

University of Surrey
Annual Open Research Culture
Event

Wednesday 31st January 2024



UNIVERSITY OF SURREY

Open Research



Dear Colleagues,

Welcome to the University of Surrey Open Research Culture event. This is the third in a series of annual events, planned for the forthcoming years. The series provides a platform for our community across all disciplines and career stages to discuss and share their knowledge, experience, and aspirations in Open Research Culture. In designing this year's event, we acted on feedback from our very successful [Surrey Open Research and Transparency Showcase](#), held in March 2021 and Surrey Open Research Culture events, held in April [2022](#) and February [2023](#). Our 2024 event includes presentations from the Open Research Award shortlist, invited lectures, networking opportunities and a panel discussion, whilst our drop-in clinics are designed for further discussion, advice and signposting. The event is for those who are just starting to think about Open Research, to those who already embed Open Research practices into their research.

The 2024 Annual Open Research Culture event could not have come together without the help and contribution of many people. First, we would like to thank our invited speakers, Dr Steven Hill and Prof Marcus Munafo who have given up their time and travelled to be here with us. I would also like to thank our University of Surrey contributors, Dr Carla Bonina (panel member), Prof Mark Plumbley (representing Haohe Liu), Dr Sacha Beniamine, Dr Ana Andries and Dr Mojtaba Soltanlou (Open Research Award shortlist presenters). We also thank our Pro Vice Chancellor Research and Innovation, Prof Lisa Collins, for supporting our Open Research culture, and for opening the event. I would personally like to thank the rest of the organising team: Anna Korzeniowska (lead), Peter La, Montserrat Rodriguez-Marquez and members of the [CoGDeV](#) lab.

We hope that this is an inspiring event and that you enjoy the day.

Professor Emily Farran

[UK Reproducibility Network \(UKRN\) Institutional Lead](#) for University of Surrey.

Your key Open research resources:

- [Open Research checklist](#), an at a glance guide to recommended Open Research actions
- [Open Research handbook](#) with all useful links
- [Open Research website](#) with guidance on open practices
- [Open Research training](#), bookable and drop-in training opportunities
- [Open Research module](#) available to all staff and PGRs. A badge (micro-credential) is awarded on completion
- [Open Research across disciplines](#), with examples of good practice (now also a [UKRN webpage](#)).

Social media: #SurreyOR2024 @SurreyOR

Contents

Event Schedule	4
Invited Speakers	5
Open Research Award 2024 shortlist	8
Open Research drop-in clinics.....	15
Appendix: Open Research Award 2024 Abstract Submissions	17

Event Schedule

Time	Event	Location
9:00am	Registration	Austin Pearce (AP) Foyer
9:30am	Opening: Prof Lisa Collins, Pro Vice Chancellor Research and Innovation	Austin Pearce (AP) Lecture theatre 3-4
9:40am	Dr Steven Hill, Director of Research, Research England; Chair of the REF 2029 steering group	Austin Pearce (AP) Lecture theatre 3-4
10:40am	Coffee break	Austin Pearce (AP) Foyer
11:10am	Prof Marcus Munafò, Head of UK Reproducibility Network; Associate Pro-Vice Chancellor, Research Culture, University of Bristol	Austin Pearce (AP) Lecture theatre 3-4
12:10pm	Lunch Networking Open Research drop-in clinics	Austin Pearce (AP) Foyer Seating in Austin Pearce (AP) Lecture theatre 1-2
13:00pm	University of Surrey Open Research Award shortlist 2024	Austin Pearce (AP) Lecture theatre 3-4
14:00pm	Discussion panel: Dr Steven Hill, Prof Marcus Munafò, Dr Carla Bonina (Open Research champion, Surrey Business School) Chair: Prof Emily Farran, UK Reproducibility Network Institutional Lead.	Austin Pearce (AP) Lecture theatre 3-4
14:45pm	Open Research Award presentation, Prof Lisa Collins	Austin Pearce (AP) Lecture theatre 3-4
14:50pm	Round-up of the event: Dr Laura Evans, Nifty Fox Closing: Prof Emily Farran	Austin Pearce (AP) Lecture theatre 3-4
15:00pm	Close	Austin Pearce (AP) Lecture theatre 3-4

Invited Speakers

Dr Steven Hill

Director of Research, Research England; Chair of the REF 2029 steering group



Research Culture: from policy to implementation

Abstract

There has been an increasing focus on research culture in recent years. It is becoming more recognised that the conditions within which research is conducted, and the associated impacts on people and their careers, are important components determining the quality and impact of research. In this presentation, I will consider some of the evidence relating to research culture and the policy response, including the Government's R&D People and Culture Strategy and policy and initiatives from UKRI and Research England. I will also survey the importance of research assessment as a factor in research culture, and provide an update on the People, Culture and Environment component of REF 2029.

Biography

Professor Steven Hill is Director of Research at Research England. Steven was formerly Head of Research Policy at the Higher Education Funding Council for England (HEFCE) and leads on all aspects of research policy and funding. Steven is responsible for research funding, including quality-related funding (QR), general capital funding and the UK Research Partnership Investment Fund (UKRPIF). He also leads Research England's research assessment and policy work, and is the chair of the steering group for the 2029 Research Excellence Framework (REF). Policy responsibilities include

research integrity, public engagement and open research, and Steven contributes to debates and discussions at home and overseas on the enhancement and assessment of research impact. His team also includes Research England's analysis function. Steven has a degree in Natural Sciences from the University of Cambridge, and a PhD in Plant Biology from the University of Edinburgh. After a number of research posts, he was appointed to a lectureship in plant science at the University of Oxford, where he taught and researched for eight years. He then shifted career tracks into the world of policymaking, with two roles at the Department for Environment, Food and Rural Affairs, focusing on bringing science and research into policymaking. Subsequently, he was appointed Head of the Strategy Unit at Research Councils UK, and then took up the post of Head of Research Policy at HEFCE in 2013.

Professor Marcus Munafò
Chair of UK Reproducibility Network (UKRN); Associate Pro-Vice
Chancellor, Research Culture, University of Bristol



Research Ecosystems and Research Quality

Abstract

There is growing interest in how we can ensure research outputs are robust and high quality. This interest has extended both upwards – for example to the House of Commons Science Innovation and Technology Committee – and sideways – from the original biomedical literature where many of these conversations began to the arts and humanities. I will discuss – primarily from a biomedical perspective – how and why we can easily generate findings that are not robust, leading to error and waste, and slowing the speed with which the knowledge we generate benefits society. In particular, I will discuss how open research practices can help us create a research ecosystem that is more inherently trustworthy; one where the research process can be relied upon to ensure that our final research outputs are high quality.

Biography

Professor Marcus Munafò is Professor of Biological Psychology and Associate Pro-Vice Chancellor (Research Culture) at the University of Bristol. Professor Munafò's research focuses on understanding pathways into, and the consequences of, health behaviours and mental health, with a particular focus on tobacco and alcohol use. He also has interests in the role of incentive structures in science, and the extent to which these shape the robustness and reproducibility of scientific research. Professor Munafò is the Chair of the UK Reproducibility Network Supervisory Board.

Open Research Award 2024 shortlist

(Presenter in bold)

Talk 1: Building the research community with open-source practice

Haohe Liu¹, **Mark D. Plumbley**¹, Wenwu Wang¹

¹Faculty of Engineering and Physical Sciences; University of Surrey

Contact information:

Haohe Liu

Postgraduate Researcher

Faculty of Engineering and Physical Sciences

haohe.liu@surrey.ac.uk

<https://orcid.org/0000-0003-1036-7888>

Abstract

Our research employs artificial intelligence to pioneer advanced audio models that empower communication and kindle the creativity of content creators. Notable projects include AudioLDM, AudioLDM 2, and AudioSR, acclaimed models on GitHub with 2100+, 1800+, and 700+ stars respectively, and 25k+ downloads across HuggingFace and Zenodo. The AudioLDM, lauded on HuggingFace and featured in a popular space in February 2023, has been cited 131 times since its ICML acceptance.

The essence of our open research practice is accessibility and reproducibility for diverse audiences. For instance, AudioLDM's web demos and APIs engage the public interactively, supplemented by explanatory videos that garnered over 700 likes on Twitter. We provide source code for enthusiasts to modify, and comprehensive documentation for experts to replicate our work effortlessly.

Hosting the web API posed a significant challenge due to GPU resource demands. Collaboration with organizations such as HuggingFace, Replicate and Deque AI led to solutions for robust, cloud-hosted web demos.

Our technology has captivated the machine learning and audio production communities. Its application ranges from movie production to music sample generation and sound effects for gaming, illustrating the vast potential of our open-source technology. Key lessons we learnt from our open research practice include the importance of audience-tailored presentations, contingency planning for resources, proactive industry engagement, and the impactful role of demonstration videos. Active participation in open-source discussions and presentations has been crucial for community engagement and visibility.

Talk 2: Leading Open Data Practices in Linguistics

Sacha Beniamine¹, Matthew Baerman¹, Mae Carroll³, Grev Corbett¹, Matteo Pellegrini², Erich Round¹, Helen Sims-Williams¹

¹Surrey Morphology Group, SLL; ²U. of Milano; ³University of Melbourne

Contact information:

Sacha Beniamine

Research Fellow

Faculty of Arts and Social Sciences

s.beniamine@surrey.ac.uk

<https://orcid.org/0000-003-2584-3576>

Abstract

Introduction: The Surrey Morphology Group is a world-leading research group, which has published Open language data since the 1990s and is now at the forefront of innovative data practices in Linguistics. We devised the DeAR principles to bolster open science: Decentralized standardisation for interoperability; Automated validation for data quality; and Revisable pipelines to keep data up-to-date.

Practices: Many valuable databases are trapped in legacy software. We have salvaged, standardized, cleaned and openly released the Romance Verbal Inflection Dataset 2.0 (<https://doi.org/10.5281/zenodo.3611076>) and Surrey Morphological Complexity Database (<http://dx.doi.org/10.15126/SMG.23/1>). To foster FAIR and DeAR data, we created the Paralex standard (<https://paralex-standard.org/>, Beniamine, 2023) and tools for its use (<https://pypi.org/project/paralex>) and automated publication (<https://pypi.org/project/gitlab2zenodo/>).

Challenges

1. Standardization must impose norms without restricting scientific decisions. Paralex imposes strict formats, but leaves freedom for alternate analyses.
2. Short-term funding hinders long-term data maintenance. We prioritize low-maintenance open infrastructures (static sites) leveraging archiving services (zenodo; Surrey Open Research Repository) and version management (git) to ensure longevity.

Benefits: Paralex enjoys world-wide uptake (<https://zenodo.org/communities/paralex>) and salvaged datasets have supported publications (Beniamine 2021, Cathcart 2022, Herce 2023). Gitlab2zenodo (1,670 downloads) enjoys use from immunology to agriculture. New collaborations have arisen (IAS Fellowship, Matteo Pellegrini), and DeAR principles feature in graduate classes from Uppsala to Canberra.

Lessons: International cooperation is key for documenting the world's 7000 languages. Machine-readable downloads are essential, yet human-friendly visualisations remain crucial for qualitative research. We devised workflows to deliver both, borrowing methodologies from software development, including continuous pipelines, test suites, and versioning.

Conclusion: We have led an Open Data transformation through new standards, open tools and data-rescue, attracting enthusiastic early-adopters.

Talk 3: Space4Nature (S4N)

Ana Andries¹, Stephen Morse¹, Richard Murphy¹

¹Centre for Environment and Sustainability (CES), University of Surrey

Contact information:

Ana Andries

Research Fellow

Faculty of Engineering and Physical sciences

ana.andries@surrey.ac.uk

<https://orcid.org/0000-0002-6042-269X>

Abstract

Space4Nature (S4N) is a collaborative project spanning three years, dedicated to addressing critical biodiversity challenges in Surrey County, UK. Focused on predictive habitat mapping and habitat quality assessment, the project integrates advanced technologies, including very high-resolution (VHR) satellite data and citizen science, to strategically restore habitats and establish Nature Recovery Networks (NRN).

S4N aligns with Open Research (OR) objectives, aiming to enhance accessibility, transparency, and reproducibility in scientific endeavours. Embracing open practices, the project utilises VHR satellite data and citizen-collected ground truth data, employing various Machine Learning (ML) methods for transparent habitat classification prediction and assessing quality. Citizen scientists, actively participating in ground-truthing surveys, contribute to the open and participatory nature of the research.

In the pursuit of openness, S4N faced challenges in ensuring the reliability of results. To address this, the project incorporated open peer review practices involving ecologists from Surrey Wildlife Trust and Buglife. Physical visits by ecologists and repeated citizen science surveys were employed to validate results collaboratively, establishing a transparent model for scientific findings. The open practices and methodologies of S4N contribute to the reproducibility and transparency of research processes. By sharing historical surveys spanning two decades openly, the project supports knowledge exchange and collaboration. The use of comprehensive evaluation methods, including supervised classification metrics and field validation, ensures the robustness of the methodology. S4N's experience underscores the importance of collaborative validation for transparent research. Involving citizen scientists and ecologists in open peer review processes ensures the reliability of findings. The project also emphasises the significance of sharing methodologies, data sources, and validation processes openly on the LandApp platform. S4N exemplifies a

commitment to OR practices, promoting transparency, accessibility, and collaboration in the pursuit of biodiversity conservation and sustainable land use practices. The project's approach offers valuable lessons for future research endeavours, emphasising the benefits of open methodologies and collaborative validation processes

Talk 4: The Learning Brain Lab

Mojtaba Soltanlou¹, Arcan Altinar¹, Elizaveta Ivanova¹

¹School of Psychology, University of Surrey

Contact information:

Mojtaba Soltanlou

Lecturer

Faculty of Health and Medical Sciences

m.soltanlou@surrey.ac.uk

<https://orcid.org/0000-0003-1959-1384>

Abstract

The Learning Brain Lab <https://www.surrey.ac.uk/learning-brain-lab> investigates mathematical cognition through various behavioural and neuroimaging methods. In the Learning Brain Lab, we are committed to Open Research practices, including high-quality, easily reproducible, and shareable research.

Here are some of our Open Research activities:

- Registered Report, Stage 1 (IPA), Journal of Experimental Psychology: General
- Registered Report, Stage 1 (IPA), Cortex
- Registered Report, Stage 1 (in revision), Peer Community In
- Registered Report, Stage 1 (under review), Journal of Experimental Psychology: General
- Preregistration of a cross-cultural study on OSF and dissemination in nurseries
- Preregistration of a systematic review/meta-analysis on PROSPERO and OSF
- Three ongoing preregistrations of our new projects
- Involvement in three multi-site open-data projects
- Publication of over 10 open-access peer-reviewed papers
- PI's editorial board membership in an open-access journal
- Equity, Diversity, and Inclusion efforts through workshops in South Africa and Iran

By following Open Research practices, we encountered barriers and learned valuable lessons:

(i) Maintaining Open Research practice across the lab: We established this by implementing protocols of conduct, e.g., quality-checking.

(ii) Time-consuming pre-registration/Registered Report and accumulated time pressure: We have learned to have these discussions early on in the project's development to account for the time commitment needed.

(iii) Worry about mistakes affecting transparency: We have navigated this concern by being open about our needs, reviewing each other's work, and seeking to upskill when needed.

Open Research within our lab has inspired, educated, and evolved our practice. Each obstacle has created a learning opportunity, and we are excited about the future of our work aligned with Open Research. In the Learning Brain Lab, we strongly believe that Open Research significantly helped us to further build a reputation in scientific societies and allowed our lab to grow within ourselves and our work.

Open Research drop-in clinics

Clinic 1: Open Research Institutional Repository

You are cordially invited to a short practical session on how to use the University's Open Research repository. Bring along your laptop and a member of the Open Research team will show you how to register your ORCID, create a record for your publications and add your files, and upload research data sets or create records for your research data. As you know, the Library helps you check publishers' policies on open access of your deposited papers to ensure that you meet funder requirements. We are also happy to chat with you about open access issues, researcher identifiers, etc. This clinic will be available at registration, coffee break and lunchtime.

Clinic 2: Faculty Librarian team and the Library Research Hub

Meet with the Faculty Librarian team who will demonstrate the Library Research Hub, reveal all about the library resources and tools that can support your research interests and how to make the most of your Open Research journey at Surrey. This clinic will be available at lunchtime.

Clinic 3: Research Integrity and Governance (RIGO)

The RIGO team work with the University Ethics Committee to provide support, advice and guidance to the University's researchers. Come and speak to us to find out about the launch of Ethics RM, the new system for making ethics and governance applications. Ethics RM will streamline the administrative process and allow researchers to track the progress of their applications. To aid collaboration, applications can be shared between researchers and supervisors will receive automatic notification to approve submissions by students. The RIGO team look forward to sharing with you how Ethics RM will simplify the application process. This clinic will be available at registration, coffee break and lunchtime.

Clinic 4: Research Data Management

The Research Data Manager, Dr Michelle Willows, will be available to answer any queries students or staff may have regarding all aspects of Research Data Management, from how to collect, store and share qualitative and quantitative data, to how to create Data Management Plans. Michelle can advise on uploading data to our Surrey Open Research repository and many other data related activities. So please come along on the day and just ask! This clinic will be available at lunchtime.

Clinic 5: Institute for Sustainability

This Institute aims to increase the scale and enhance the excellence of the University's research and innovation in priority areas of sustainability research. Representatives of the Institute will attend the Open Research Culture event to help increase the visibility of Sustainable Development Goals

Appendix: Open Research Award 2024 Abstract Submissions

(1) Being open to fix the climate faster!

Damon Hart-Davis; Center for Environment and Sustainability, University of Surrey

Contact information:

Damon Hart-Davis

Postgraduate Researcher

Faculty of Engineering and Physical Sciences

d.hart-davis@surrey.ac.uk

<https://orcid.org/0000-0003-4744-8702>

Abstract

Having spent the last ten years designing a better domestic heating control in a startup, I now want to continue to help with robust decarbonisation of UK home heating from a research perspective. Part of that is being as open as possible with code and data (my current research model and data is public on GitHub, more generally I publish lots on my own site) and discussing with practitioners in the field (heat pump designers, installers, influencers) what the realities are. My current work seems to have found a small but significant hole in the formal literature and in practitioners' understanding, which could be resulting in discarding maybe 20% or more of very cheap central heating energy savings, and some householder comfort, when retrofitting from gas boiler to heat pump. Being open may have made for a small quick win for the climate and people's pockets. I have long been an advocate of open source software and hardware, and it seems the same benefits flow from open research.

(2) Using Open Research tools to uncover a new therapeutic target for an aggressive blood cancer

Maria Teresa Esposito; School of Biosciences, University of Surrey

Contact information:

Maria Teresa Esposito

Senior Lecturer Biochemistry

Faculty of Health and Medical Sciences

Mariateresa.esposito@surrey.ac.uk

<https://orcid.org/0000-0002-9147-6906>

Abstract

Using publicly available datasets, we discovered a gene called SET as over-expressed in acute myeloid leukaemia (AML) and correlating with poor prognosis. By analysing additional freely accessible datasets, we uncovered that SET expression highly correlates with the expression of HOXA genes, identified as part of the “molecular signature” of an aggressive subtype of AML (known as KMT2A-AML), prompting us to investigate the role of SET in this disease in our laboratory models. We discovered that SET is required for the survival of KMT2A-leukaemia and that silencing SET reduces the expression of HOXA genes, identifying SET as a new player and a potential therapeutic target for KMT2A-AML. The clinical data we accessed through the public datasets were fundamental to prompt us to pursue this research project and to make these important discoveries. Moreover, they enabled us to reproduce and validate the data extracted from the public datasets. The analysis of some of these datasets was done using software that are freely available and user-friendly, others required a dedicated bioinformatician and computer power; this can limit their application and usability. Given the significance of our data, we made freely available our own datasets, generated as part of this project, through the platform NCBI and PRIDE, the pre-print manuscript on Research Square and presented the data at the EACR congress. This was instrumental to attract the attention of AstraZeneca, and to start conversations on partnerships. Our paper is now published open access in Oncogene, a hybrid journal covered by an agreement with our university. As SET is over-expressed also in several solid human and canine tumours and in Alzheimer’s disease, I hope the accessibility of our research will enable us to establish new collaborations in these research areas.

(3) Bridging Gaps: The Transformative Power of Data and Code Sharing in Multidisciplinary Infectious Disease Research

Natalia Martinez; Department of Comparative Biomedical Sciences, School of Veterinary Medicine, University of Surrey

Contact information:

Natalia Martinez

Postgraduate Researcher

Faculty of Health and Medical Sciences

n.martinezmunoz@surrey.ac.uk

<https://orcid.org/0009-0007-3214-8076>

Abstract

Bridging Gaps: The Transformative Power of Data and Code Sharing in Multidisciplinary Infectious Disease Research

Clinical trials pose significant challenges in epidemiology and infectious disease analysis due to the high costs, time, and effort involved. However, the landscape is evolving with technological advancements, the implementation of open research policies, and an increasingly multidisciplinary collaborative approach. This paradigm shift has not only streamlined the acquisition of diverse datasets but has also fostered a seamless integration of surveillance data, health electronic records, and production parameters across various dimensions of population dynamics. This integration has significantly contributed to the accessibility and abundance of relevant data.

How can we harness the power of these diverse datasets to foster a global dialogue with clinical trials, facilitating the prediction and comprehension of variations, influences, and consequences within populations?

A crucial ally in this endeavour has proven to be mathematical and statistical models capable of synthesizing evidence and shared data into predictive tools that aid in informed decision-making.

My academic journey at the University of Surrey reflects a commitment to the collaborative nature of scientific progress. The research methodology involves extracting meaningful information, working with confidential data, developing analytical codes, and sharing these codes with the broader scientific community.

This collaborative approach ensures that others facing similar datasets can benefit from established methodologies and anticipate potential outcomes.

My work underscores the transformative impact of sharing data and the analytical tools and codes that facilitate its interpretation. The collaborative ethos that underpins this practice accelerates scientific advancement, fostering a collective gain rather than individual accomplishment. By illustrating the iterative process of building upon shared knowledge, this kind of quantitative research emphasizes the indispensability of code sharing in the multidisciplinary pursuit of understanding and combatting infectious diseases. The findings highlight the necessity for continued collaboration and transparency in scientific endeavours to address the complex challenges posed by infectious diseases on a global scale.

(4) Open Research Practices in the study of rare genetic conditions

Gloria Yoshkova; School of Psychology, University of Surrey

Contact information:

Gloria Yoshkova

Postgraduate Researcher

Faculty of Health and Medical Sciences

g.yoshkova@surrey.ac.uk

<https://orcid.org/0009-0003-7402-4320>

Abstract

The proposed case study will focus on the importance and challenges associated with conducting open research in developmental psychology, particularly in the context of rare conditions. Two ongoing longitudinal studies will be used as examples to demonstrate the benefits and drawbacks of using open practices. These studies aim to follow the early motor development of children with Williams Syndrome - a rare genetic condition that affects 1 in 18,000 people in the UK. Children with WS are born with heart defects, developmental delays, and motor deficits among other characteristics. While motor skills are essential for almost everything we do, little is known about the impact of early motor impairment on later outcomes in individuals with WS. By designing open studies, we aim to address this gap and provide international guidance on how movement develops in WS and its association with other skills in children. We have created two public pre-registrations of the studies mentioned above. Committing to a publicly shared plan on how each step of the research takes place, demonstrates thorough and thoughtful planning of the study design, and more credible results as initial goals and later outcomes can be easily compared. This open practice can enhance reproducibility by informing others of the used methodologies. In the context of rare conditions, this can also prevent redundant studies, and allow often limited resources to be utilised in areas that are less explored. Pre-registrations also create opportunities for cross-disciplinary collaborations and raise awareness in the public and stakeholders who could support further research in the area.

Researchers considering the use of pre-registrations should be aware of some drawbacks associated with this open practice. Pre-registering study designs can require a lot of time, effort and preparation. Therefore, pre-registrations can be perceived as an administrative burden to small research teams, where time and

resources are limited. Studying rare conditions means working with small populations, which are not always easy to reach and where unpredicted circumstances can easily arise. This may lead to the need to adapt methodologies, and thus invest further resources into editing or creating new pre-registrations. In addition, this open practice may not be suitable for all types of research, for example, qualitative or exploratory studies. Finally, pre-registrations may discourage researchers from publishing unexpected results, the outcomes of which may be seen as a failure to follow the pre-registered plan.

In summary, pre-registrations for studies on rare genetic conditions ensure rigour and quality of research, prevent publication bias, and promote transparency and accountability, which are essential ethical considerations when working with vulnerable groups. However, pre-registering research can increase workload, may not always be suitable, and may prevent us from publishing important findings that were not part of the initial plan. Despite these challenges, the benefits of thorough and effective pre-registrations outweigh the drawbacks and promote research that is honest, meaningful and safe for those conducting it and for those taking part in it.

(5) A day in the life of a “Research Parasite”: added value by secondary analysis of published genomics datasets

Arnoud H.M. van Vliet; School of Veterinary Medicine, University of Surrey

Contact information:

Arnoud H.M. van Vliet

Senior Lecturer in Microbiology

Faculty of Health and Medical Sciences

a.vanvliet@surrey.ac.uk

<http://orcid.org/0000-0003-0203-1305>

Abstract

In January 2016, the editor-in-chief of the prestigious New England Journal of Medicine (NEJM) published a commentary describing their disapproval of "Research Parasites": scientists who do replication analyses but were not involved in the original publication, and were accused of "possibly stealing from the research productivity planned by the data gatherers", and the even worse crime of trying to "disprove what the original investigators had posited". They wanted co-authorships being given to the authors that had already published their data! This may have been the acid reflux of established academics from a bygone era struggling to adapt to the modern times, but it does show their underappreciation of the value of replication studies and independent secondary analysis of published data, a key concept of Open Science.

I use the term "Research Parasite" as an honorific, and my research focuses on the secondary analysis of genomics datasets of pathogenic bacteria. A requirement for publication is that these are deposited in repositories at EMBL and NCBI, and as such are publicly available. This has led to large datasets being publicly available; for example, there are >0.5million genome sequences of Salmonella isolates from all over the world. This allows for secondary analyses of trends in antimicrobial resistance, identification of novel factors involved in causing disease, and for the generation of novel diagnostic tests and the formulation of new treatments and One Health approaches to protect the efficacy of antibiotics for as long as possible.

The NEJM editors felt the heat and did qualify their comments, but it shows the continuous need to fight such bad views and practices, and highlights the advantages of Open Science. Next steps should include ensuring data analysis opportunities are as shared as data, and ensuring those in resource-limited countries are protected against the exploitative kind of research parasites.

(6) Modifying a self-efficacy tool for people with learning disabilities: inclusive, open and accessible research

Anna Cox¹, Athena Ip¹, Sally Kendall², Ali Jabeen, Scott Watkin

¹School of Health Sciences, University of Surrey; ²University of Kent

Contact information:

Anna Cox

Senior Lecturer, School of Health Sciences

Faculty of Health and Medical Sciences

a.cox@surrey.ac.uk

<https://orcid.org/0000-0001-5254-1296>

Abstract

Introduction: People with learning disabilities can be good parents with the right support. Interventions exist to support parents-to-be with learning disabilities, but evaluation of their impact is limited by a lack of accessible outcome measures. Our study used a process of inclusive research to modify a self-efficacy tool (TOPSE ante-natal) to make it accessible and meaningful as an outcome measure for people with learning disabilities.

Open research practices: The study aimed to be inclusive, open, and accessible. Inclusivity was achieved by fully involving people with learning disabilities in the research process. A father with learning disabilities was co-investigator, and specialist parent advocates supported the research. The breadth of experience in the research team and advisory group, and the focus on inclusivity and a sharing of power in the research process, created a tool that is accessible and meaningful to parents-to-be with learning disabilities. However, accessibility is important in terms of both the product and its dissemination. We published the paper as an open access article and made the modified tool available to others as a free download. However, 'open' doesn't necessarily mean 'accessible'. Therefore, we also recorded a short discussion of the paper and attached it to the article online. This discussion involved two authors, a researcher and a father with learning disabilities.

Benefits and challenges: This approach to accessibility took very little time or money. Following previous advice not to script people with learning disabilities, the film we made is relaxed and chatty. Our accessible tool measures the construct of 'self-efficacy' but in the discussion the term 'self-confidence' is used, a construct often used

interchangeably with self-efficacy. The challenge is achieving accessibility without compromising meaning, balancing lay and academic language.

Conclusion: Accessible dissemination gives opportunity to share why our work matters and what it means, with people it is designed to benefit. Accessible written summaries or recorded discussions should be widely adopted.

(7) Utilising open research practices to conduct a systematic review of healthcare data

Claire Price^{1,2}, Debbie Cooke¹, Freda Mold¹, Nadia A.S. Smith², Martyn Winn³,
Agnieszka Lemanska^{1,2}

¹University of Surrey, ²National Physical Laboratory, ³Science and Technology
Facilities Council

Contact information:

Claire Price

Postgraduate Researcher

Faculty of Health and Medical Sciences

Claire.a.price@surrey.ac.uk

<https://orcid.org/0000-0003-3060-9821>

Abstract

Introduction: Pancreatic cancer is rare but has dismal survival rates meaning it's the sixth cause of UK cancer mortality. It often presents with non-specific symptoms. This makes early diagnosis challenging. However, because weight loss occurs in most patients with pancreatic cancer, it could be a useful diagnostic marker. Our aim was to improve its utility as a marker for pancreatic cancer by quantifying pre-diagnosis weight loss. We undertook a systematic review with meta-analysis.

Open Research Practices: We registered the study protocol on PROSPERO (CRD42022302985) and followed PRISMA guidelines and FAIR principles. We used ROBINS-I to assess bias. Extracted data has been deposited with Zenodo and will be submitted as a Research Elements article. Results will be published as a preprint and then open access in a peer reviewed journal.

Challenges: Extracting data for meta-analysis was difficult due to heterogenous study types and a lack of transparency. Many original studies lacked meta-data or only presented data graphically. Authors of the original studies often could not clarify results. Without meta-data to enable data reuse this meant some studies had to be excluded from meta-analysis.

Benefits: Transparency of methodology enables reproducibility, preventing duplication of effort and cost. Depositing extracted data maximises the findability and accessibility of the data. This encourages data reuse practices which supports data minimisation. This is especially important for potentially invasive research.

Lessons: Original datasets that consist of healthcare records cannot be made open access. However, open research practices can still be adopted. It is important to present meta-data to provide context for the research. This improves standardisation across study types and facilitates study combinability.

Conclusion: Research using healthcare records needs to be approached carefully, due to privacy and other ethical considerations. However, maximising open research practices enables data to be shared to maximise patient benefit. This avoids repetition of work and increases reproducibility.

(8) MINDMAP

**Philip Dean¹, Juha Silvanto¹, Kathrin Cohen Kadosh¹, Ines Violante¹, Simon Evans¹,
Rocco Chiou¹, Tibor Auer¹, Shachar Hochman¹, Roi Cohen Kadosh¹**

¹School of Psychology, University of Surrey

Contact information:

Philip Dean

Research Fellow

Faculty of Health and Medical Sciences

p.dean@surrey.ac.uk

<https://orcid.org/0000-0002-2362-513X>

Abstract

MINDMAP is the name of the project created for Undergraduate dissertations within the Brain & Behaviour Section, School of Psychology. The objective was to make the data collected from dissertations more reproducible, improve their quality & potential to add to the research literature, and enable students to co-operate in a large research setting. This allows them to ask broader questions using data collected by other students and observe this data collection even if not a core part of their project, furthering their research experience. As the data is anonymised, the data, code, questionnaires & tasks will also be available as an open dataset, deposited in an open repository, freely available for future research by all researchers.

In practice, we created a structure that aims to combine all dissertations into one comprehensive, well-powered dataset, with certain measures taken for every participant (core measures, including consent, demographics & screening). Students can then create studies using these core measures on their own or in combination with a larger array of harmonised measures (e.g., surveys, behavioural tasks, neuroimaging, intervention) pooled and agreed across the researchers within the section. As all data have the same participant ID, all students can use all data collected. The project will run across dissertations in each year, but also across years to collect a large, longitudinal dataset.

We used a series of workshops using SurreyLearn to introduce students to the concept of MINDMAP, as well as the benefits of well-powered and reproducible research for their study question. Aspects of reproducibility and techniques used in open data were also covered, such as Open Science Framework, Brain Imaging Data Structure and data management. Students are given autonomy and ownership of their own project

within MINDMAP but have support for the new aspects they are learning. This support can now be both at a supervisor level, and as a group level within the workshops.

This is the first year we have run MINDMAP and there were, and still are, challenges and lessons learned in this format that we are adapting to. Setting up a project of this scale to cover all possible dissertations, recruitment style, and data whilst anonymising it has logistical challenges and requires some time to setup and check initially, and some adaptation to new challenges. In addition, the level and amount of information given to students would need adapting in coming years. For instance, using pre-registration tools such as aspredicted.org could work within the workshops for presenting study ideas when they formulate their study plan.

The benefits so far are that students like the autonomy this gives them, as well as the feeling of being in a research group together. It enables them to start straight away (ethics are already covered) and gives them some hands-on experience of a wider research setup. As the data comes in, we hope these benefits rise in the scope of their research possibilities and support, as well as being able to create adequately powered and potentially publishable outputs.

In conclusion, MINDMAP was setup to allow open and reproducible data to be collected within undergraduate dissertations over many years, whilst teaching aspects of general and open research techniques. There are already aspects we can see that have worked and are looking forward to the first cohort finalising their research in order to review to what extent the rest of our aims were achieved.

(9) COMFOCUS

Philip Dean¹, Monique Raats¹, Lada Timotijevic¹, Charo Hodgkins¹, Lisa Mohebat¹

¹School of Psychology, University of Surrey

Contact information:

Philip Dean

Research Fellow

Faculty of Health and Medical Sciences

p.dean@surrey.ac.uk

<https://orcid.org/0000-0002-2362-513X>

Abstract

COMFOCUS (<https://comfocus.eu/>) is a project involving institutes in different disciplines across Europe with an interest in food consumer science. The aim is to integrate and harmonise the measures used to measure data in food consumer science, allowing the data to be FAIR (Findable, Accessible, Inter-operable & Re-usable), and using RRI (Responsible Research & Innovation) data principles. The use of open science and harmonised datasets will then allow large scale data science and AI based analysis to be performed on adequately powered, well documented, harmonised and good quality research data across disciplines, countries and topics.

The main objective was therefore to make research in this area more reproducible by increasing the quality and documentation of information in published research, whilst enabling open data, code, material and publication where possible. This was done by first creating harmonised protocols for good design and documentation (metadata and reporting guidelines) and harmonised measures across a range of data types: self-report / subjective measures, emerging measures (psychophysiology, eye-tracking, VR, face reading), longitudinal and data mining. These would be then used in two open calls for research proposals, first with just self-report measures, and second with a call for researchers to visit the COMFOCUS institutes as “fellows” to learn how to use the emerging measures within their research, and apply the harmonised and Open Science principles (including pre-registration such as <https://aspredicted.org/>).

In practise, the tools we use for this are the harmonised protocol and measures documents, which are highly integrated into an online “COMFOCUS toolbox” which creates harmonised study and survey setups for ease of use for researchers, as well as helping create standard metadata. At the end of the two open calls the data will be

deposited with COMFOCUS to test and refine whether the harmonised measures have enabled ease of integration and data-sharing across the many different projects. The data will also be deposited within an open repository for other researchers to use.

There are many challenges and lessons learned with integrating data design, capture and storage across countries, equipment setup and disciplines. As such, in many areas, COMFOCUS acted as a first step, where we move as close to harmonisation as possible within the limitations of current setup and without placing heavy burden on researchers. These difficulties are more pronounced in the emerging measures due to the wider array of possible uses and setups.

Within the making of the documentation, even the core researchers have benefited from new insight from different areas. The COMFOCUS fellows have not yet had their research visits, but they are already discussing the benefits of the created structure in terms of clarity and design.

In conclusion, COMFOCUS attempts to harmonise protocol, metadata and measures in food consumer science to better integrate this research area. It hopes to then allow well-documented open data to be available to all researchers, and it is this phase it is now entering with the Open Call 2 COMFOCUS Fellows generating data at their host institutes.

(10) Co-developing open-source optimisation software for better decision-making in energy systems and pharmaceuticals
Michael Short; School of Chemistry and Chemical Engineering, University of Surrey

Contact information:

Michael Short

Senior Lecturer

Faculty of Engineering and Physical Sciences

m.short@surrey.ac.uk

<https://orcid.org/0000-0002-9719-3368>

Abstract

My group develops decision-support software based on optimisation formulations for sustainability challenges in energy, chemicals, and supply chains. During my postdoc at Carnegie Mellon University, I was lead developer of KIPET, an open-source package for kinetic parameter estimation from spectroscopic data. The package was co-developed with chemists at Eli Lilly, Pfizer, and Dow, with manuals built automatically from well-documented code (using Sphinx in Python) and made available via read-the-docs. The accompanying articles used template examples, meaning the work was fully reproducible[1,2].

Since joining Surrey, my group has built on these practices. An example is our easy-to-use, open-source regional energy planning software (DECO2) supported by 2 British Council projects. The software was co-developed with Malaysian and Philippine industry and government partners. It was built by international collaboration, with main developer in Malaysia collaborating closely with myself and Foundation for Public Code to learn best-practice. The software was disseminated via workshops and seminars, with open code and databases[3,4,5]. Since, the developer secured a position at Oxford due to showcasing his work via his open portfolio.

My group publishes open-access papers with accompanying code hosted on open repositories as standard process[6,7,8]. This leads to more trust in our work, make collaborations easy to begin and maintain, and saves time when new group members develop extensions to work. I also teach these open practices to undergrad students at Surrey and to academics via workshops.

The main challenges are making the specialised dependencies and commercial solvers accessible for users. This may lead to frustrations for inexperienced users.

Furthermore, since we do not develop GUIs, users (industrial and policymakers) may be disinterested despite the number of sample problems and documentation.

Overall, we believe that open research practices are worth the additional effort due to large increases in impact, and benefits that increase communication and collaboration.

References:

- [1] Schenk, C., Short, M., Rodriguez, J.S., Thierry, D.M., Biegler, L.T., Garcia-Munoz, S., Chen, W., 2020, Introducing KIPET: A novel open-source software package for kinetic parameter estimation from experimental datasets including spectra. *Computers & Chemical Engineering*, 134, 106716.
- [2] Short, M., Biegler, L.T., Garcia-Munoz, S., Chen, W., 2020, Estimating variances and kinetic parameters from spectra across multiple datasets using KIPET, *Chemometrics and Intelligent Laboratory Systems* 203, 104012.
- [3] Nair, P.N.S.B., Tan, R.R., Foo, D.C.Y., Gamaralalage, D., Short, M., 2023, DECO2— An Open-Source Energy System Decarbonisation Planning Software including Negative Emissions Technologies, *Energies*, 16(4), 1708.
<https://doi.org/10.3390/en16041708>
- [4] Hameed, G., Nair, P.N.S.B., Tan, R.R., Foo, D.C.Y., Short, M., 2023, A novel mathematical model to incorporate carbon trading and other emission reduction techniques within energy planning models, *Sustainable Production and Consumption*, 40, 571-589. <https://doi.org/10.1016/j.spc.2023.07.022>
- [5] https://openresearch.surrey.ac.uk/esploro/outputs/99769066302346?institution=44SUR_INST&skipUsageReporting=true&recordUsage=false
- [6] De Mel, I.A., Bierkens, F., Liu, X., Leach, M., Chitnis, M., Liu, L., Short, M., 2023, A decision-support framework for residential heating decarbonisation policymaking, *Energy*, 126651. <https://doi.org/10.1016/j.energy.2023.126651>
- [7] De Mel, I.A., Klymenko, O.V., Short, M., 2024, Complementarity reformulations for the optimal design of distributed energy systems with multiphase optimal power flow, *International Journal of Electrical Power & Energy Systems*, 155(B), 109610.
- [8] De Mel, I.A., Klymenko, O.V., Short, M., 2024, Discrete optimal designs for distributed energy systems with nonconvex multiphase optimal power flow, *Applied Energy*, 353, 122136.

(11) Open Research Practices in Astronomy: Insights from the DELVE and SMASH Surveys

Noelia E. D. Noël¹, Mahdiah Navabi¹, Joanna Sakowska¹

¹University of Surrey

Contact information:

Noelia E. D. Noël

Senior Lecturer

Faculty of Engineering and Physical Sciences

n.noel@surrey.ac.uk

<https://orcid.org/0000-0002-8282-469X>

Abstract

The DECam Local Volume Exploration Survey (DELVE) and the Survey of the MAGellanic Stellar History (SMASH) seek to understand the faintest galaxies in the Universe offering valuable insights into various astrophysical phenomena. The adoption of open research practices in DELVE/SMASH was driven by a commitment to transparency, accessibility, and the collaborative expertise of the scientific community. Everyone is welcome to participate in projects and to write papers using the data; there is no requisite to be listed as a co-investigator on the initial DELVE/SMASH proposals to get involved. New members are welcome and can join either as junior participants (undergraduate students, PGRs, or postdocs) or through the approval of the Management Committee. The process to join is open (https://delve-survey.github.io/docs/DELVE_PolicyGuidelines.pdf).

The data are publicly available through the NOIRLab Astro Data Lab via (i) the Data Lab Table Access Protocol (TAP) service that gives an access layer for the catalogs' database tables; (ii) the Data Lab Query Manager, which provides a Python API to Data Lab database services including anonymous and authenticated access, asynchronous queries, personal database storage, and storage; (iii) the Data Lab Simple Image Access (SIA) service that provides a fast way to retrieve cutouts from images processed with the community pipeline; and (iv) the Data Lab Jupyter Notebook server (user-authenticated service) that contains examples of how to access and visualize the catalogs.

Establishing data standards and user-friendly tools streamlined the research process and led to more than 100 peer-reviewed publications since the inception of the surveys. To amplify the dissemination of our findings, our research papers are

consistently uploaded to the arxiv.org preprint server, a pioneering platform that played a pivotal role in ushering open access into the scientific landscape.

Although issues concerning data quality and cross-platform compatibility occasionally surface, we meet these challenges through the implementation of robust quality control measures and the development of user-friendly software tools. The collaborative spirit of open research in the SMASH and DELVE consortiums has consistently proven to be key in tackling these challenges effectively.

Lessons learned from my experience include the importance of clear data documentation, slack channels to share, and the necessity of fostering an open research culture.

(12) Open research on progressive collapse in building structures

Giacomo Caredda^{1,2}, Nirvan Makoond², Manuel Buitrago², Juan Sagaseta¹, Marios Chryssanthopoulos¹, Jose M. Adam²

¹University of Surrey, ²Universitat Politècnica de València

Contact information:

Giacomo Caredda

Civil Engineer

Faculty of Engineering and Physical Sciences

gc00890@surrey.ac.uk

<https://orcid.org/0000-0002-6888-0409>

Abstract

OUR AIM

Extreme events, as well as design and construction errors, often cause local damage to building structures, which can lead to progressive collapse. Some progressive collapses, such as the collapse of the World Trade Center towers in New York (2001), can have a profound impact worldwide. We can learn from past failures and existing databases of building collapses or partial collapses have identified key factors contributing to risk of disproportionate collapse. However, little information is available in these databases about the propagation of failure in relation with the triggering event. Our research focused on creating an open-access database of structural collapses to show, for each case study, a systematic treatment of initial failures, collapse propagation, and related consequences so that future studies could build upon this knowledge.

OUR METHODOLOGY

The database gathered the analysis of 40 partial or complete structural collapses. Each case contained specific data with contextual information, structural typology, geometry, failure, and economic and human consequences. For each case study included in the database, we carried out an in-depth study of multiple academic and forensic engineering approaches. As can be expected, the available information for each case varied significantly, depending mostly on the nature and number of past studies available. As such, for greater transparency towards potential future users of this database, we decided to introduce a field containing an evaluation of a utility factor of the information collected.

THE OUTCOME

This research is part of an ongoing PhD collaboration between University of Surrey and Universitat Politècnica de València UPV for project “Endure” (ERC Consolidation Grant led by UPV). The research was disseminated through the publication of a journal paper [1] and database [2] in open-access spaces, hoping to provide a useful information tool for practitioners and researchers in the field of progressive collapse.

[1] Caredda G, Makoond N, Buitrago M, Sagaseta J, Chryssanthopoulos M, Adam JM. Learning from the progressive collapse of buildings. *Developments in the Built Environment* 2023;15:100194. <https://doi.org/10.1016/j.dibe.2023.100194>.

[2] Caredda, G., Makoond, N., Buitrago, M., Sagaseta, J., Chryssanthopoulos, M., & Adam, J. M. (2023). Database of past progressive collapse of buildings (Versione 2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8131026>

(13) Ampere-hour-scale soft-package potassium-ion hybrid capacitors enabling 6-minute fast-charging

Yi Gong; Advanced Technology Institute

Contact information:

Yi Gong

Postgraduate Researcher

Faculty of Engineering and Physical Sciences

Yg00541@surrey.ac.uk

<https://orcid.org/0009-0005-4602-8177>

Abstract

In this work, a holistic optimization for potassium-ion-based devices was conducted to achieve fast chargeable 1Ah PIHCs pouch cells with high energy density, which possess superb comprehensive

performance of high specific energy of 140Whkg^{-1} (based on the whole mass of the device), excellent performance (fully charged in 6 min) and good cycling stability. This breakthrough demonstrates significant potential in extending the lifespan and enhancing the charging speed of electric vehicle (EV) batteries.

This outcome has been published in an OA journal (Nature Communication) to make it more accessible to other researchers.

While conducting this research, we needed to refer to some previous work, but the challenge was that we were not able to download these papers. To overcome this difficulty, we send emails to the authors to request these papers.

Based on this experience, I learnt that sharing is very important to the advancement of technology. The Open Research should be greatly encouraged.

(14) Collocaid

Ana Frankenberg-Garcia; University of Surrey

Contact information:

Ana Frankenberg-Garcia

Associate Professor of Translation Studies

Faculty of Arts and Social Sciences

a.frankenberg-garcia@surrey.ac.uk

<https://orcid.org/0000-0001-9623-7990>

Abstract

The ColloCaid project (AHRC) investigated how to help writers expand their English academic writing vocabulary with minimal disruption. It compiled a very large database of over 32,000 effective word combinations (called collocations) sourced from expertly written formal/academic texts and edited by lexicographers, then developed a prototype that integrates interactive collocation assistance into a text editor to help people improve word pairings as they write. The prototype is free and can be accessed at <https://collocaid.uk/prototype/editor/public/>. Well over 8500 users from all over the world have registered to use the tool independently or encouraged by academic writing tutors at various universities across the globe. We also shared the following datasets openly:

Frankenberg-Garcia, A., Rees, G. (2021). "ColloCaid Academic Collocation Errors and Other Problems". Available at <https://doi.org/10.6084/m9.figshare.13640624> (355 downloads, 2084 views)

Sharma, N., Butcher, P. W. & Roberts, J.C. (2021). "ColloCaid Code". Available at: <https://doi.org/10.6084/m9.figshare.14170988.v1> (134 downloads, 557 views)

Frankenberg-Garcia, A., Rees, G., & Lew, R. (2020). "ColloCaid Sample Data". Available at: <https://doi.org/10.6084/m9.figshare.13028207> (170 downloads, 1322 views)

We also shared the full collocation data with the European Lexicographic Infrastructure (ELEXIS) project, funded Horizon 2020, who have reused our data in Word of Games app.

Ten open-access research publications related to the project are available at <https://www.collocaid.uk/publications/> and have been attracting a growing number of citations by researchers in lexicography and writing pedagogy.

We have recently applied for funding to extend the benefit of ColloCaid by:

1. Transforming the ColloCaid prototype into free add-ins compatible with mainstream text editors in response to user requests.
2. Raising professional/academic writers' awareness of collocation and collocation assistance and empowering writers and writing instructors in an age of AI writing.

(15) Nutrient recovery from domestic wastewaters for circular economy and sustainable development in cities and communities

Carla Mae Pausta¹, Dr Devendra Saroj¹

¹School of Sustainability, Civil and Environmental Engineering

Contact information:

Carla Mae Pausta

Postgraduate Researcher

Faculty of Engineering and Physical Sciences

c.pausta@surrey.ac.uk

<https://orcid.org/0009-0006-8177-8018>

Abstract

The PhD research would particularly address the increasing challenges on eutrophication in water resources, wastewater management, and depletion of non-renewable phosphorus for agri-food systems in the context of a sustainable development and circular nutrient economy. Nutrient recovery from wastewaters have been studied in some countries in Europe but has never been explored in developing countries like the Philippines, where the adverse impacts are also evident. Utilising the past and existing research on nutrient recovery systems to improve and develop a technology that is sustainably feasible for another geographical boundary leads to a positive impact in promoting the open research culture. In this research, a pilot-scale nutrient recovery reactor was installed at a farm in the Philippines, wherein wastewater is processed for treatment and to recover an alternative fertiliser for local crop production. The data and results from this research, including reactor operations and life cycle assessment, were submitted to open access journals, presented to conferences, and promoted in social networking sites for general public awareness and replicability. Consequently, the technology was also presented to the relevant stakeholders, including the local communities, through workshops and focus group discussions for research outcomes dissemination. In general, practicing open research would lead to further research improvements and implementation of appropriate solutions to the identified challenges.