

Open research prize 2026 – Case studies

The Oxford Open Research Prize celebrates and rewards researchers and research support staff who use open research practices to make research more accessible, transparent and reproducible. The prize is funded by the Bodleian Libraries and administered through the Bodleian's Office for Open Research, in partnership with the Research Practice team and Reproducible Research Oxford.

Open research improves the quality, integrity, visibility, and societal value of research. By making publications, data, software, protocols, materials, methods, and engagement practices more openly available, researchers increase the reach, reproducibility, and impact of their work. Open research also supports public trust in scholarship and helps ensure that research outputs can be reused and built upon by others.

The University of Oxford already demonstrates strong commitment to open research through its policies, infrastructure, training, and support services. This prize aims to recognise excellent examples of open practice across all disciplines and career stages, while encouraging a research culture in which openness is recognised and rewarded.

Entries below are ordered by prize status, with winners and shortlisted entries appearing first.

Open research prize 2026 – Case studies ©2026 – Is licenced under the CC BY 4.0 International licence (<https://creativecommons.org/licenses/by/4.0>)

Principle author: Jacqueline Thompson

Co-Authors: see individual case studies

NARAIN / IDDO TEAM - ENABLING DATA REUSE TO FIGHT INFECTIOUS DISEASES (WINNER)	5
MUELLER – THE SNACKER TRACKER (WINNER – DPHIL PRIZE).....	7
AZAD – INVISIBLE EAST DIGITAL CORPUS (HIGHLY COMMENDED)	9
MARTINEZ-PANEDA – OPEN CODES AND SOFTWARE TO ACCELERATE THE ENERGY TRANSITION (HIGHLY COMMENDED).....	11
ASIF ET AL – FAIR EPIGRAPHY TEAM – OPEN INFRASTRUCTURE FOR ANCIENT WORLD INSCRIPTIONS (HIGHLY COMMENDED)	13
LUPPI - SYSTEMATIC EVALUATION OF FMRI DATA-PROCESSING PIPELINES FOR CONSISTENT FUNCTIONAL CONNECTOMICS (HIGHLY COMMENDED)	15
KHAN ET AL - HUMAN BONE MARROW ORGANOIDs: AN OPEN PLATFORM FOR TRANSLATIONAL RESEARCH AND INNOVATION (SHORTLISTED)	17
XU - SHINYTONE: AN OPEN TOOLKIT FOR THE PHONETICS OF TONE LANGUAGES (SHORTLISTED)	19
BERNABEU - SECURE AND SCALABLE SPEECH TRANSCRIPTION FOR LOCAL AND HPC (SHORTLISTED)	21
THE GOVERNMENT OUTCOMES LAB – THE INDIGO IMPACT BOND DATASET: MAKING DATA ON SOCIAL IMPACT BONDS OPEN, TRACEABLE AND REUSABLE (SHORTLISTED)	23
YUAN TAN - TEXTCOURT: UNLOCKING THE TEXTUAL WORLDS OF CHINESE COURT DRAMA THROUGH OPEN RESEARCH INFRASTRUCTURE (SHORTLISTED)	25
CLARKE - OPEN TOOLS AND STANDARDS FOR REPRODUCIBLE NEUROCHEMICAL IMAGING (SHORTLISTED)	27
PIT – ROBUSTNESS IN MISINFORMATION RESEARCH (SHORTLISTED).....	29
JEONG ET AL - A GLOBAL GEOSPATIAL DATASET OF RENEWABLE ELECTRICITY SUPPLY AND NETWORKS (SHORTLISTED).....	31

DIDI - FLIP2 OPEN BENCHMARKS FOR PROTEIN ENGINEERING (SHORTLISTED)	33
JURASKOVA - MLP-TRAIN PACKAGE: AUTOMATED ACTIVE LEARNING FOR REACTIVE MACHINE LEARNING INTERATOMIC POTENTIALS	35
GLADSTONE - MAKING OXFORD'S WATER SECURITY RESEARCH OPENLY AVAILABLE VIA ORA AND RESHARE	37
STRAUB - PHENOFHY: OPEN, REUSABLE TOOLING FOR PHENOTYPE DATA IN OUR FUTURE HEALTH	39
AULICINO - BUILDING AND SUSTAINING SHARED INFECTION BIOLOGY CAPACITY AT THE TIDU	41
ALKIRE ET AL - OXFORD POVERTY AND HUMAN DEVELOPMENT INITIATIVE (OPHI) – SHARING THE GLOBAL LEVELS AND TRENDS OF MULTIDIMENSIONAL POVERTY AS A GLOBAL PUBLIC GOOD.....	43
BERNABEU - REPRODUCIBLE AND EXPANDABLE MODULAR WORKFLOWS FOR THE CREATION OF FAIR EXPERIMENTAL STIMULI IN PSYCHOLINGUISTICS	45
HIARI - NAVIGATING UK ETHICS POLICY IN OVERSEAS FIELDWORK.....	47
JOSAN - OPEN EYE RESEARCH	48
THOMAS - GLASS: GEOCHEMICAL LABORATORY APPROACHES FOR SHARING SCIENCE.....	50
ABUSAMRA - SYSTEMATIC REVIEW: CIRCULATING TUMOUR CELLS (CTCS) AS A WINDOW INTO LETHALITY IN PROSTATE CANCER	52
KARIYAWASAM - OPEN DATA FOR BETTER HEAT-RISK DECISIONS: UNDERSTANDING HOW CITIES EXPERIENCE AND RESPOND TO EXTREME HEAT	53
BERNABEU - ADVOCACY FOR OPEN DATA GOVERNANCE IN INDUSTRY-FUNDED RESEARCH: AMENDING DATA AGREEMENTS	55
MUTUNAYAGAM - UNDERSTANDING HOW TWO COMMON RESPIRATORY INFECTIONS INTERACT IN THE NOSE OF HEALTHY ADULTS: RESPIRATORY SYNCYTIAL VIRUS AND STREPTOCOCCUS PNEUMONIAE	56

TAO - INCIDENCE, PREVALENCE AND MORTALITY OF MATERNAL CARDIAC DISEASE WORLDWIDE	58
DESEILLIGNY - LIQUID-LIQUID PHASE SEPARATION IN NANOCONFINED GEOMETRIES	60
HIGSON-SWEENEY - DEVELOPING MINDS DIGEST: CO-DESIGNING A NEW RESOURCE FOR RESEARCH DISSEMINATION IN YOUTH MENTAL HEALTH	62
FATEMEH - CORTICAL AND ALLELIC DETERMINANTS OF AGE-RELATED COGNITIVE DECLINE: AN INTERDISCIPLINARY APPROACH TO THE NEUROSCIENCE OF AGEING..	64
RAJA - QUADMET-FORMER: PRE-TRAINING ON QUANTUM MECHANICAL PROPERTIES IMPROVES ADMET PREDICTION.....	66
WALKER SILVERMAN - CAN SOCIAL MEDIA CHANGE ATTITUDES TOWARDS REFUGEES?	68
COMPUTER SCIENCE TEAM - INCLUSIVE RESEARCH DESIGN FRAMEWORK.....	70
JACOB - BUILDING AN OPEN MODEL OF CHILDHOOD BRAIN CANCER TO ACCELERATE TREATMENT DISCOVERY	71
ALFURAIH - POROUS ARCHIVING: METHODOLOGICAL RESPONSE TO POLES OF ARCHIVAL HESITATION AND LOSS ANXIETY	73
PU - OPEN RESEARCH IN EARLY DRUG DISCOVERY.....	75
CAPEL ET AL - DEPT OF STATISTICS - SABDAB2: THE STRUCTURAL ANTIBODY DATABASE IN THE AGE OF MACHINE LEARNING	77

Narain / IDDO team - Enabling data reuse to fight infectious diseases (WINNER)

What open practices were used?

Infectious Diseases Data Observatory (IDDO) has developed an open research model that makes individual participant data (IPD) genuinely reusable. We convene scientific advisory groups to identify priority questions, assemble IPD from studies worldwide, and curate datasets to the internationally recognised CDISC SDTM standard. Dataset context is preserved, persistent identifiers support FAIR data, and harmonised datasets are available for free to non-profit researchers. We share pseudonymised rather than fully anonymised data to retain scientific detail for robust analyses. Our repository's infrastructure and governance framework are certified by CoreTrustSeal, and publicly share coding that supports researchers to work with repository data.

Why were these open research practices chosen?

These practices were chosen because sharing clinical data does not guarantee reuse: datasets can be difficult to find, inconsistent, or inaccessible without appropriate governance. Harmonisation reduces repeated curation, while pseudonymisation preserves the detail needed to analyse important subgroups, including pregnant women and young children. The model also supports research equity: contributors retain ownership, receive credit for reuse and can participate in collaboration, while non-profit researchers receive harmonised data at no cost.

How were these open research practices implemented?

We work with global scientific communities to identify questions that cannot be answered by individual studies alone. We bring together relevant IPD, preserve its context, harmonise data to CDISC SDTM standards, assign persistent identifiers, and apply a governance framework that balances openness with participant protection and contributor rights. Applications are reviewed independently and data are transferred securely, turning fragmented datasets into a reusable shared resource.

What challenges were encountered?

The main challenge is balancing openness with responsible data protection. Pseudonymised data retain more scientific value than fully anonymised data, but require more demanding governance. We therefore established robust legal, ethical and operational processes for contributor ownership, access decisions, secure transfer and compliance. Harmonising heterogeneous datasets while preserving their context also requires substantial work.

What difference or impact did these open research practices make?

Our model has enabled up to 90% of contributed datasets to be reused. IDDO pooled analyses have changed global malaria treatment recommendations for first-trimester pregnancy, enabled middle-income country-based researchers to demonstrate simplified age-based dosing, and to produce a large global analysis of antimalarial resistance. Overall, the project increases the value of existing research, enables new discoveries from shared evidence, strengthens collaboration, and supports evidence-based changes in global health policy.

Mueller – The Snacker Tracker (WINNER – DPHIL PRIZE)

Marissa Mueller

What open practices were used?

An open-science ethos drove the development, validation, and publication of the *SnackerTracker*—a novel system we created to continuously and non-invasively measure laboratory mouse feeding behaviour (**Figure 1**). The device presents as a refinement to animal research by improving data quality, supporting animal welfare, and reducing experimental costs; one *SnackerTracker* can be built for ~£120/cage, while commercial systems cost ~£2,000-£10,000/cage. Rather than simply publishing a device description, we sought to create a freely available resource so others can reproduce and expand upon our work.

Why were these open research practices chosen?

We adopted a comprehensive range of open research practices. We published the [resulting manuscript in Wellcome Open Research](#)—a journal which provides immediate open access followed by transparent peer review. We shared *SnackerTracker* construction and experimental protocols through the [Arduino Project Hub](#) and [protocols.io](#). We deposited hardware schematics to the [NIH 3D Print Exchange](#), enabling laboratories to adapt and build the device or to send design files to electronics shops for contracted fabrication. Source code is available via [GitHub](#), and versioned releases are archived via [Zenodo](#) to support collaborative development. We deposited additional *SnackerTracker* documentation to [Figshare](#) to ensure that all project components are available in a singular location for ease of reference.

How were these open research practices implemented?

This approach mandated careful planning beyond the preparation of a conventional manuscript. We invested in producing clear instructions such that researchers with different levels of technical expertise can build devices, perform experiments, and analyse data. We provided checklists to promote transparent reporting so that future researchers who use *SnackerTrackers* may understand how data is acquired and which processing variables may compromise comparisons between studies.

What challenges were encountered?

Financial concerns both motivated *SnackerTracker* design and complicated its production—I funded initial prototype development using savings from my DPhil stipend. In this respect and more broadly, I believe access to cost-effective and technically accessible equipment should be more commonplace in research.

What difference or impact did these open research practices make?

Open science practices have increased the *SnackerTracker*'s potential impact. Researchers can freely access everything needed to reproduce our findings, critically evaluate our system, and adapt it for future applications. The *SnackerTracker* demonstrates how combining open-source hardware, software, and peer review can strengthen research quality and foster a more collaborative research culture.

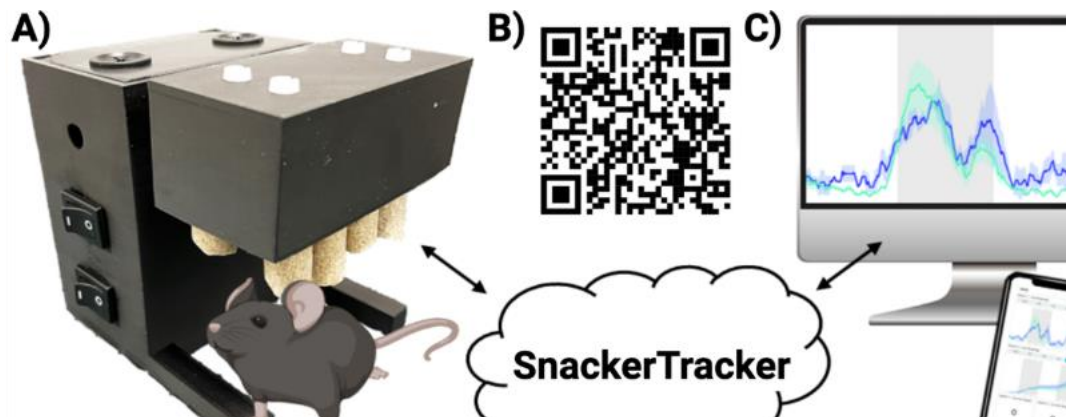


Figure 1. *SnackerTracker* design and operation. Illustrates **A)** the final prototype with a roughly to-scale cartoon mouse, a **B)** QR link to a *SnackerTracker* overview video, and **C)** data analysis and cloud-based capabilities including remote operation and monitoring.

Azad – Invisible East Digital Corpus (HIGHLY COMMENDED)

A. Azad, E. Shawe-Taylor, M. Arghandehpour, H. Smith, N. Saeed

What open practices were used?

Four key practices were used: 1) a freely searchable digital corpus, with data and source code open for download and reuse; 2) interoperable data incorporated into larger resources, including the Open Islamicate Texts Initiative (OpenITI- a repository of more than 5 billion words of medieval Arabic texts); 3) a bilingual English-Persian interface; and 4) open-access research, including three books with Edinburgh University Press and 23 freely available Documents of the Month peer reviewed online publications.

Why were these open research practices chosen?

Our aims were scholarly, ethical and practical. **Scholarly:** we wanted to encourage multiplicity, enabling other projects to reuse our methods, code and data, and the IEDC to connect with larger research infrastructures. **Ethical:** our documents have suffered from displacement, uncertain provenance and limited accessibility; making them freely available contributes to the digital repatriation of cultural heritage, particularly for communities in Afghanistan, Iran and Central Asia. **Practical:** open access helps overcome language and financial barriers and fulfils our ERC (No. 851607) and AHRC funding requirements.

How were these open research practices implemented?

We convened the Future Philology: Digitisation and Beyond (2021) conference to learn from comparable projects, then worked with a specialist digital humanities company to build a lightweight, low-bandwidth Django platform supporting rare and non-Latin scripts. Records are downloadable in DOCX, JSON and TEI-XML, enabling integration with projects like OpenITI. Code and data are openly available through GitHub and Zenodo. We also provide online training and outreach and prioritise open-access publishing.

What challenges were encountered?

The multilingual corpus (over 10 languages and scripts) presented technical challenges, particularly non-Latin and right-to-left scripts. Digitisation required decisions about philological and digital-markup standards, while image permissions were sometimes difficult to secure. We collaborate across politically complex contexts, including Afghanistan, Iran, Israel, China, Russia, Europe, the UK and USA. Uncertain provenance and the location of manuscripts in Western and private collections also raise ethical questions. Transparency about these histories while increasing access remains central to our approach.

What difference or impact did these open research practices make?

By June '26, the IEDC had reached over 40,000 users, including from Iran, Afghanistan, and Central Asia. The bilingual interface allows many of these users to reach the database in a familiar language. Our open approach has led to collaborations with other research groups, including inter-institutional panels and seminars. The IEDC received the Resource Award at the DH Awards 2025, recognising its contribution as a digital research resource.

Martinez-Paneda – Open codes and software to accelerate the energy transition (HIGHLY COMMENDED)

Emilio Martinez-Paneda / Mechanics of Materials Lab

What open practices were used?

All the codes developed by members of the Mechanics of Materials Lab (Department of Engineering Science, University of Oxford) are immediately and freely released (see <https://mechmat.web.ox.ac.uk/codes>).

The codes span a wide range of programming languages and platforms, from ‘add-ons’ to the commercial finite element packages used by industry to open-source codes in C++, Julia, Matlab and Python developed fully in-house. In addition, we have developed a number of pieces of freely available software, such as TDS Simulator (<https://tdssimulator.web.ox.ac.uk/>) and Abaqus2Matlab (<https://abaqus2matlab.web.ox.ac.uk/>).

Why were these open research practices chosen?

Members of the Mechanics of Materials Lab conduct interdisciplinary research to address fundamental scientific challenges that can accelerate our transition to a net-zero carbon society and assist in tackling our energy and climate crises.

The vast majority of our work involves developing computer models that can allow us to identify when critical green energy infrastructure will fail, preventing catastrophic failures and enabling safe operation and optimisation.

Our commitment to open science is intended to facilitate scientific and technological progress while strengthening reproducibility, transparency and trust. This is particularly important when addressing something as important and global as the fight against climate change.

How were these open research practices implemented?

All codes developed by members of the Mechanics of Materials Lab are immediately and freely released through our website and associated project websites. To facilitate their use by others, members of the Lab inspect their code for bugs, clean it, and develop extensive documentation.

We also support users after release, addressing email queries related to our codes and their application to other problems. We currently receive approximately 20 such emails per week.

What challenges were encountered?

This commitment to open science and reproducibility comes at a cost. Preparing codes for public use, developing documentation and providing ongoing support require significant time from members of the Lab.

This attitude is not common in our scientific community. Other research groups argue that their codes constitute their 'know-how', providing a competitive advantage and allowing them to secure additional research projects and, particularly, contracts with industry. While these are fair points, at Oxford's Mechanics of Materials Lab we are happy to sacrifice our time and potential revenue if this helps to foster a culture of open science and facilitates scientific and technological progress around the world.

What difference or impact did these open research practices make?

The impact has been profound, with more than 20,000 downloads and the codes being actively used by hundreds of research groups and dozens of companies around the world.

The audience is the broadest possible: from energy engineers to researchers in developing countries who do not have the means to pay for commercial software licences. Making our codes freely available also enables researchers to reproduce the results of our journal publications.

Quite frequently—approximately once per month—we receive an email from a student in a developing country stating that only thanks to our generosity have they been able to carry out their PhD thesis. It cannot get more rewarding than that.

Further examples and resources:

<https://mechmat.web.ox.ac.uk/codes>

<https://tdssimulator.web.ox.ac.uk/home>

<https://abaqus2matlab.web.ox.ac.uk/>

<https://www.empaneda.com/codes/>

Asif et al – FAIR Epigraphy team – Open Infrastructure for Ancient World Inscriptions (HIGHLY COMMENDED)

Imran Asif (RSE, Digital Scholarship, University of Oxford) & Jonathan Prag (Merton College, University of Oxford)

What open practices were used?

The FAIR Epigraphy project (inscriptions.org) developed open infrastructure for digital epigraphy: (1) controlled vocabularies published as Linked Open Data (LOD) using SKOS, with stable URIs and downloadable RDF Turtle files under CC-BY-SA-4.0; (2) a citable bibliographic platform with stable LOD URIs; (3) a comprehensive CIDOC-CRM-based ontology covering the full epigraphic workflow; and (4) a live SPARQL endpoint and browser enabling cross-project data integration. All code and data are published across nine public repositories in the FAIR-epigraphy GitHub organisation (<https://github.com/FAIR-epigraphy>), and findings are published open-access in the Journal of Open Humanities Data (2025). Materials are also being deposited, as they are released, on [Zenodo](https://zenodo.org).

Why were these open research practices chosen?

Existing digital epigraphy projects publish data through legacy HTML interfaces with limited reuse potential. A Linked Open Data approach was chosen as the most neutral path: flexible enough to integrate with diverse existing workflows without requiring projects to abandon their current systems, while enabling genuine cross-project interoperability for the first time.

How were these open research practices implemented?

The project began with a community scoping survey of 13 partner projects, published openly on Zenodo (<https://doi.org/10.5281/zenodo.6610696>). The full technical infrastructure — ontology serialisation scripts, triple store deployment, SPARQL endpoint, [EpigraphyBrowser](#), vocabulary tooling, and [Identifier Resolver](#) — was designed and built by Imran Asif. Vocabularies were developed using SKOS and reviewed with twelve major epigraphic databases and the Epigraphy.info community. The bibliographic platform was co-developed with the Association Internationale d'Épigraphie Grecque et Latine (AIEGL).

What challenges were encountered?

Previous community efforts (EAGLE vocabularies; EpOnt ontology) had stalled due to resourcing and consensus difficulties. FAIR Epigraphy addressed this by building on existing resources rather than from scratch, combining structured community engagement with a dedicated RSE role focused on delivery and publication — ensuring

outputs reached completion and were formally published under open licences. The challenge for the future (for which a further grant application is currently under review) is to facilitate and support the use of these resources by existing and future projects, in a context of academic conservatism and limited resources.

What difference or impact did these open research practices make?

FAIR Epigraphy delivers what two prior community efforts could not complete: a fully published, openly licensed, interoperable infrastructure layer for an entire research discipline. The CHANGE project at Oxford actively uses the Identifier Resolver to map Trismegistos IDs across epigraphic databases. The AIEGL has adopted the bibliographic platform on its own website (<https://www.aiegl.org/grepiabbr.html>). The Epigraphy.info international community engaged directly in vocabulary development. The open-access JOHD publication ensures the methodology is citable and reproducible, enabling other disciplines to adopt the same approach. Together, these outputs make interoperable ancient world data a practical reality.

Luppi - Systematic evaluation of fMRI data-processing pipelines for consistent functional connectomics (HIGHLY COMMENDED)

Andrea Luppi

What open practices were used?

This project used open research practices to address a major reproducibility challenge in neuroimaging: arbitrary analysis choices lead to different conclusions from the same functional MRI data, jeopardising credibility. We established 5 criteria for trustworthy data processing. We applied these criteria to systematically evaluate 768 neuroimaging data-processing pipelines: the most ever compared. Our openly available results identify few solid options in a sea of variability: a striking 98% of pipelines failed at least one criterion. Inappropriate data-processing can produce results that are not only misleading, but systematically so. This work provides a reusable benchmark for functional MRI studies, helping researchers choose methods based on openly available evidence and supporting more reliable neuroscience.

Why were these open research practices chosen?

This work involved open science practices at multiple levels. The main contribution to open science is the freely available [Interactive Pipeline Selection Tool](#). The tool is made fully and freely accessible to the entire community to inform best practices in neuroimaging. All results are made publicly available, providing a full breakdown of 768 pipelines' performance across each criterion and each dataset, totalling >13,000 unique evaluations. We made the evidence behind each of our recommendations fully transparent through Open-Access publication (Luppi et al., 2024 *Nature Communications*: <https://doi.org/10.1038/s41467-024-48781-5>), sharing of source data, extensive supplementary materials that detail all analyses, and clearly documented data and software access routes, enabling others to inspect, reuse, and challenge the recommendations. These practices were chosen because the central scientific problem was itself about transparency and reproducibility. If these choices are not made explicit, researchers and readers cannot know whether a finding reflects biology, disease, or method.

What challenges were encountered?

This positive impact was achieved in the face of considerable challenges, including one reviewer who labelled our unbiased benchmarking of hundreds of pipelines, rather than following conventional wisdom, “a waste of time and electricity”. Another open-science practice, positing of a preprint, enabled us to alert the community to our findings and

recommendations already in 2021, well before final publication in 2024. We chose to persevere because the existence and prevalence of systematically misleading pipelines is a threat that would have remained unknown to the neuroscience community, had it not been for this work.

What difference or impact did these open research practices make?

The impact is both technical and cultural. Technically, the project provides tangible criteria for guiding robust and reproducible data-processing choices. However, these criteria are not left as an empty “wish list” for each researcher to figure out by themselves. On the contrary, we also provide the most extensive evidence base to date, to enable users to make appropriate choices with confidence. Not only do we identify which pipelines meet all criteria: we also explain where every other pipeline falls short, and what are the consequences of making an inappropriate choice. An inappropriate pipeline can produce results that are not only misleading (statistically significant in the opposite direction as the truth), but replicably so across multiple datasets. This can turn replicability against researchers, with devastating consequences: an intervention that makes disease worse may be systematically mis-identified as curing it. Culturally, this work promotes a shift from pipeline choice by convention to pipeline choice by transparent, openly available evidence. This improves research quality, reproducibility, interpretability, and trustworthiness, and gives researchers, reviewers, and trainees a practical resource for more reliable neuroimaging. The impact of this work on the field and research quality is already noticeable, reflected in widespread adoption (>100 citations) and over 25,000 reads since publication in June 2024.

Khan et al - Human bone marrow organoids: an open platform for translational research and innovation (SHORTLISTED)

The Oxford Bone Marrow Organoid Team (Abdullah Khan, Yuqi Shen, Aude-Anaïs Olijnik French, Zoe Wong, Jasmeet Reyat, Eleanor Murphy, Bethan Psaila), MRC WIMM, Radcliffe Department of Medicine

What open practices were used?

We developed the first human induced pluripotent stem cell (iPSC)-derived bone marrow organoid – a three-dimensional model of the marrow – as a shared platform. We combined open-access publications, sequencing data and analysis code with hands-on training and an open community forum, making both research outputs and practical expertise accessible.

Why were these open research practices chosen?

Realistic human bone marrow models constrain research into often-incurable blood cancers. We aim to create a platform that other laboratories could adopt and extend. Publication alone cannot reliably transfer a complex protocol: reagents, iPSC lines, culture conditions and operator technique can all affect outcomes. Therefore, beyond accessible papers, data and code, we provided practical training and ongoing community support for successful adoption.

How were these open research practices implemented?

We built three connected routes: 1) Open-access publications, data and code: We published the original model (Cancer Discovery, 2023), a step-by-step protocol (Nature Protocols, 2024) and the combined bone and lympho-myeloid organoid (Cell Stem Cell, 2026), with sequencing data in GEO and analysis code on GitHub to support independent scrutiny and reproduction. 2) Hands-on training: Since 2023, we have delivered five intensive, two-week training camps. Researchers from Oxford, UCL, Imperial, King's, Birmingham, DKFZ Heidelberg, Harvard and Erasmus MC practised critical steps under supervision. 3) Community support: We established an open Google Group in 2023, that now connects over 130 members who exchange protocol updates, troubleshooting advice and adaptations.

What challenges were encountered?

Protocol sensitivity was the main challenge: written methods could not capture every practical variable. We addressed early adopters' failures through protocol refinement, online troubleshooting and support, sharing solutions across the community.

Sustaining this work requires substantial time with limited formal recognition, constraining programme expansion.

What difference or impact did these open research practices make?

More than 60 laboratories worldwide have independently established the model, turning a single-laboratory innovation into a shared international resource. The original paper has received over 230 citations, but adoption provides the clearest evidence of its reach. Open data and code enable independent scrutiny and reanalysis, while hands-on training and public troubleshooting through open forum reduce technical barriers and duplicated effort. This has also led to collaborations with Oxford spin-outs, AstraZeneca, GSK and other partners, extending the platform's translational reach.

Xu - Shinytone: An open toolkit for the phonetics of tone languages (SHORTLISTED)

Chenzi Xu

What open practices were used?

Shinytone is a free web application and R package that takes a citation tone phonetic study from raw recordings to publication-ready tone transcriptions, analysis, and visualisation. Every layer is open: source code under GPL-3 on GitHub; an archived release on Zenodo with a DOI; public documentation with a tutorial and a reference for all 35 exported functions; an openly bundled dataset (38,808 f0 samples, 13 speakers, 6 tones); and open methods, since the tool implements the recommendations of our cross-linguistic review of citation-tone methodology (Xu & Zhang 2024, JASA). Two features make each individual analysis open as well: every plot and model panel prints the R code that produced it, and every manual correction of a pitch track is written to a downloadable edit log.

Why were these open research practices chosen?

Our review of 136 citation-tone studies across 129 tonal varieties found methods diverging at almost every step: f0 extraction settings unreported in 130 of 136 studies, corrections rarely described. This limits how well tone inventories can be compared across languages and leaves most published descriptions unreproducible. About half of the world's languages use pitch to distinguish words, and many are documented by fieldworkers and students without a phonetics lab. We needed a tool that lowers the barrier while raising the floor on transparency, so that the easy route is also the documented, reproducible one.

How were these open research practices implemented?

Shinytone is written in R and Shiny, with the analytical functions packaged separately so the same code runs in the browser, in a script, or in an R Markdown report. Open infrastructure carries the rest: GitHub for source and issues, GitHub Actions rebuilding the documentation site on every commit, Zenodo's GitHub integration archiving each release with a DOI, and a privacy policy (uploads are processed in memory, never stored). Interoperability is deliberate: recordings (WAV), Praat Pitch files and TextGrids are read directly, every intermediate result is exported as CSV, figures as PNG, and a Praat script reproduces the extraction in Praat with the user's settings.

What challenges were encountered?

The central tension is between the convenience of a point-and-click tool and the reproducibility of a script. Our answer was to make the interface generate the script: the

“Show R code” buttons exist. Pitch tracking was a second problem: automatic f0 is unreliable on creaky or breathy voice, and manual correction is usually invisible in publications, so we built correction into the tool but made every edit logged and exportable. Practically, hosted uploads are capped and some recordings cannot leave an institution for ethical reasons, which is why the identical app can be run locally.

What difference or impact did these open research practices make?

Shinytone replaces a patchwork of scripts and programs with one open pipeline whose outputs are comparable across studies and languages. Reproducibility: analyses leave the app as code and data, and corrections leave an audit trail. Accessibility: a phonetician, fieldworker, or student needs only a browser, and the sample data mean a first analysis takes minutes. Transparency: the review that motivates the tool and the tool itself are both public, citable, and versioned. Collaboration and teaching: it has a pedagogical function enabling efficient demonstration of tones in class. Culture: We believe it will change the default practice for how tone data are analysed and reported.

Bernabeu - Secure and scalable speech transcription for local and HPC (SHORTLISTED)

Pablo Bernabeu

Manual speech transcription creates a major time and financial bottleneck in research, yet it is routinely treated as a minor preparatory chore and summarised in a single line of a methods section. This lack of transparency conceals technical and analytical decisions from readers. Off-the-shelf commercial chatbots restrict transcription to very short recordings, prohibit automated post-processing and obscure processing choices, which severely limits the traceability and reproducibility of output transcripts. To resolve these limitations while maintaining scientific rigour, this project established an openly licensed, fully documented transcription workflow that standardises execution and subjects the entire processing pipeline to inspection and audit. Data privacy forms a core arm of the framework, ensuring automated efficiency does not compromise ethical standards or institutional compliance.

Because uploading audio to commercial cloud or chatbot services risks breaching ethics approvals, participant expectations and institutional data handling, the workflow keeps raw audio entirely within institutional control. Deployment options depend on data sensitivity. Highly sensitive or identifiable recordings are processed locally on encrypted workstations to maintain strict data sovereignty. For large-scale datasets where institutional cluster policies and ethics approvals permit multi-tenant execution, the workflow scales to high-performance computing infrastructure.

The open repository contains source code, installation steps, cluster submission scripts, command-line parameters, sample outputs, licensing terms and citation metadata. The workflow can be tailored seamlessly by pointing to any transcription model hosted on the Hugging Face database, including OpenAI Whisper models. To address the technical constraints of off-the-shelf tools, the framework incorporates automated post-processing modules, including audio enhancement, repetition filtering, speaker diarisation, language-specific text cleaning and multilingual personal-name masking. Each processed recording generates a dedicated log file recording the exact model checkpoint, execution settings, hardware environment, applied post-processing parameters and validation metrics alongside the transcript. Standardising these execution parameters establishes process-level reproducibility, providing an auditable trail to accompany structured human review.

To evaluate performance systematically, the workflow undergoes ongoing empirical validation against established speech transcription benchmarks. Finally, because automated model outputs remain susceptible to errors, we must acknowledge the persistent necessity of human validation, especially to assess domain-specific

terminology, subtle conversational nuances and automated privacy masking prior to final analysis.

Outputs

Project description: <https://pablobernabeu.github.io/2025/speech-transcription-python/>

Repository:

https://github.com/pablobernabeu/secure_local_HPC_speech_transcription

Archived release: <https://doi.org/10.5281/zenodo.17624830>

Oxford e-Research Centre case study: <https://oerc.ox.ac.uk/ai-centre/ai-in-research-case-studies/secure-private-transcription-at-scale-with-whisper-and-github-copilot>

The Government Outcomes Lab – The INDIGO Impact Bond Dataset: making data on social impact bonds open, traceable and reusable (SHORTLISTED)

Juliana Outes Velarde, on behalf of the Government Outcomes Lab (GO Lab), Blavatnik School of Government

What open practices were used?

The Impact Bond Dataset was designed as an open resource following FAIR principles, making data Findable, Accessible, Interoperable and Reusable. The dataset is freely available in CSV and JSON formats, with each data point linked to its original source. Identifiers to connect projects with other public databases are used where possible, alongside documentation, data dictionaries and a GitHub repository explaining the dataset's structure and variables.

Why were these open research practices chosen?

Information about social impact bonds was historically scattered, inconsistent and difficult to access, limiting researchers', governments' and practitioners' ability to compare projects and learn from previous experience. The open practices therefore had two aims: to make the GO Lab's own research transparent and reproducible, and to create a practical evidence base that others could freely use for research, policy and practice.

How were these open research practices implemented?

The team collected and published information on social impact bonds, creating an evidence base now covering more than 320 projects across 42 countries. Data are provided in reusable formats with source links, standard identifiers and supporting documentation. Importantly, the dataset is developed collaboratively with organisations across the sector. The team acts as a **data steward**, working with partners to establish what information can be shared openly while respecting legitimate confidentiality concerns.

What challenges were encountered?

A major challenge has been data ownership. Information about these partnerships is often generated by multiple organisations with different expectations about publication. The team therefore needed to balance openness with confidentiality and organisational concerns. Information is only published when the organisations involved agree to its release, ensuring that openness is based on trust and collaboration.

What difference or impact did these open research practices make?

The project has transformed access to evidence in a previously fragmented field. Researchers, governments, investors and public-service organisations can now access a common evidence base to support research, programme design and policy decisions. Since 2023, the dataset pages have received more than 138,000 views. The resource has informed new programmes, helped organisations adapt successful approaches across countries and inspired further knowledge-sharing initiatives. Collaborative data stewardship has also strengthened relationships between researchers, governments and practitioners and encouraged a broader culture of openness and shared ownership of evidence.

* AI use statement: ChatGPT (OpenAI) was used to support editing, summarisation and refinement of text for clarity, accessibility and word limits. All substantive content was authored, reviewed and verified by the applicant, who takes full responsibility for the final text.

Yuan Tan - TEXTCOURT: Unlocking the Textual Worlds of Chinese Court Drama through Open Research Infrastructure (SHORTLISTED)

Tian Yuan Tan, Ming Tak Ted Hui, Ewan Macdonald, Cheuk Yee Wai

What open practices were used?

TEXTCOURT built an openly accessible digital archive for Chinese court theatre, bringing together dramatic texts, collections, entities, foreign records, and related visual material. The project publishes TEI-encoded digital editions, structured metadata, and linked data, with datasets preserved through Oxford Sustainable Digital Scholarship (SDS) and research outputs deposited in ORA. Open-access publications, workshops, public events, and a student practicum extend the infrastructure beyond the core database.

Why were these open research practices chosen?

The corpus contains more than 20,000 surviving texts, dispersed across collections and often treated as anonymous performers' texts. We chose open infrastructure so that researchers can discover, compare, and cite materials without reconstructing the corpus independently. FAIR principles, persistent identifiers and recognised digital standards make the resources findable, accessible, interoperable, and reusable, while open publication supports scrutiny, teaching, international collaboration, and future research beyond the original project team.

How were these open research practices implemented?

Literary scholars, digital humanists, librarians, SDS specialists, and the Faculty IT team developed the platform together. TEI encoding and structured metadata make editions machine-readable, while authority data and persistent identifiers link scripts to plays, collections, people, objects and places. Search and browse tools, visualisations, and cross-links make these relationships discoverable.

What challenges were encountered?

The project had to reconcile a large, heterogeneous corpus gathered from dispersed sources, including variant titles, incomplete texts and records in different languages and scholarly traditions. It also had to make analytical groupings useful for navigation without presenting them as definitive classifications. Keeping the platform sustainable and publicly accessible as the data continued to grow required iterative workflows, cross-disciplinary coordination, careful documentation, and ongoing support from SDS and the IT team.

What difference or impact did these open research practices make?

TEXTCOURT gives researchers, students, and the public a shared starting point for studying Chinese court theatre. The linked database supports questions about textual transmission, literary connections, people, objects, places, and cross-cultural encounters that are difficult to pursue from isolated catalogues. Open datasets, digital editions, and publications make the work inspectable and reusable. The platform's visual tools and workshops support teaching and engagement. The project's principal open research output is therefore a durable scholarly infrastructure that enables new research beyond the original project and its publications.

Clarke - Open tools and standards for reproducible neurochemical imaging (SHORTLISTED)

William Clarke

What open practices were used?

The project embedded open research in magnetic resonance spectroscopy (MRS) methods, software and standards. Its main outputs were FSL-MRS, an open-source Python toolbox for end-to-end analysis; the NIfTI-MRS community data standard; and tools for dynamic fitting, visualisation and data conversion. Code was developed in version-controlled repositories, released publicly, and archived with citable Zenodo DOIs and journal publications. Documentation, tutorials, practicals and educational material were also openly available, while NIfTI-MRS was developed through international collaboration.

Why were these open research practices chosen?

MRS workflows have historically been fragmented across proprietary formats and ad hoc pipelines, making studies difficult to reproduce, compare or reuse. Open software and a shared standard were chosen to reduce ambiguity, make methods inspectable and reusable, and connect spectroscopy with wider neuroimaging workflows. Open documentation and training were equally important so researchers could use and adapt sophisticated methods without direct support from the originating group.

How were these open research practices implemented?

FSL-MRS and associated tools were built as modular open software and tested on real datasets. Releases were versioned and archived, with documentation and tutorials published online. For NIfTI-MRS, an international expert group defined the standard and supporting software to enable reproducible pipelines across laboratories, scanner vendors and analysis packages. Tutorials, workshops, online resources and course materials supported community adoption.

What challenges were encountered?

MRS data are low-SNR, highly structured and acquired using many sequence variants, so robust open implementation required more than releasing code. We addressed this through modular design, testing on real data and iteration with Oxford users and external collaborators. Accessibility was another challenge, making clear documentation, tutorials and educational resources a core part of the project.

What difference or impact did these open research practices make?

The open ecosystem has improved reproducibility, transparency, accessibility and collaboration in neurochemical imaging. FSL-MRS has received more than 250 citations within five years, supported more than 15 internal studies, and contributed to five papers from external collaborations; its tools and standard are also used by other analysis packages. By enabling consistent data exchange and making methods inspectable and reusable, the project has lowered barriers to entry and supported a wider culture of reuse and collaboration in MRS.

Pit – Robustness in misinformation research

(SHORTLISTED)

Ilse Pit

What open practices were used?

I started this project following a journal call announced at the 2025 Society for the Improvement of Psychological Science (SIPS) conference, and I used open research practices at every stage. First, I used an openly available cross-national dataset including over 16,000 participants' sharing intentions of 41 pieces of news (including misinformation) regarding COVID-19. I categorized the news items into several categories, resulting in a new dataset that I subsequently shared, and transparently reported the categorization to allow others to check and repeat the process. I then operationalized congruence of the news in several ways, again transparently reporting each analytical decision, and sharing the R code to inspect and reproduce my analytical decisions using open source software. I used a multiverse approach, in which multiple analytical pathways are taken and reported, as opposed to just one. All material, data, and R and R Markdown code were shared with proper documentation on the Open Science Framework and assigned a DOI, and the article was published open access under a CC BY 4.0 licence in the Journal of Robustness Reports. Furthermore, the peer-review process was completely open and can be found with the published article.

Why were these open research practices chosen?

The transparent reporting of each step will allow other researchers to inspect, scrutinize and reproduce my methods, but I hope will also help others conduct similar research and analyses. Within misinformation research, a multiverse approach is relatively novel. Given the rise in research on misinformation, and with it many different ways to operationalise and analyse variables, I am hoping my approach can inspire a more transparent research culture with less flexibility in reporting only "successful" outcomes. The analysis framework I used furthermore separated people's ability to discriminate true from false news from their general tendency to share, allowing a more precise assessment of how accuracy prompts work. Finally, I pooled the posterior distributions to provide a clear summary of effect sizes across analytical specifications and to inform future misinformation research.

What challenges were encountered?

When it comes to challenges, the original analysis could not be computationally reproduced because it used licensed software unavailable to me, so I documented this limitation and developed a fully open-source alternative workflow. Defining the

boundaries of the multiverse was also challenging: I needed to distinguish plausible analytical choices from specifications that were theoretically or statistically unsuitable. The Journal's short report format created a further challenge for transparency, which I addressed by placing the full decision trail and supplementary analyses on the Open Science Framework.

What difference or impact did these open research practices make?

I think this project demonstrates the value of openly reusing published data: I was able to generate new variables and use novel analytical approaches to generate additional insights without further data collection. The method has informed other projects I am involved in, which will be presented at two upcoming conferences. More broadly, I hope it encourages researchers to report operationalisation and analysis choices more systematically, use open-source tools, reanalyse existing datasets, and treat robustness testing as a central part of credible misinformation research.

Jeong et al - A global geospatial dataset of renewable electricity supply and networks (SHORTLISTED)

Jihyeon Jeong, Dr. Fred Thomas, Prof. Jim W Hall

What open practices were used?

The project openly shared three connected research products: the manuscript, the global geospatial dataset, and the software used to construct and validate it. The dataset and code were released as versioned Zenodo archives with permanent DOIs. The live Python workflow is maintained on GitHub under the MIT License, and the manuscript was published as a preprint during peer review (Nature Scientific Data).

Why were these open research practices chosen?

The paper produces spatial estimates of renewable electricity supply and transmission infrastructure, rather than only presenting final statistics. Users need to understand how national electricity projections were translated into approximately 10 km spatial layers; how electricity use and supply were allocated to settlements; how existing and future generation facilities were mapped or modelled; and how settlements and facilities were connected through transmission networks. Sharing the data, code and modelling assumptions allows users to scrutinise these decisions, reproduce individual stages, substitute improved inputs and adapt the workflow to alternative scenarios.

How were these open research practices implemented?

The Zenodo dataset provides standardised outputs for 189 countries for 2024, 2030 and 2050. These include generation-facility locations, settlement-level electricity use and supply, transmission networks, national summary tables, and solar, wind and hydropower viability layers. Outputs are provided in reusable Parquet, GeoTIFF and Excel formats. The GitHub repository documents the complete Python workflow, from input preparation to validation and visualisation. Reproducibility is supported by a step-by-step README, a specified Conda environment, centralised model parameters and support for local or SLURM-based execution.

What challenges were encountered?

No single open source provides complete and consistent information on electricity generation, demand and transmission infrastructure across 189 countries and three time horizons. Input datasets differed in coverage, resolution, reference year, classification and quality. Future facilities also required spatial modelling because their precise locations are unknown. These challenges were addressed through systematic harmonisation, explicit modelling assumptions, automated workflows and technical

validation. Source-data licensing restrictions also mean that some inputs must be obtained separately, even though the resulting outputs and code are openly available.

What difference or impact did these open research practices make?

The open outputs make the assumptions and processing behind the maps inspectable, enabling independent checking, reuse and methodological improvement. Users can rerun selected countries or workflow stages and incorporate better local information. The dataset can also be combined with climate-hazard, population, land-use or investment data. For example, the Singapore case study overlays the electricity layers with climate hazards to identify potentially exposed infrastructure, demonstrating how an open global dataset can support locally relevant energy-resilience analysis.

Didi - FLIP2 open benchmarks for protein engineering (SHORTLISTED)

Kieran Didi

What open practices were used?

FLIP2 provides shared tests for comparing machine-learning methods that predict how changes to a protein's sequence affect its function. We openly released seven experimental datasets, 16 predefined evaluation splits, data-processing and evaluation code, baseline results and documentation. Dataset records preserve provenance, attribution and reuse licences. A public website supports access and reuse, while a Zenodo archive provides a persistent, citable record.

Why were these open research practices chosen?

Protein engineering needs reliable comparisons between predictive methods. When data preparation or testing procedures are opaque, apparent improvements can reflect different evaluation choices rather than better models. We chose to share the full workflow so researchers could inspect those choices, reproduce comparisons and evaluate new methods on identical tests without independently rebuilding the benchmark.

How were these open research practices implemented?

We harmonised experimental data from different groups into standard CSV and FASTA formats, retaining dataset-specific documentation. We defined fixed training, validation and test sets that reflect practical challenges, such as predicting mutations at unseen positions or transferring predictions to a different starting protein. We published processing and evaluation workflows on GitHub, alongside baseline implementations and results. The project website brings together dataset descriptions, loading examples and downloads. Dataset-specific licences include CC BY 4.0 and MIT terms.

What challenges were encountered?

The main challenges were reconciling heterogeneous measurements and source formats, preserving biological context, and ensuring appropriate attribution and redistribution terms. Designing realistic tests also required going beyond random data partitions. We addressed these issues through consistent schemas, explicit provenance and licensing records, documented split definitions and fixed evaluation sets.

What difference or impact did these open research practices make?

The release makes the complete evaluation process inspectable and reusable, lowering the practical barrier to reproducing results and comparing new methods. FLIP2 showed that realistic evaluation tasks are harder than random splits and that simple statistical models can match or outperform fine-tuned protein language models on several tasks.

Sharing these baselines makes limitations visible as well as successes, supporting more credible claims about progress. Researchers can reuse and extend the curated resources while retaining credit to the original data creators.

[Project website](#) | [Code](#) | [Data archive](#) | [Paper and project authors](#)

Juraskova - mlp-train package: automated active learning for reactive machine learning interatomic potentials

Veronika Juraskova

What open practices were used?

We develop and maintain open-source code to train machine learning interatomic potentials (MLIPs) to model chemical reactions in the gas phase and in solution. The code is publicly available on GitHub [<https://github.com/duartegroup/mlp-train>] with examples, documentation, and tutorials. We published research using the software as open-access articles and preprints, with datasets deposited on the Oxford Research Archive and figshare.

Why were these open research practices chosen?

We chose these practices to turn the project-specific research code into a reusable tool that other researchers could understand, modify, and contribute to. Tutorials and documentation further lower barriers to adopting the code. This accessibility increases the reproducibility of our results, facilitates validation, and allows others to adapt the workflows and code to their own research.

How were these open research practices implemented?

We used GitHub, ChemRxiv and the Oxford Research Archive to make the data accessible and archived beyond the lifetime of existing projects and funding. I introduced the Research Software Engineering (RSE) practices, including testing, formatting and continuous integration. I also organised group hackathons and delivered presentations on RSE to help researchers learn these practices and contribute to the code.

What challenges were encountered?

One of the main challenges has been helping students feel more confident sharing and developing code and explaining the importance of good research practices and code review. With the emergence of coding agents and generative AI, RSE practices are becoming even more important. It is now much easier for users to generate large amounts of code; however, careful testing, clear documentation, change tracking, and high-quality review are still needed to ensure the research is transparent, reproducible, and adaptable by others. Hackathons, tutorials and collaborative code review provided a supportive environment in which students could develop these skills by contributing to a real research tool.

What difference or impact did these open research practices make?

These open research practices lowered barriers to training and using MLIPs in research. For example, groups beyond the Duarte group now use the code to build their own training datasets and extend the applicability of MLIPs. In the Duarte group, the materials help incoming students kick-start their projects and learn to contribute to the collaborative codebase. The published datasets and preprints also serve as a blueprint to continue to incorporate open research practices in the group.

Gladstone - Making Oxford's water security research openly available via ORA and Reshare

Nancy Gladstone

What open practices were used?

Research publications, reports, working papers, datasets, and supporting materials from the REACH Water Security Programme were made openly available through trusted repositories, principally the [Oxford Research Archive \(ORA\)](#) and, for social science survey data, the UK Data Service ReShare repository. The project applied FAIR (Findable, Accessible, Interoperable and Reusable) principles by creating high-quality metadata, linking related outputs, and providing documentation to support discovery and reuse.

Why were these open research practices chosen?

The goal of the project was to increase the accessibility, transparency, and long-term value of research funded through the REACH programme. As the programme focused on addressing water security and climate resilience challenges in low-income countries, it was important that researchers, policymakers, practitioners, and partner organisations could access outputs without subscription barriers. Open access and open data approaches also aligned with the programme's commitment to ensuring that publicly funded research was openly available to support evidence-based decision-making and encourage wider reuse of data and findings.

How were these open research practices implemented?

The creation of the REACH Water Security Collection on ORA brought together more than 224 research outputs generated across multiple countries and institutions. This involved identifying and gathering relevant publications, datasets, and supporting materials from projects throughout the programme; obtaining accompanying documentation; and creating consistent, high-quality metadata. Close partnership with colleagues from ORA, alongside support from a library work-experience student, helped ensure records were thoroughly checked, uploaded, and preserved for long-term access.

What challenges were encountered?

REACH research outputs were produced across multiple countries, institutions, and disciplines. To ensure consistency, completeness, and compliance with repository requirements, preparing datasets and metadata for open dissemination required substantial effort to coordinate among a large number of contributors. Researchers and

partners needed to be consulted to confirm methodologies, variables, permissions, and documentation, often many years after data had originally been collected.

What difference or impact did these open research practices make?

The project significantly improved the accessibility, visibility, and long-term preservation of REACH research, including contributions from international partners. By removing barriers to access and preserving outputs for future use, the collection has increased the reach, influence, and practical value of REACH research while creating a sustainable legacy for the programme beyond its formal lifetime.

Straub - Phenofhy: open, reusable tooling for phenotype data in Our Future Health

Vincent Straub

Phenofhy is an open-source Python package developed to make phenotypic analysis within the Our Future Health (OFH) Trusted Research Environment (TRE) more usable, reproducible and accessible. The project responds to a practical tension in modern health-data research: TREs provide essential security and privacy protections, but can also make the exploratory, iterative work of data analysis more difficult and costly.

What open practices were used?

Phenofhy adopts open-source software development as its core practice. The package is publicly available in beta through GitHub, with its source code, documentation, quick-start guide, tutorials, key concepts and API reference openly accessible. Development is community-oriented: users can provide feedback, report bugs and request features through an embedded discussion channel. The package is also designed to be versioned, maintained and will soon be distributed through PyPI.

Why were these open research practices chosen?

These practices were chosen to address the practical barriers that researchers can encounter when working within TREs. While secure environments protect sensitive health data, researchers may face delays in launching analyses, difficulty importing or testing custom software, uncertainty around compute costs and a steep learning curve when moving into a new environment. These constraints can hinder exploratory trial-and-error work, which is an important part of detecting data problems, developing analyses and enabling discovery. Open, reusable tooling provides a way to address these barriers while retaining the security of the TRE. Phenofhy specifically aims to reduce programming and technical overhead, support local prototyping before cloud execution, improve reproducibility, and lower barriers for early-career and other researchers using TRE-based research for the first time.

How were these open research practices implemented?

Phenofhy was designed around four principles: TRE-ready, so it can be deployed directly in the cloud; available locally, allowing researchers to test code without incurring cloud compute costs; community-oriented, enabling feedback and contribution; and easy to install, version and maintain. The package provides tools for efficient phenotype preprocessing and quality control, including automated phenotype profiling before downstream analyses. It also supports local simulation of OFH-structured data, allowing researchers to prototype and test pipelines before accessing phenotypic data or committing cloud resources. The current beta release can be imported into an

approved OFH TRE project through the Airlock process, while the package is being developed further in response to community needs.

What challenges were encountered?

The central challenge was designing software that supports open, flexible research practices within a controlled and governed environment. TREs necessarily restrict how sensitive data and software are accessed, while researchers still need the ability to experiment, make mistakes, test pipelines and iterate efficiently. The project therefore had to balance openness and usability with the governance, security and computational constraints of OFH. A further challenge was recognising that compute in a TRE can carry direct financial costs, making inefficient experimentation particularly burdensome.

What difference or impact did these open research practices make?

Phenofhy provides a reusable research capability that helps make OFH phenotypic analysis more accessible, transparent and reproducible. By sharing source code, documentation and reusable workflows, it reduces reliance on bespoke scripts and individual knowledge. Local prototyping can reduce unnecessary cloud execution and associated compute costs, while automated preprocessing and quality-control functionality can support more consistent analytical practice. The community-development model also creates a mechanism for users to contribute feedback and shape future functionality as the OFH research community grows. Overall, Phenofhy demonstrates how open research software can complement secure data infrastructure: sensitive data remain protected within the TRE, while the tools, documentation and knowledge needed to work effectively with those data can be shared and improved collectively.

Key reference

Straub, V.J., Mills, M.C. Enabling secure discovery in trusted research environments with improved tooling. *Nat Med* **32**, 2712–2715 (2026). <https://doi.org/10.1038/s41591-026-04401-x>

Aulicino - Building and Sustaining Shared Infection Biology Capacity at the TIDU

Anna Aulicino

Following several years working in the CL3 facility, I took on the role of High Scientific Coordinator after the departure of the researchers who had previously supported its scientific operation. I have since helped maintain the facility, its equipment, training and compliance, while developing shared microbiology infrastructure within the TIDU for Level 1 and Level 2 bacterial research.

What open practices were used and why were they chosen?

My approach has been to share infrastructure, equipment, expertise, protocols and training rather than restrict resources to individual research groups. Infection biology requires specialist facilities and equipment that can be difficult and expensive to establish independently. Shared resources therefore improve access, reduce duplication and allow researchers without established microbiology facilities to incorporate infection biology into their projects.

Within the CL3, I coordinate the maintenance and scientific use of shared equipment, including the Chromium 10x system and Sony MA900, which is also available to external users. Within the TIDU, I helped establish a shared bacteriology laboratory, currently used by two groups and supporting many researchers and students over the past five years. Salmonella SOPs, training records and competency assessments are maintained in a shared repository accessible to CL3 users.

How were these open research practices implemented? What challenges were encountered?

As High Scientific Coordinator, I work closely with Health and Safety and Biological Safety Officers to maintain the CL3 and coordinate equipment maintenance, user support, training and documentation. A key challenge was maintaining continuity after the departure of the previous CL3 researchers while developing microbiology capacity within a predominantly translational research environment.

The shared bacteriology infrastructure has also enabled new research. In particular, I coordinated the installation of an anaerobic cabinet in the only suitable location within the Unit, enabling a new *Helicobacter pylori* research project.

What difference or impact did these open research practices make?

The infrastructure has enabled many researchers and students to access infection biology facilities, while supporting both internal and external users. It has expanded the

Unit's research capacity beyond *Salmonella* and enabled new projects that would otherwise have required external infrastructure.

The facilities have also become an important training environment for early-career researchers. Two Master's students completed *S. Typhi* projects in the CL3; an Engineering PhD student conducted *S. Typhi* research that contributed to a collaborative publication; and an Immunology Master's student worked on *S. Typhimurium*, subsequently progressing to a PhD position in Dublin.

Overall, this work has created a more accessible and collaborative infection biology environment, where infrastructure, expertise and knowledge can be shared while enabling new research and training the next generation of infection biologists.

Alkire et al - Oxford Poverty and Human Development Initiative (OPHI) – Sharing the global levels and trends of multidimensional poverty as a global public good

Submitted by the Oxford Poverty and Human Development Initiative (OPHI)

Sabina Alkire, Rakesh Mishra, Lhachi Selden, Maya Evans, Juliana Yael Milovich, Corinne Mitchell, Ellie Duncan and Maarit Kivilo.

What open practices were used?

Since 2010, OPHI has shared the global Multidimensional Poverty Index (MPI) as an open-access global public good. It publishes open data tables, analytical and country briefings, methodological documentation and computational files. Since 2018, annual Stata do-files have enabled users to reproduce published tables. An interactive databank and MPI toolbox support access and reuse. Database and replication materials are licensed under CC BY 4.0.

Why were these open research practices chosen?

These practices make internationally comparable evidence on multidimensional poverty accessible, transparent, reproducible and reusable. They allow researchers, governments, NGOs, civil society organisations, journalists and businesses to scrutinise the evidence, reproduce results, conduct further analysis and build on the research. Openness also extends the reach and public value of the global MPI in policy and international monitoring.

How were these open research practices implemented?

OPHI releases the global MPI annually with open-access estimates, data tables, reports, country briefings and methodological documentation. Computational and data-preparation do-files show how results are generated and how harmonised indicators can be produced from household-survey microdata. The interactive databank provides harmonised trends and disaggregated analysis, while the MPI toolbox supports replication and further programming. Outputs are also shared through ORA, UNDP, the World Bank, Humanitarian Data Exchange, Our World in Data and Worldmapper.

What challenges were encountered?

The underlying household surveys are owned by multiple providers, so OPHI must balance transparency with licensing and access conditions. In 2025, funding for the Demographic and Health Surveys programme was threatened; it provides data covering around 70% of people living in multidimensional poverty (763 million people),

potentially affecting the global coverage of the global MPI. Short-term funding has since been secured.

What difference or impact did these open research practices make?

The global MPI is an officially recognised poverty statistic with a specific target in Sustainable Development Goal 1 and is used to monitor the Third UN Decade to Eradicate Poverty (2018-2027). It is shared through major open-access platforms and supports collaboration with UNDP and others. In 2025, the global MPI report was downloaded around 21,000 times from the UNDP website and 2,000 times from the OPHI website, and was cited in 314 international media outlets within two weeks. Academic engagement averages around 80 publications a year, and the methodology has inspired tailored national MPIs in around 50 countries.

Bernabeu - Reproducible and expandable modular workflows for the creation of FAIR experimental stimuli in psycholinguistics

Pablo Bernabeu

The Linguistic Economy through transfer Source Selectivity (LESS) project is a longitudinal study of third language acquisition led by Dr Jorge González Alonso at UiT The Arctic University of Norway. I led the preparation of the experimental materials, which required stimuli for six counterbalanced sessions across multiple languages and research sites. To ensure full reproducibility, every item was generated programmatically using free, open-source software. The resulting materials are presented via OpenSesame, maintained under version control with public release notes, and published on OSF and GitHub.

Large-scale experimental stimuli are usually assembled manually or via unreleased, single-use scripts, contributing to low reproducibility rates in experimental linguistics. Unreleased materials cannot be inspected, corrected or extended by external researchers or original authors in later years, allowing methodological errors to persist unnoticed. Programmatic generation resolves this by creating an auditable path from raw linguistic inventories to the precise sequences presented to participants.

The generation framework uses a modular architecture, combining atomic components into complex stimuli. Language-specific elements reside in site-specific base files. Scripts dynamically construct items from these files, counterbalancing gender and number across conditions, rotating words across properties and sessions, and controlling word frequency against reference corpora. Automated validation scripts verify output distributions, alerting researchers to departures from target balances. Traceable code assigns participant-specific parameters, while a custom Python integration within OpenSesame time-locks electroencephalographic recordings to stimulus onset.

To maintain flexibility across diverse languages and locations, core generation logic was decoupled from language-specific data. New research sites can adopt the processing workflow simply by substituting their base language files.

This transparent structure enables methodological errors to be identified and rectified in public view. For example, release 1.1.0 corrected an electroencephalographic trigger fault that had rendered two experimental conditions indistinguishable, a modification impossible to audit within static file folders. By construction, the materials operationalise FAIR criteria: findable and accessible in public repositories, interoperable through open-source plain text and reusable through modular design.

Serving as an adaptable template for other laboratories, this framework ensures publicly funded outputs remain cumulative scientific assets. I am grateful to Dr Jorge González Alonso for supporting these open science proposals.

Outputs

Materials and R project: <https://osf.io/974k8>

Code: https://github.com/pablobernabeu/materials_L3_transfer_longitudinal

Poster: <https://osf.io/h83yq>

Audio overview: <https://www.youtube.com/watch?v=NG9G1gQdOEo>

González Alonso, J., Bernabeu, P., Silva, G., DeLuca, V., Poch, C., Ivanova, I., & Rothman, J. (2025). Starting from the very beginning: Unraveling third language (L3) development with longitudinal data from artificial language learning and EEG. *International Journal of Multilingualism*, 22(1), 119–142. <https://doi.org/10.1080/14790718.2024.2415993>

Hiari - Navigating UK Ethics Policy in Overseas Fieldwork

Sandra Hiari

The entry is an open-access paper, titled Ethical mediation: Navigating research ethics and frictions in overseas fieldwork by Hiari and Razem. It examines the application of UK-based research ethics policy for overseas fieldwork.

What open practices were used?

The open practice used was (gold) open-access publishing in a top-tier journal in geography for Royal Geographical Society (with the Institute of British Geographers).

Why were these open research practices chosen?

This publishing route was chosen so the research can reach the broadest audience possible without posing any limitations on researchers who cannot access subscription to the journal.

How were these open research practices implemented?

This research made use of the University of Oxford's publisher deal with Wiley, and the research was published under the gold full open access.

What challenges were encountered?

The first author coincidentally learnt about the open access options available through the University of Oxford and publishers. Luckily, she had attended the Bodleian training course Fundamentals of Open Access to understand the types of open access options available.

What difference or impact did these open research practices make?

This research improvement engagement with ethics issues. Upon publishing the paper as open access, there were three immediate impacts. First, multiple opportunities to engage in academic debates emerged. The first author, who is a student at the University of Oxford, was contacted to review two papers, one specifically focusing on ethics, for the International Development Planning Review and Urban Studies, another journal in her field, Urban Studies. Second, the first author was contacted by the UK Research Integrity Office (UKRIO) to participate in a roundtable on ethical issues in international and interdisciplinary collaborations. Third, the paper was circulated among members of an international working group on policy mobilities within urban geography wherein the first author presented the paper and shared experiences.

Josan - Open eye research

Amandeep Singh Josan

What open practices were used?

My role as a research Optometrist involved in clinical trials has led me to develop several software tools to facilitate and enhance research practices in Ophthalmology. The tools became so useful within our lab group that I published them open-source on Github and developed freely available downloadable programs for other clinicians and researchers to benefit from.

Why were these open research practices chosen?

Front line eye research has for decades been hampered by proprietary software developed by imaging device manufacturers. There is significant expertise in Ophthalmology research but an inability to perform deep research due to these proprietary tools.

How were these open research practices implemented?

Software was initially developed for personal and lab research use. The tools became useful for many clinicians and, in particular, for small medical student (FHS) research projects. I decided to turn them into standalone programs freely available via my website and Github.

What challenges were encountered?

The main challenges initially were around being able to open and interpret image files and make the process smooth and reproducible. Once correctly reading image data was resolved, the remaining challenges were around making the software accessible to all. This involved learning about Github repositories and, more challenging, converting my code into standalone locally run programs. Patient data protection was paramount, so running things from the cloud was not possible, making locally run programs essential.

What difference or impact did these open research practices make?

Once developed, the software quickly became a core resource for research projects analysing patient retinal imagery and visual field results. The programs have also begun to gain traction in the wider community and have been used in top line results for several clinical trials.

More widely, significant challenges surround outcome measure selection in retinal gene therapy and other novel therapy clinical trials. Many historical trials (including by our group) have failed due to insensitive endpoints rather than lack of efficacy. However, the

FDA and EMA are understandably hesitant to approve novel endpoints that are not well established or shown to provide direct patient benefit. This has made outcome measures one of the biggest issues in retinal clinical trials and a widely published topic in Ophthalmology.

The open-source tools I have developed provide a crucial analytical tool for examining the link between retinal structure (OCT imaging) and retinal function (retinal sensitivity via perimetry). This is helping to validate more novel and sensitive outcome measures which should form the basis of primary endpoints in future clinical trials.

Thomas - GLASS: Geochemical Laboratory Approaches for Sharing Science

Richard Thomas

GLASS is an open-science initiative to make experimental geochemistry more accessible, reproducible and clear.

What open practices were used?

The idea is to share as much of the experimental record as we reasonably can: what we did, how we did it, the data we produced, how those data were analysed, and the physical samples that remain afterwards. This should make it easier for other researchers to check each other's results, compare them with their own, or use the work as a starting point for new research. At the centre is an open repository containing experimental methods and procedures, raw and processed data, analytical methods, instrument and calibration information, data-processing workflows, sample information, and supporting documentation. These are deposited through platforms such as ORA and GitHub. We also want the repository to be accessible to researchers with different levels of computing experience, so wherever possible we provide files in familiar formats such as Word and Excel, rather than requiring specialist software or coding skills to access the work. READMEs and supporting notes explain what the files contain, how the experiments were carried out and how the data were processed.

Why were these open research practices chosen?

In experimental geochemistry, sharing data is only part of the problem. Results often depend on the actual materials used: geological samples, starting materials, standards and the glasses produced during experiments. Where possible, we therefore catalogue and retain our experimental products and make them available to other researchers. Someone could, for example, analyse one of our glasses using a different technique, compare measurements between laboratories, or independently check a published result using the original material. This is important because reproducing an experiment from scratch can be expensive and time-consuming, and sometimes the original material simply cannot be recreated. Not every researcher has the same access to experimental facilities, analytical equipment or funding. Sharing existing samples and experimental records can make it possible for researchers to test or build on previous work without first having to reproduce the entire study.

What challenges were encountered?

There are practical difficulties. Preparing data so that somebody outside the lab can actually understand and use it takes time. Samples need to be labelled, catalogued and stored, and some are too small, valuable or sensitive to be freely circulated. There are

also costs associated with maintaining an archive and sending materials between laboratories. This work is important, but it is not always recognised in the same way as publications or new experimental results.

What difference or impact did these open research practices make?

We would also like GLASS to grow beyond our own laboratory. Other experimental geochemistry groups are encouraged to take part and share their experiments, methods, data and samples in the same spirit. If more laboratories do this, it should become much easier to compare results between groups and work out whether differences come from the experiments themselves, the analytical methods, or the way the data have been processed. It could also save time by making existing experiments and materials easier to find and reuse.

One of the most useful outcomes of GLASS so far has therefore been a fairly simple change in how we think about an experiment. Recording methods properly, organising data and looking after samples are part of doing the research, rather than things to sort out once the paper has been written. Ultimately, GLASS is about making it easier for someone else to see exactly what we did, access what we produced, check our results and do something new with them. The hope is that, by encouraging other laboratories to do the same, experimental geochemistry can become easier to reproduce, compare and build upon.

Abusamra - Systematic Review: Circulating tumour cells (CTCs) as a window into lethality in prostate cancer

Sophia Abusamra

What open practices were used?

The review combined four open research practices: prospective preregistration of the protocol on PROSPERO; transparent, PRISMA-based reporting of the full search and study-selection methods; publication of the extracted evidence as structured data tables; and open-access release of the accepted manuscript under a CC BY licence in the Oxford University Research Archive (ORA).

Why were these open research practices chosen?

Preregistration commits a review team to its question and methods in advance, reducing the risk of selective reporting and making our intentions publicly inspectable. Transparent methods allows others to scrutinise, reproduce and update the work. Sharing the extracted data, allows others to use the review as a resource to build on. Open access through ORA removes the paywall so clinicians and researchers anywhere can use the findings.

How were these open research practices implemented?

The protocol was registered before screening began. An information specialist co-author designed and ran reproducible searches across four databases (MEDLINE, Embase, the Cochrane Central Register and Web of Science); we report the complete search strings, database coverage, the number of records at every stage, links to the saved searches, and a PRISMA flow diagram documenting screening from 16,006 records to the included studies. Records were screened in Covidence by two independent reviewers with conflicts resolved by consensus, and the extracted evidence was published as structured supplementary tables. Because the journal is subscription-based, we used the University's rights-retention route to deposit the accepted manuscript openly in ORA under CC BY.

What challenges were encountered?

The CTC literature is highly heterogeneous as terminology, isolation platforms and reporting standards vary widely, making a defensible, repeatable search was demanding to construct. Achieving open access depended on the green rights-retention route rather than the published version as the target journal was pay-walled.

What difference or impact did these open research practices make?

Preregistration and method reporting make the review reproducible and updatable, as others can re-run or extend the search. Data sharing makes the knowledge accessible for future research. Open deposit under CC BY widens access, permits reuse, and supports compliance with our funders' open-access policies. Working with an information specialist brought reproducible-search expertise. These practices model transparent evidence synthesis, strengthening trust in the findings and lowering the barrier for others to reuse and extend them.

Project outputs: Review article, <https://doi.org/10.1038/s41585-025-01121-8>; open-access manuscript (CC BY 4.0), <https://ora.ox.ac.uk/objects/uuid:ab9f5d43-3d68-4128-80db-6fcd8c19b862>; preregistration, PROSPERO CRD42023438337.

Kariyawasam - Open Data for Better Heat-Risk Decisions: Understanding How Cities Experience and Respond to Extreme Heat

Nethmi Jayaratne Kariyawasam

What open practices were used?

This project uses open global data and transparent analytical methods to develop heat-risk assessments for comparison and reuse globally. Data sources include ERA5 and ERA5-Land climate reanalysis, WorldPop population data, green and blue infrastructure datasets, and open socioeconomic indicators. Processing steps, assumptions and analytical decisions are documented to support reproducibility.

Why were these open research practices chosen?

Extreme heat is a growing global challenge, but access to data and modelling resources for cities across the globe remains highly uneven. Many assessments rely on proprietary datasets or locally specific methods that are difficult to reproduce. Open data and transparent methods reduce these barriers, enable scrutiny, and make the framework accessible to researchers, policymakers and practitioners.

How were these open research practices implemented?

The research uses globally available datasets and standardised processing methods. Choices about variable selection, harmonisation, normalisation and aggregation are documented throughout the workflow. This makes the analytical pathway visible, allowing the framework to be replicated, updated with new data and adapted across cities, countries and regions.

What challenges were encountered?

A key challenge was combining datasets with different spatial resolutions, temporal coverage and reporting scales. Some indicators, such as electricity prices, are only available at a national level, while environmental variables may be much finer. Careful harmonisation and documentation of these limitations were essential for transparency, comparability and interpretability.

What difference or impact did these open research practices make?

Open research strengthened accessibility, credibility and reproducibility by enabling others to inspect, reuse and adapt the framework. Standardised methods support comparison and future reproducibility. The work also led to broader conversations on heat risk on UN Disaster Risk Reduction (UNDRR) platform, widening policy and practitioner engagement. Publication: Kariyawasam et al. (2026), Sustainable Cities and Society, Moving beyond exposure: a globally comparable framework for heat risk assessment in cities. <https://doi.org/10.1016/j.scs.2026.107535>

Bernabeu - Advocacy for open data governance in industry-funded research: Amending data agreements

Pablo Bernabeu

Upon joining the Learning in Families through Technology project, I reviewed its data agreement, a joint controllership contract with a corporate funder. The contract permitted the publication of analyses but did not extend to the public sharing of raw research outputs. Because this restriction governed the entire project, study-by-study workarounds were impossible. Readers require raw outputs to verify analyses or reproduce results. Whilst industry partnerships grant access to large-scale data, underlying contracts are rarely drafted with reproducibility in mind, leaving science to bear the technical debt later.

To resolve this issue, I proposed amending the core contract so that anonymised datasets and analysis code could be released for all subsequent studies. Commercial partnerships typically default to caution regarding data sharing. I addressed these concerns by demonstrating how anonymisation protects users' privacy and stakeholders' interests whilst fulfilling academic commitments to open research. Submitted to the principal investigators, led by Professor Victoria Murphy, the proposal demonstrated how open data would enhance research credibility, offer long-term project visibility and provide reputational value to the funder. To facilitate legal review, the proposal outlined concrete terms, including examples of data repositories and licences.

With the support of project leadership, the amendment was formally submitted. The review lasted seven months, during which colleagues updated data protection impact assessments. Sustaining progress across this period required persistent management to keep the item active on the review agenda.

The review concluded in formal approval to amend the original contract, establishing a project-wide open data framework. Anonymised datasets and reproducibility scripts may be deposited publicly. This outcome demonstrates that industry funding need not preclude open science, an argument further detailed in a public blog post as a model for similar barriers. I am grateful to Professor Victoria Murphy, the project leads and Research Services for supporting this proposal and advancing it with the funder.

Outputs

Amended data agreement between the University of Oxford and the corporate funder, placing the project under an open data framework.

Bernabeu, P. (2026). Beyond the paper: Why open science is the new research standard. <https://pablobernabeu.github.io/2026/open-science-new-research-standard/>

Mutunayagam - Understanding how two common respiratory infections interact in the nose of healthy adults: Respiratory Syncytial Virus and Streptococcus pneumoniae

Sanjita Brito Mutunayagam

What open practices were used?

Public and patient involvement (PPI) in protocol and document development; prospective trial registration (ISRCTN); open-access publication of a detailed protocol paper (PLOS One); open sharing of supporting study documents (participant information leaflet, informed consent form, safety and education leaflets); transparent documentation of protocol deviations; and planned dissemination of findings back to participants.

Why were these open research practices chosen?

PPI was chosen to keep the research relevant, inclusive and responsive to participants' and the public's needs. Prospective registration and open publication of the protocol were chosen to create a transparent public record, reduce the risk of selective reporting, and strengthen research integrity and reproducibility. Sharing supporting documents was chosen to help other researchers understand, appraise and potentially replicate the work.

How were these open research practices implemented?

PPI directly shaped study design, notably the decision to allow home self-isolation rather than in-patient isolation. The study was registered on ISRCTN before starting. A full protocol paper covering rationale, design, procedures, outcomes and statistical analysis was published open-access, with key participant-facing documents shared alongside it. Analyses and reporting followed the pre-specified protocol, with any deviations documented systematically. Findings will also be shared directly with participants.

What challenges were encountered?

Preparing documents for public sharing took extra time and effort to ensure clarity, completeness, and compliance with governance and confidentiality requirements. Balancing transparency with participant privacy and data protection required careful review before materials were released.

What difference or impact did these open research practices make?

PPI-informed design (particularly home self-isolation) improved recruitment to a hard-to-recruit challenge study, enabling completion within one year. Open registration and publication increased transparency and confidence in the findings while reducing risk of selective reporting. The protocol paper's reach (1,087 views, 714 downloads, 3 citations) demonstrates real engagement, supporting replication and future collaborative work, and improved accessibility for the wider research community.

Tao - Incidence, Prevalence and Mortality of Maternal Cardiac Disease Worldwide

Ling Tao

This systematic review and meta-analysis, published as an immediate Open Access article in *The Lancet Obstetrics, Gynaecology & Women's Health* (April 2026), provides the first global synthesis comparing maternal cardiac disease morbidity and mortality between high-income countries (HICs) and low- and middle-income countries (LMICs).¹ Drawing on 114 studies, it found that maternal cardiac disease prevalence in LMICs is roughly four times that in HICs, with different disease profiles between settings: arrhythmias were most common in HICs, and rheumatic heart disease was most common in LMICs. The review also exposed major evidence gaps, particularly for LMICs.

What open practices were used?

Open research principles ran through the whole project, from design to conduct to dissemination, to maximise transparency, reproducibility and accessibility. The review protocol was prospectively registered on PROSPERO (CRD42023487679) before completion of the review, reducing reporting bias and ensuring the study followed a pre-specified plan.² The review was conducted and reported according to the PRISMA 2020 guideline. The final article was published immediate Open Access under a CC BY 4.0 licence. This was chosen because maternal cardiac disease is a global health issue with a substantial burden in LMICs, where many institutions cannot afford subscription journals. Open access means clinicians, researchers and policymakers everywhere can read and use the findings.

How were these open research practices implemented?

accessible supplementary materials document the complete search strategy for every database, study selection process, statistical methods, subgroup analyses, certainty assessments, and detailed incidence, prevalence and mortality estimates for every disease subtype and included study. This lets readers examine every stage of the review and reproduce or extend the analyses. Funding sources, author contributions and protocol registration were openly reported, and the underlying extracted data are available upon reasonable request.

What challenges were encountered?

Funding open access publication was not a challenge for this project because the University of Oxford's read-and-publish agreement covered the journal in full, removing what is often the biggest practical barrier to open access publishing. The real challenge was balancing how much data to present in the main text against keeping the article

concise and readable. This was addressed by moving lengthy or secondary data, such as full search strategies, study characteristics, risk-of-bias assessments and detailed subtype estimates, into the freely accessible supplementary materials, and by stating clearly in the article that the underlying extracted data are available from the corresponding author upon reasonable request.

What difference or impact did these open research practices make?

These open practices have improved the quality, transparency and reach of the research. The open-access publication and supplementary materials provide a reusable resource for future systematic reviews, guideline development and health economic evaluations, while transparent reporting allows readers to critically appraise and reproduce the work. Unrestricted access helps get this evidence to the countries where maternal cardiac disease carries the greatest burden. It directly supports the paper's calls for standardised disease coding, more LMIC-based research, and cost-effectiveness evaluation of cardiac screening in pregnancy. By making both the findings and the methodology openly available, this project shows how open research can strengthen scientific integrity, speed up knowledge translation and increase the global impact of maternal health research.

References

1. Tao L, Ramaiya A, Ramakrishnan R et al.

Incidence, prevalence, and mortality of maternal cardiac disease: a systematic review and meta-analysis

The Lancet Obstetrics, Gynaecology, & Women's Health, 2026; 2, e298-e309

2. Tao, L · Ramaiya, A · Ramakrishnan, R · et al.

The incidence, prevalence, and mortality of maternal cardiac disease: a systematic review and meta-analysis

https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023487679

Deseilligny - Liquid-liquid phase separation in nanoconfined geometries

Lauriane Pierrot Deseilligny

What open practices were used?

Open research practices were integrated throughout the project, from the use of open-source software to the dissemination of the results. The complete analysis workflow and its dependencies are documented alongside the openly accessible arXiv manuscript. The developed code and representative experimental datasets are publicly shared through GitHub and permanently archived in the Oxford Research Archive with a DOI.

Why were these open research practices chosen?

I consider that research progresses faster when methods, code, and data are openly available. My own work depends extensively on open-source software, so releasing my own contributions publicly is scientifically efficient and ethically consistent. I deliberately deposited the material on both GitHub and ORA. GitHub maximises visibility, usability, and collaborative development. ORA guarantees long-term preservation, citation, and institutional archiving.

How were these open research practices implemented?

Open research was built into the project from the outset. The analysis pipeline uses only open-source software. I work exclusively under the free, open-source operating system **Linux (Ubuntu)**. Image processing relies on the free, **open-source MMVII C++** library together with **Python**. All software dependencies are explicitly documented and linked in the manuscript (currently on **arXiv** <https://arxiv.org/abs/2606.24809>, under review). The code I developed and test datasets made from my experiments are all openly available on **GitHub** (<https://github.com/cyber-Lolo/InterfeRIOT>) and in the **Oxford Research Archive** with a permanent DOI (<https://ora.ox.ac.uk/objects/uuid:dff54cd1-7779-44a4-9789-316428cd857b>).

What challenges were encountered?

These choices required substantial additional work beyond developing the experimental method itself: cleaning research code, documenting the software, preparing reusable datasets, and organising long-term archiving. This effort is an integral part of producing robust scientific research rather than an optional extra.

What difference or impact did these open research practices make?

I consider that making my research open source acts as a safeguard for scientific honesty by exposing the complete workflow to scrutiny, while also facilitating collaborations and scientific discussion. This has already proved useful for recent collaborations.

Higson-Sweeney - Developing Minds Digest: Co-designing a new resource for research dissemination in youth mental health

Nina Higson-Sweeney

What open practices were used?

The main open research practice was citizen science, specifically co-production and public involvement. We actively collaborated with parents, teachers, clinicians, and researchers through stakeholder workshops to co-design each element of the website, including its design, content, functionality, and implementation. The website is also openly accessible to all and has been designed in a way that facilitates transparency and reproducibility.

Why were these open research practices chosen?

These practices were chosen to make research more accessible to a wider audience without losing accuracy or credibility. Proactively involving the intended audience from the outset enabled us to understand what information they needed, how it should be presented, and what would make the website a trusted resource.

How were these open research practices implemented?

We implemented citizen science by facilitating five online co-design workshops with a diverse range of stakeholders, where we presented the initial concept of a research dissemination website and asked questions in relation to content, design, functionality, credibility, and implementation. This feedback was collated and used to develop a prototype website, which was reviewed by stakeholders prior to the official website launch. Open access was implemented by ensuring that the website is freely available to an international audience. Transparency was supported through the development of guidance that describes how summaries are written and reviewed prior to publication, allowing users to understand how information arrives on the website. Research is translated into concise, jargon-free summaries that retain links to the original publications, enabling users to access the full article where available. Each summary identifies the contributing researcher, highlights open access and public involvement, and clearly outlines practical implications.

What challenges were encountered?

One challenge was balancing accessibility with scientific accuracy. Stakeholders consistently requested summaries that were simple and engaging yet included enough information to avoid misinterpretation or sensationalism. We addressed this by

developing clear reporting guidance that prioritise plain English, accurate representation of findings, and direct links to the underlying evidence.

What difference or impact did these open research practices make?

These practices strengthened the quality and accessibility of the website, enhancing the impact of the website with the intended audience and increasing engagement.

Fatemeh - Cortical and allelic determinants of age-related cognitive decline: an interdisciplinary approach to the neuroscience of ageing

Dr Fatemeh (Simeen) Tabassi Mofrad

What open practices were used?

The research was published open access under a CC BY-NC-ND 4.0 license, with transparent reporting of study design, MRI processing, participant selection, statistical modelling and age-, sex- and region-specific analyses, supported by detailed methods and supplementary materials to facilitate scrutiny and reproducibility. Using 46,111 UK Biobank participants, we developed a comprehensive normative atlas of subcortical grey-matter ageing, providing a reusable benchmark for research on healthy ageing and neurodegeneration. The project applied responsible open-science and data-stewardship principles, enabling reproducible engagement with the analytical framework through established UK Biobank governance. Findings were further disseminated through international conferences, Oxford Psychiatry News & Impact and social media, extending their reach across scientific, clinical and public audiences.

Why were these open research practices chosen?

These practices were chosen to maximise accessibility, transparency, reproducibility and long-term value. Given the study's scale and interdisciplinary relevance, open dissemination and detailed reporting enable researchers across neuroscience, ageing and neurodegeneration to evaluate, reproduce and build upon the work. The atlas also provides a reusable benchmark for studying healthy brain ageing and neurodegeneration beyond the original publication.

How were these open research practices implemented?

Open research was embedded throughout the study through transparent methodological reporting, responsible data governance and targeted dissemination. Detailed methods, MRI processing, quality control, statistical modelling and supplementary materials supported reproducibility, while UK Biobank governance ensured responsible access to participant data.

The findings were subsequently shared through major international conferences, the University of Oxford Neuroscience Symposium, social media and Oxford Psychiatry News & Impact, including '*Charting the Ageing Brain: Insights from 46,000 MRI Scans*'.

What challenges were encountered?

The main challenge was balancing openness with responsible governance of large-scale human data, while communicating complex, sex-specific ageing trajectories clearly. This was addressed through rigorous methodological reporting, supplementary materials and governed data access, alongside targeted dissemination across academic, clinical and public audiences.

What difference or impact did these open research practices make?

These practices supported greater transparency, reuse and collaboration, helping make the findings more accessible and useful to researchers working on healthy ageing and neurodegeneration. The atlas provides a valuable benchmark for studying patterns of healthy brain ageing and supporting future research into neurodegenerative change. Dissemination through international conferences, Oxford Psychiatry and social media further extended the reach of the findings, engaging scientific, clinical and public audiences.

Raja - QuADMET-Former: Pre-training on Quantum Mechanical Properties Improves ADMET Prediction

Arun Raja, Hongtao Zhao, Christian Tyrchan, Eva Nittinger, Michael M. Bronstein, Charlotte M. Deane, and Garrett M. Morris

What open practices were used?

My project was shared openly at every stage of the research process. The preprint was posted on ChemRxiv (doi.org/10.26434/chemrxiv.15002429) ahead of journal submission, with an earlier version of the work shared at the International

Conference for Machine Learning 2024 AI for Science Workshop and NYC AI for Chemistry Summit. All code for pre-training, fine-tuning, baselines, and configuration files is available on GitHub under an MIT licence (<https://github.com/arunrajahub/quadmetformer>), and the pre-trained model checkpoints are available on HuggingFace (<https://huggingface.co/arunraja007/quadmetformer>) so that others can reuse the models without repeating the expensive pre-training step. My work was built entirely on openly available datasets and software: QMugs for quantum mechanical pre-training targets, and Therapeutics Data Commons, Biogen ADME (via Polaris) and OpenADMET for the 23 downstream benchmarks, with the TorchMD-NET equivariant transformer as the model backbone. Benchmarking was carried out against published open baselines (MapLight+GNN and Quantum Informed Pre-training). Therefore, my project is completely reproducible and openly accessible by anyone.

Why were these open research practices chosen?

Machine learning for ADMET prediction has a well-known reproducibility problem: reported performances often rest on favourable data splits, unreleased model weights, or proprietary datasets that are publicly inaccessible. Since my main thesis here is that quantum-mechanical pre-training beats strong existing methods, it is only credible if a reader can rerun the comparison themselves. Therefore, we prioritised openness right from the beginning. Releasing the pre-trained model checkpoints also had a practical motivation: quantum-chemistry pre-training is computationally costly and energy-intensive, and sharing the trained models means the community does not have to bear the costs. As an academic-industry collaboration between Oxford and AstraZeneca, we made the deliberate choice to openly release everything so that our methods and models could be used by groups without industry-scale compute or data.

How were these open research practices implemented?

Datasets were selected on the basis that they are publicly available. Evaluation used challenging spectral splits rather than random splits, so that reported performance reflects genuine generalisation to structurally distinct molecules. The code for the splitting procedure is provided in the GitHub repository. The repository is organised so that others can easily follow the code based on the structure of the preprint. I have also provided a documented environment specification, separate pre-training and fine-tuning entry points, and scripts for the baseline comparisons. Statistical significance was reported alongside raw performance (16 of 23 benchmarks best-performing, 13 of those significant). The preprint was updated across successive versions as the work developed, keeping the public record current rather than frozen at first release.

What challenges were encountered?

Reproducing published baselines took substantial effort, since these methods were often evaluated under different splits and preprocessing, requiring re-implementation and re-evaluation on the same splits. Combining three benchmark collections (TDC, Biogen ADME, OpenADMET) with differing formats and endpoint definitions was time-consuming. In addition, preparing a research codebase for public use by pinning dependencies for equivariant GNNs with CUDA requirements, writing documentation, and hosting large checkpoint files took a lot of time.

What difference or impact did these open research practices make?

The transparency we displayed throughout the research process directly improved the quality of the science. Committing to open splits and open baselines removed the temptation toward favourable evaluation choices, and the use of spectral splits produced a harder, more honest test than the field's default of random or scaffold splits. Also, by reporting significance rather than bare wins we did not overstate the performance of models and methods. All aspects of our project are openly accessible, and this lowers the barrier for smaller academic groups to use quantum-chemistry-informed models. Pre-printing accelerated the distribution of our work and invited the community's feedback ahead of formal review. And most importantly, the project demonstrates that an academic-industry collaboration involving a major pharmaceutical company can be conducted and released fully in the open, which is still far from the norm in AI-driven drug discovery.

Walker Silverman - Can social media change attitudes towards refugees?

Emma Walker-Silverman

What open practices were used?

This study, which involves naturalistically exposing participants to different kinds of Instagram content to investigate their effects on attitudes towards refugees in Turkey, has received in-principal acceptance as a Stage 1 Registered Report. This form of publication involves peer review before study implementation, and a corresponding commitment from the journal to publish regardless of the results, so long as the study follows the agreed upon design and planned analysis. This requires specifying and registering the planned design and analysis in great detail. The practice greatly increases transparency throughout the research process and reduces both researcher degrees of freedom as well as publication bias. It also requires rigorous statistical analysis with much higher power for each hypothesis and more strenuous checks against p-hacking practices than are customary. Any changes made to the design after in-principal acceptance must be approved by the journal before implementation, and data will subsequently be made as open as possible, while respecting privacy constraints.

Why were these open research practices chosen?

I chose to design and submit this study as a Registered Report in order to benefit from reviewer feedback before implementation, at a stage when suggestions can be far more impactful. I also wanted the study to benefit from the level of methodological rigor the approach requires.

How were these open research practices implemented?

I wrote up the introduction and methods section of the paper, including a planned analysis with decision trees for all eventualities, and submitted this in September of 2024. I received in-principal acceptance for the paper in late April 2026, after 19 months of iterative revisions.

What challenges were encountered?

In all honesty, this process has proved much more difficult and time consuming than I could have imagined. The extended review process (which followed 12 months already spent going through ethics review processes) has significantly delayed my degree timeline. Despite their timelines, neither process resulted in fundamental design changes. Moreover, during this extended review process the political context that the intervention was designed for shifted significantly, making elements of the design no

longer appropriate. The process for seeking approval for adjustments made in response to these contextual shifts has proven equally arduous.

What difference or impact did these open research practices make?

While one of the two most notable effects of the Registered Report approach has been the considerable delays mentioned, the other has been a substantial expansion of my statistical analysis capabilities. It is this skillset that I hope will improve not only this but also future research I undertake.

Computer science team - Inclusive Research Design Framework

Rosa Arias-Yague, Reuben Binns, Leto Luana Riebel, Emma Dunlop, Kshitij Mohan, Jennifer Watson

What open practices were used?

Co-design, iterative feedback, openly sharing the .pdf and interactive tool, continued community development and adaptation.

Why were these open research practices chosen?

To support transparency and a more inclusive research design, bring different research and professional perspectives into its development, and create a resource that could be adapted beyond Computer Science.

How were these open research practices implemented?

Review of existing University resources → funding through an MPLS Research Culture Award → development of initial framework → co-design workshop → further academic/research support input → field-specific tailoring → interactive tool → further field-specific tailoring (ongoing)

What challenges were encountered?

Creating a common framework that remained meaningful across very different areas of Computer Science, from Quantum Computing to Health Informatics, while keeping it concrete, relevant and simple enough for researchers to use.

What difference or impact did these open research practices make?

Feedback led to a simpler response structure, clearer guidance and field-specific case studies, while further collaboration led to the development of the interactive online tool. The resulting open resource supports more transparent and robust research practice by helping researchers articulate assumptions, limitations and applicability. Its open and adaptable format supports wider reuse, further adaptation and continued development, contributing to a more collaborative and open research culture.

Jacob - Building an Open Model of Childhood Brain Cancer to Accelerate Treatment Discovery

John Jacob

What open practices were used?

We shared our medulloblastoma organoid study as a preprint and an open-access paper (DOI: 10.1093/noajnl/vdae218), deposited the single-cell data in a public repository, and released the analysis code on GitHub with a stable Zenodo archive. Detailed methods, quality-control results and supporting analyses accompanied the paper. These practices supported the FAIR principles of making data Findable Accessible, Interoperable and Reusable. We also reused open patient and tumour-model datasets to assess how closely our model resembled the disease in patients.

Why were these open research practices chosen?

We wanted others to be able to check our conclusions and take the work further. Single-cell analysis involves many decisions that a paper alone cannot fully describe, while generating organoids and sequencing data takes considerable time and resources. Sharing the methods, data and code helps that investment benefit more than one study. Reusing independent patient datasets also provided a stronger test of whether our model captured features of human disease.

How were these open research practices implemented?

We deposited raw and processed sequencing data, with sample information, in GEO (GSE254917). R Markdown analysis notebooks were shared on GitHub and preserved as a versioned, citable archive in Zenodo. We documented laboratory protocols, software, quality-control thresholds and cell-line authentication, and deposited the paper and preprint in the Oxford University Research Archive (ORA). This gave readers a route from the published findings to the underlying data and analysis.

What challenges were encountered?

A key challenge was integrating our organoid co-culture data with open single-cell datasets from patient tumours generated by other laboratories. Our initial approach blurred distinct tumour-cell states, so we chose a method that corrected technical differences while better preserving biological variation. We documented and shared the analysis code so others could scrutinise these decisions. Reproducing the laboratory model still requires specialist organoid expertise, even with detailed protocols.

What difference or impact did these open research practices make?

Sharing made our work available for others to check and build on (e.g. DOI: 10.1242/dmm.052893). The paper attracted independent citations and was recognised among Neuro-Oncology Advances' most-cited recent articles. It led to an invited, fee-waived open-access review in the journal, 'Onco'. Our public UKRI record also prompted an enquiry for a James Lind Alliance study mapping research against patient priorities.

Alfuraih - Porous archiving: Methodological response to poles of archival hesitation and loss anxiety

Rawan Alfuraih

What open practices were used?

The accepted paper for publication presents a theoretical and methodological design of 14 principles and 38 implementation steps for ethical archiving in research, based on four years of fieldwork examining current laws, 'data ideology', self-perceptions, and the duties of practitioners across different disciplines. It engages with data reproducibility in open science by planning from an early stage for the future deposit of my field recordings. The submission for publication is also intended as a preregistration before fieldwork, while remaining dynamic and with plans to report deviations after the doctorate.

Why were these open research practices chosen?

Fieldwork revealed isolation and polarisation between ethics and sustainability. Open science advocates data accumulation and reproducibility, while archives emphasise their existential role in providing long-term open access, but researchers may dismiss archiving because of ethical concerns. To restore balance between ethics and sustainability, and connection between key actors, rather than pursuing 'open science', Porous archiving proposes thoughtful openness through ongoing ethical thinking. These practices were also chosen to address the problem of late archival planning, where researchers typically begin thinking about archiving only after fieldwork, often encountering ethical and logistical difficulties.

How were these open research practices implemented?

While applied to my own research, the framework is designed for other researchers to replicate and situate. The implementation steps extend from before fieldwork begins, throughout the research lifecycle, and beyond, within three roadmaps for gaining consent, building context, and conducting sensitivity checks. The accepted paper is only the beginning of a wider research agenda and further experimentation in ethical archiving.

What challenges were encountered?

While the fieldwork was initially undertaken to resolve the technical challenge of archiving by building a digital platform for uploading and browsing field recordings worldwide, it revealed a prior need to address behaviours and practices amid polarisation across key actors, ethical gaps in existing universal laws, conflicting

institutional missions, and funding requirements that disciplinary training does not adequately prepare researchers to meet.

What difference or impact did these open research practices make?

Beside connecting dialogue between isolated actors, the paper speaks to researchers about the broader possible impact of archiving field recordings, seeking to shift mindsets and equip them with practical principles and implementation steps. More broadly, it advances discussion on sustainable research data often neglected by social scientists.

Pu - Open Research in Early Drug Discovery

Hua Pu

What open practices were used?

The candidate embedded a range of open research practices throughout his work, including open-access publishing, open data, shared protocols and materials, reproducibility measures, open peer review, and the development of an open research tool (1, 2).

Why were these open research practices chosen?

These open research practices were used to strengthen reproducibility, meet funder requirements, encourage collaboration, improve accessibility and transparency, and maximize the impact of the research.

How were these open research practices implemented?

All experiments were independently repeated at least twice to strengthen reproducibility. Source data were published alongside the figures to support research integrity, and large datasets were deposited in public databases to ensure they are accessible and reusable. Detailed methods and protocols were also published to enable replication and verification by other researchers. The work was made available through open access publishing, ensuring that the research is freely accessible to the public. In addition, the compound CCM was developed and shared freely as an open tool for the research community.

What challenges were encountered?

The main challenge is to maximise accessibility and transparency while protecting future research opportunities and intellectual property rights. This requires careful balance, for example, between filing patent applications and sharing open research tools, or between promoting openness and protecting confidential information during drug development.

What difference or impact did these open research practices make?

What difference or impact did these open research practices make?

These open research practices have increased the visibility, recognition, and impact of the work. CCM has been adopted by research groups in the United States, Canada, the Netherlands, Switzerland, and the United Kingdom, demonstrating its value as a research tool for studying circadian modulation. The work was featured in a *Nature Chemical Biology* News & Views article, recognised by Web of Science as a highly cited

paper, and described by leading researchers as a “breakthrough”, “a significant conceptual advance”, “a new frontier”, and “promising”(3-5).

Reference

1. H. Pu et al., Nature chemical biology 21, 736–745 (2025).
2. K. Ondrova et al., Nature communications 14, 2728 (2023).
3. V. Senthilvadivel, T. Hirota, Nature chemical biology 21, 614-616 (2025).
4. K. Panda et al., FASEB J 40, e71947 (2026).
5. K. V. Smirnova et al., Brain Sci 15, 2076-3425 (2025).

Capel et al - Dept of Statistics - SAbDab2: The structural antibody database in the age of machine learning

Henriette L. Capel*, Odysseas Vavourakis*, Benjamin H. Williams, Christopher R. Taylor, Charlotte M. Deane

* These authors contributed equally

What open practices were used?

The Structural Antibody Database (SAbDab) collects and organises all public-domain data on the molecular structure of antibodies. These immune proteins are critical to medicine, and now widely used as therapeutics, diagnostics, and in vaccine development. SAbDab serves this data up freely on the Web and, at 45,000 annual visitors, is the central resource of its kind. Our latest update, SAbDab2 (<https://sabdab.opig.stats.ox.ac.uk/>), restructures the database in accordance with FAIR data principles to facilitate machine-learning applications and open data access.

Why were these open research practices chosen?

SAbDab data is widely used to train AI models that will help design the next generation of antibody drugs. It is in everyone's interest to make this data as broadly accessible as possible, to standardise and make it re-usable, and to keep it continuously updated, both to accelerate progress and deepen our understanding of antibody structure, binding, and design.

How were these open research practices implemented?

SAbDab2 is designed as a resource for the field for years to come. It introduces support for novel (synthetic) antibody formats, which are of growing interest, and makes it easy to compare all known structures of a given antibody. All data and annotations can be freely accessed, searched, and downloaded via an intuitive web interface. Using a dedicated API endpoint, SAbDab2 integrates bi-directionally with the Protein Data Bank, with PDB records updating automatically. It also maintains a versioned set of high-quality structures suitable for training and testing AI models without further processing, giving the field standardised data cleaning and splitting and helping to ensure fair comparison between models.

What challenges were encountered?

Considerable effort went into designing a data model flexible enough to support the wide array of synthetic constructs now in development, and an annotation pipeline that allows for flexible intervention by database maintainers. All aspects of the project, from back-end performance to UI design, were informed by end-user needs and community

feedback, to ensure the data is findable, accessible, and useful to researchers and the public alike.

What difference or impact did these open research practices make?

SAbDab2 continues to support structural antibody research worldwide. This update guarantees long-term sustainability of the database, while keeping step with emerging trends in immunoinformatics and supporting AI-enabled drug discovery. The data is updated weekly, and we are confident adherence to FAIR principles will ensure its long-term impact. Versioned data splits for ML with individual DOIs are released six-monthly.