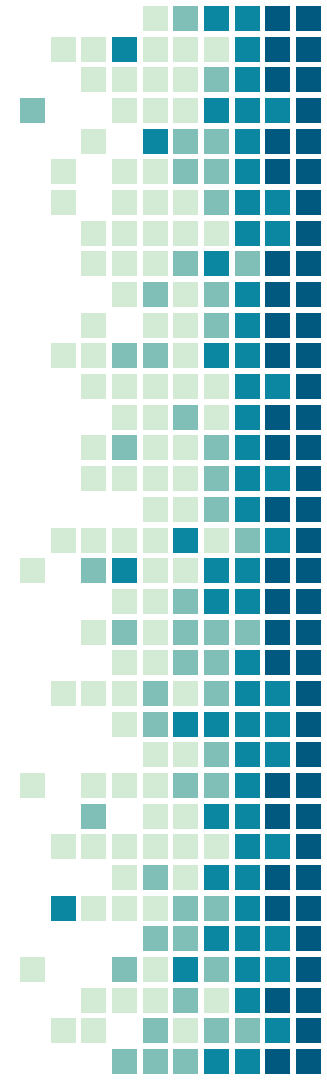
HPC and AI Support

- AI

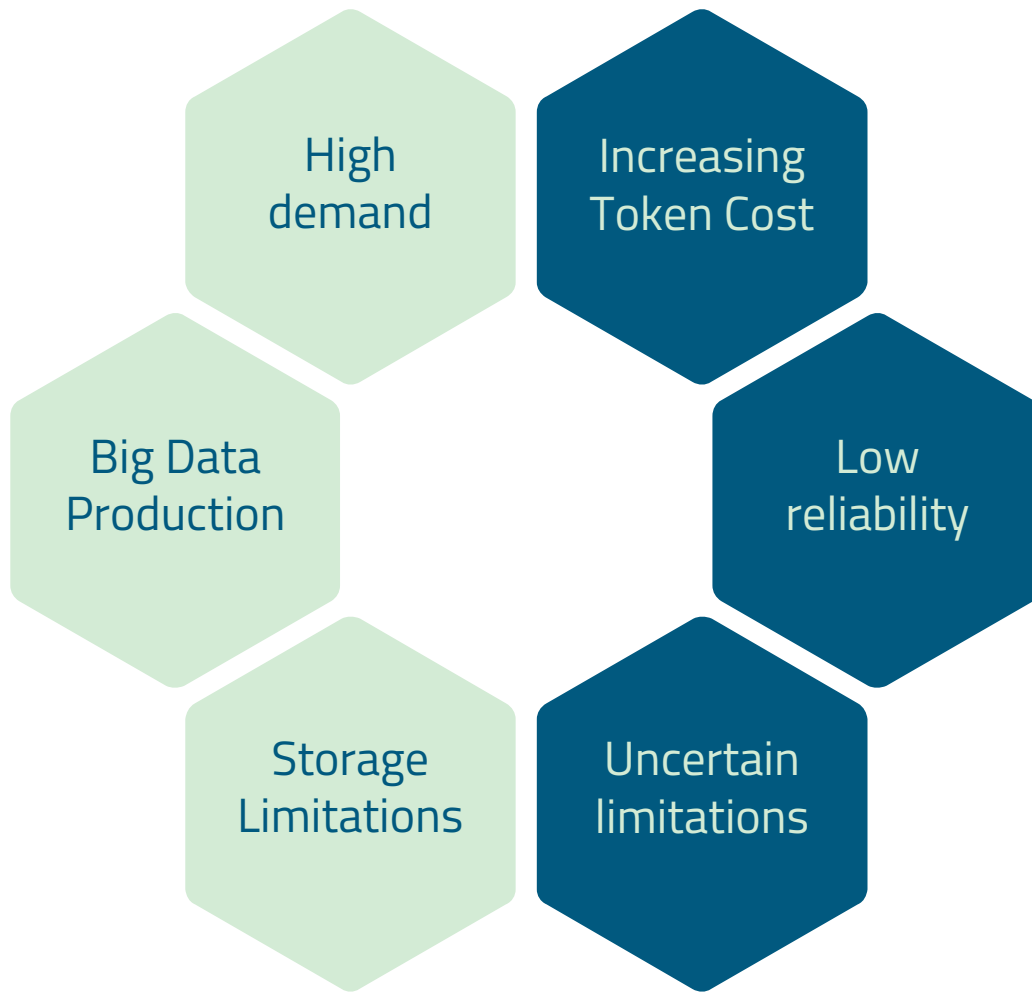
- Data conversion
- Documentation
- Containerisation
- Data
Review/Checking
(before publication)
- Tool Generalisation

- HPC

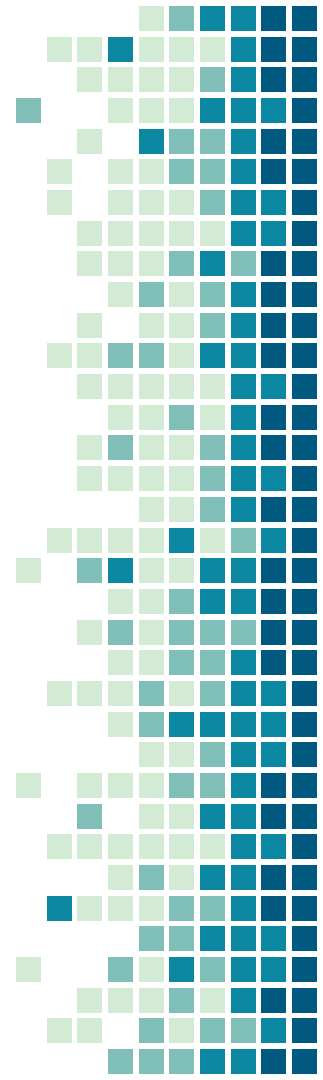
- Data conversion
- Tool application



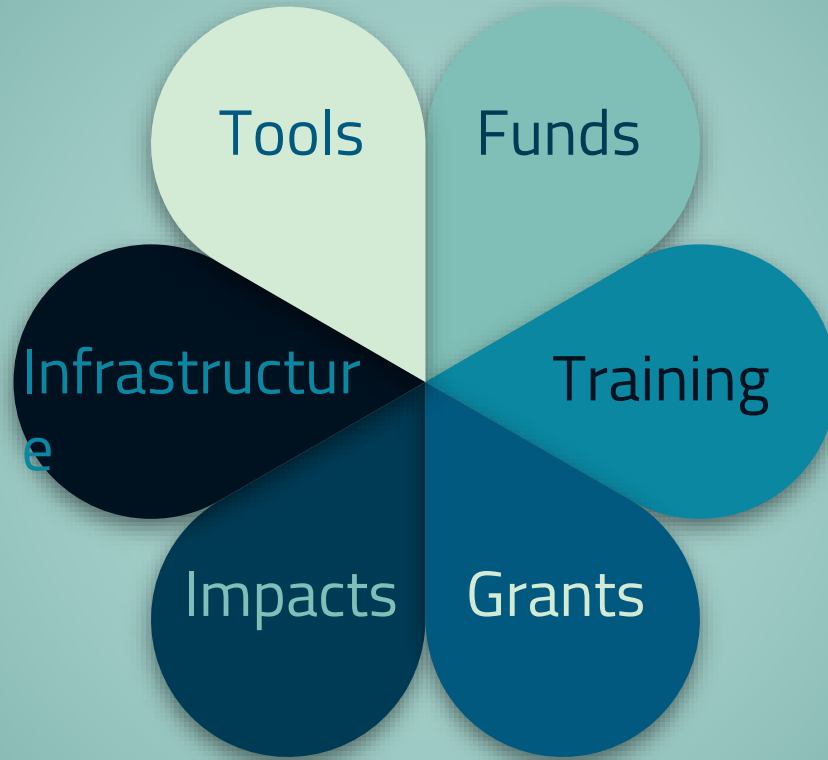
HPC



AI

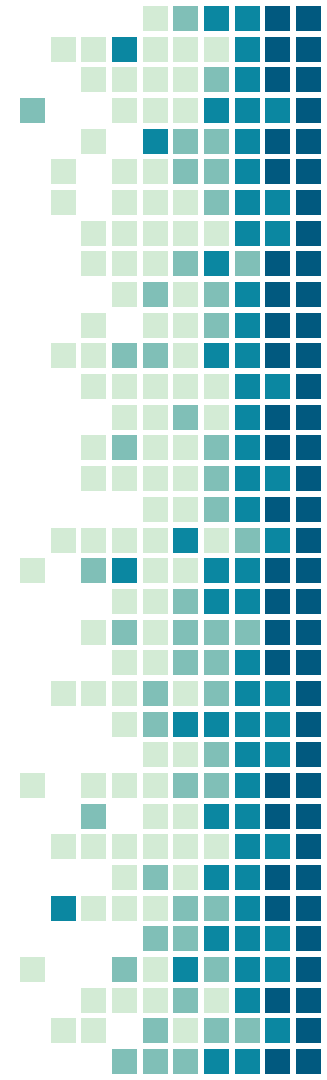


Support

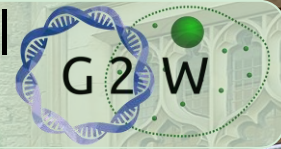


Conclusion

- Great options exist but need more development
 - Great practitioners exist but need more time
 - Great solutions exist but need better funding
 - Academia loves novelty but we need consistency more
- If you want action from academics...
 - You need a free lunch
 - Lunch is the only free time many academics have left



Many Thanks to Genome2Wholecell Workshop 2024



AIBIO UK network under BBSRC grant BB/Y006933/1,
the EPSRC-SFI grant EP/V027395/1+2, the Science Foundation
Ireland grant 21/EPSRC/3754, and the EPSRC CCPBioSim grant
EP/T026308/1.

Thanks to University of York, IT Services for travel funding.

Presentation template by [SlidesCarnival](https://www.slidescarnival.com/)



UNIVERSITY
of York