Festival of Digital Research, Innovation & Scholarship 2026

Programme and Poster Information

UCL

Festival of Digital Research, Innovation & Scholarship, 2026

Programme

Arrival & Coffee (Networking)

Time: 9:00– 9:30
Location: UCL Roberts Building Foyer, G02

Session 1

Time: 09:30–11:05
Location: UCL Roberts Building, G06 Lecture Theatre
Session Chair: Prof James Hetherington, Director of UCL Advanced Research Computing Centre (ARC)

Running Order

Time	Activity	Speaker
09:30–09:45	Welcome & ARC Opening	Dr Claire Ellul & Prof James Hetherington
09:45–10:20	The CCMI CDT – Graduate Training on the Interface of Software, Data, and Computation	Prof Timo Betcke
10:20–11:05	Speed Postering Session 1	Chaired by Dr Claire Ellul

Welcome & ARC Opening

Speakers

Dr Claire Ellul
Reader in Geographical Information Science, ARC Community PO

Prof James Hetherington
Director of UCL Advanced Research Computing Centre (ARC)

Keynote Talk

The CCMI CDT – Graduate Training on the Interface of Software, Data, and Computation

Speaker

Prof Timo Betcke
Professor of Computational Mathematics, UCL
Director, EPSRC CDT in Collaborative Computational Modelling at the Interface (CCMI)

Abstract

The EPSRC Centre for Doctoral Training in Collaborative Computational Modelling at the Interface (CCMI) is a newly created Doctoral Training Centre between UCL and Imperial College. Its vision is to create a modern graduate training programme that lifts Research Software Engineering from being a Professional Service to being a core academic activity of a Computational Researcher. In our first cohort we had 20 students and are soon welcoming 16 more students for the second cohort, aiming to train over 70 students in the lifetime of the CDT. In this talk I will give an overview of what the CDT is and how we try to tackle the core training challenges on the interface of data, software, and computation.

Speaker Bio

Prof Timo Betcke is Professor of Computational Mathematics in the Department of Mathematics at UCL, and an Academic Fellow of the UCL Advanced Research Computing Centre (ARC). He works at the interface of computational mathematics, research software, and applications in the physical and engineering sciences. He is the Director of the EPSRC Centre for Doctoral Training in Collaborative Computational Modelling at the Interface (CCMI) and has been substantially funded by EPSRC as part of the ExCalibur and other software programmes.

Speed Postering Session 1

Session Chair

Dr Claire Ellul

Coffee Break & Posters (Networking)

Time: 11:05– 11:35
Roberts Foyer, G02

Session 2

Time: 11:35–12:50
Location: UCL Roberts Building, G06 Lecture Theatre
Session Chair: Dr Owain Kenway, Principal Research Infrastructure Developer, UCL Advanced Research Computing (ARC)

Running Order

Time	Activity	Speaker
11:35–12:05	Keynote— DataVita: *AI Growth Zones and Secure, Sustainable, Sovereign Capabilities*.	Richard Woodfield, DataVita
12:05–12:50	Speed Postering Session 2	Chaired by Dr Claire Ellul

Keynote Talk

DataVita: AI Growth Zones and Secure, Sustainable, Sovereign Capabilities.

Speaker

Richard Woodfield, DataVita

Abstract

The AI Growth Zone is a key standard for the kind of Datacentre that the UK needs. The lack of capacity is becoming an urgent issue, at the same time as the environmental and financial economics of High-Power Compute to support machine learning and AI applications become more complex. Recent political/security interventions from the USA are adding to this phenomenon.

Speaker Bio

I am a business leader with over three decades of experience driving tech start-ups and successful exits globally. My foundational expertise in AI stems from early academic work in the 1980s, exploring neural networks through mathematical and philosophical research. Today, based in Scotland with DataVita, I head solution development in support the Lanarkshire AI Growth Zone. My focus is on delivering sustainable, sovereign high-performance-compute platforms to provide access to researchers and the broader market.

Speed Postering Session 2

Session Chair Dr Claire Ellul

Session Chair

Dr Claire Ellul

Lunch break & Posters (Networking)

Time: 12:50– 13:50
UCL Roberts Foyer, G02

Session 3

Time: 13:40– 15:10
Location: UCL Roberts Building, G06 Lecture Theatre
Session Chair: Dr George Svarovski, Principal Research Software Engineer, UCL Advanced Research Computing (ARC)

Running Order

Time	Speaker	Talk
13:50 – 14:20	Prof David De Roure	Trusted Research Environments Software Stewardship (TRESS): Learning How to Share and Collaborate when Security Matters
14:20 – 14:50	Lily Venables & Alice Nerurker	Accelerating Action on the SDGs through Business / Academic Collaboration
14:50 – 15:20	Eric Fauvet	The Journey to National Compute Resource (NCR)
15:20 – 15:50	Oscar McDonald & Sam Balch	Science-led Urban Design and Planning: Applying Research in Industry

Talk 1

Trusted Research Environments Software Stewardship (TRESS): Learning How to Share and Collaborate when Security Matters

Speaker

Prof David De Roure
Professor of e-Research, University of Oxford

Abstract

“Trusted Research Environments Software Stewardship (TRESS) – Learning How to Share and Collaborate when Security Matters” examines a current cross-institutional partnership between the University of Oxford and UCL to co-develop secure research infrastructure. With one of the key elements being Oxford’s adoption of UCL’s TRE codebase through a ‘shared-source’ approach, the talk highlights how both institutions are working together to balance collaboration with strict security requirements. It explores the practical realities of shared development, including governance, trust-building, and coordinating technical work across organisations. The session reflects on challenges encountered, lessons learned, and emerging good practices for stewarding TRE software as critical infrastructure. Looking ahead, it considers how this partnership can evolve to support wider collaboration, reduce duplication, and strengthen secure, data-driven research across the sector.

Speaker Bio

Prof David De Roure is Professor of e-Research at the University of Oxford and Academic Director of Digital Scholarship. His work spans digital scholarship, research infrastructure, software sustainability, data stewardship and AI security. He has played leading national roles across UKRI, the Software Sustainability Institute, DARE UK and Digital Research Infrastructure initiatives.

Talk 2

Accelerating Action on the SDGs through Business / Academic Collaboration

Speakers

Lily Venables
UN Global Compact Network UK

Alice Nerurker
UCL Grand Challenges / SDG Reactor

Abstract

This session will introduce the UN Global Compact UK Network and its role in supporting businesses to deliver the Sustainable Development Goals (SDGs), with a particular focus on responsible business practice, data and digital innovation. It will outline collaboration to date with UN Sustainable Development Solutions Network UK (hosted at UCL), including joint work through the SDG Reactor, which has acted as a platform for bringing together academia, business and policy to address complex sustainability challenges through applied, challenge-led research. Looking ahead, the session will explore opportunities to deepen collaboration around data and digital, including responsible AI, data for sustainability, and approaches to measuring and scaling impact. It will highlight the role of universities as testbeds for innovation and the value of cross-sector partnerships in translating research into practice. The discussion will also point to wider opportunities for engagement, including student internships, knowledge exchange, joint projects and further convening activity.

Speaker Bios

Lily Venables is a Senior Project Manager at the UN Global Compact Network UK, where she leads work across governance and the Sustainable Development Goals (SDGs). Her role focuses on supporting businesses to align with the SDGs through practical tools, partnerships and programmes, helping translate sustainability ambition into meaningful action. Lily has contributed to initiatives such as sustainability toolkits and sector-focused reports, and works closely with partners across academia, industry and civil society to drive responsible business practice. She brings experience in stakeholder engagement, programme delivery and cross-sector collaboration, with a particular interest in how governance, data and innovation can support sustainable development outcomes.

Alice Nerurker is part of the UCL Grand Challenges team and serves as Coordinator of the SDG Reactor, where she supports the development of collaborative, challenge-led research partnerships focused on advancing the Sustainable Development Goals. Her work centres on convening academics, businesses and external partners to co-create practical solutions to complex sustainability challenges, with a particular emphasis on data, digital innovation and interdisciplinary approaches. Alice plays a key role in fostering links between UCL, SDSN UK and external partners, including the UN Global Compact UK Network, helping to translate research insights into real-world impact through initiatives such as the SDG Reactor.

Talk 3

The Journey to National Compute Resource (NCR)

Speaker

Eric Fauvet
Senior HPC & AI Sales Specialist, Hewlett Packard Enterprise (HPE)

Abstract

In his talk, Eric will reflect on his engagement to date with UCL and will address HPE's approach in proposing a response to the UKRI National Compute Resources (NCRs) hosted by UCL. This £19.5m supercomputing facility, one of four UK facilities will provide compute power not only in the hands of researchers but also undergraduate students!

Speaker Bio

Eric Fauvet is a Senior HPC& AI Sales Specialist working for HPE - he has an engineering background and led Operational IT teams in Research, Commercial, Central and Local governments organisations before joining the vendor landscape. He is married with two children and lives in Greater London.

Talk 4

Science-led Urban Design and Planning: Applying Research in Industry

Speakers

Oscar McDonald
Urban Intelligence Lab (Sidara)

Sam Balch
Director, UCL Grand Challenges

Session Summary

This will cover a few key things:
What is changing in digital planning and spatial analysis (spoiler – a lot, but maybe not where everyone thinks). Including:
The potential for barriers to entry to academic research to be lowered e.g. its easier to find appropriate cutting-edge research even if it isn't so much easier to use or develop (we still very much need academia)
The potential for LLMs to improve public engagement and buy-in, through allowing more nuanced input across larger populations (most major problems are political as much as technical)
The potential for remote sensing and AI assisted coding to improve equality in how cutting-edge research is implemented e.g. addressing the gulf between the places that really need spatial intelligence and the places that can afford to have it
The potential for sophisticated slop - all of the above need addressing with care and intelligence – we are seeing a huge rise in the ability to bullshit in some depth and this is a serious issue
Why this makes bridging academia and industry even more important
What initiatives are being undertaken to do this (UIL / UCL partnerships and others, the

Grand Challenges)
What UIL is doing specifically (and what others are doing)

Speaker Bios

Oscar McDonnald: I'm an architect, spatial analyst, and urban planner. I've spent most of my career at the intersection of academia and industry.
Starting out I founded a non-profit focusing on evidence-based planning in the Maldives, based on my masters thesis project. This has led to working recently as an expert advisor on the Maldives National Development Plan, working on spatial analysis and digital platforms.
I was until recently at Space Syntax (a UCL spin-out) where I was head of the Strategic Design Studio, working on large-scale regional and city plans. This focused on taking Space Syntax techniques, developed through UCL and in-house R&D, and applying them to solve big spatial problems.
I have recently joined the Urban Intelligence Lab which I'm co-leading. This is a standalone company within a wider group (Sidara) that forms one of the biggest built-environment consultancy groups in the world. We will be working to link this network and others to the best science-based analysis and design techniques from both UCL and MIT, particularly focusing on implementation in developing contexts.

Sam Balch is the Director of UCL Grand Challenges programme. This is a key academic initiative that supports challenge-led, interdisciplinary research and innovation at UCL and with partners around some of the biggest issues facing society.

Coffee Break

Time: 15:10– 15:40
Location: UCL Roberts Foyer, G02

Session 4

Time: 15:40–16:30
Location: UCL Roberts Building, G06 Lecture Theatre
Session Chair: Dr Claire Ellul

Running Order

Time	Activity	Speaker
15:40–16:10	Keynote— Ordnance Survey: From shared photogrammetry roots to AI-era collaboration: 75 years of OS–UCL partnership shaping the future of geospatial research.	Andy Gardiner & Dr Stefano Cavazzi
16:10–16:30	Closing Remarks & Poster Prizes	Dr Claire Ellul & Prof James

Keynote Talk

From Shared Photogrammetry Roots to AI-era Collaboration: 75 Years of OS–UCL Partnership Shaping the Future of Geospatial Research

Speakers

Andy Gardiner
Head of Research, Ordnance Survey

Dr Stefano Cavazzi
Principal Innovation & Research Scientist, Ordnance Survey

Abstract

Celebrating 75 years of collaboration between Ordnance Survey (OS) and UCL, this talk traces a journey from shared photogrammetry roots to today's cutting edge, AI-driven research. It brings to life how OS, as the national mapping agency of Great Britain, has worked hand in hand with academia to push the boundaries of geospatial science. Through case studies spanning PhD research, innovative collaborations, and real-world proofs of concept, the session highlights tangible impact and looks ahead to how this partnership will continue shaping the future of mapping and digital innovation.

Speakers Bio

Andy Gardiner is Head of Research at Ordnance Survey with a career spanning 25 years at OS, including photogrammetry, remote sensing consultancy, and computer vision-based research. He holds a bronze 50m swimming badge and is by far the least decorated member of his team.

Dr Stefano Cavazzi is Principal Innovation & Research Scientist at Ordnance Survey, leading AI-driven automation research. With 20 years' experience in R&D he delivers customer focused solutions. He holds a PhD in Spatial Statistics, is an Honorary Lecturer at UCL, and co-chairs the Arthritis UK research committee driving real-world impact.

Closing Remarks & Poster Prizes

Session Chair

Dr Claire Ellul

Celebratory Drinks & Nibbles (Networking)

Time: 16:30– 19:00
Location: UCL Roberts Foyer, G02

Posters

Speed Postering Session 1

Poster Running Order

Poster Abstracts and Lead Presenter Information. PhD Students and Postdocs are eligible for poster prizes.

Name	Last Name	Poster Title & Abstract	Affiliation
Ross	**Ah-Weng**	**Statistical Degradation of Dynamical Systems on Computers (initially: Birkhoff Averages in Finite Precision: Can we Trust Statistics from Numerical Trajectories?)**	**Phd Student**
Many physical systems of interest, such as the climate, the solar system, and molecular dynamics, are chaotic; due to sensitivity to initial conditions, we cannot accurately predict trajectories over long time horizons. Rather, we estimate invariant statistics of these systems from time averages of observables over simulated trajectories. Yet under floating-point arithmetic, the continuum is replaced by a finite, discrete state space. As a result, every orbit is eventually periodic, and the finite-precision model is not chaotic. Nevertheless, in practice one often relies on numerical simulations for accurate statistical predictions. We study truncated Birkhoff averages of chaotic dynamical systems under finite-precision arithmetic through the lens of functional graphs, and quantify how decreasing floating-point precision degrades the accuracy of statistics. We also assess alternative approaches, including stochastic rounding and analogue computing architectures.			
Shany	**Neeman**	**Learning PDE solutions through Chebychev encoding**	**PhD Student**
We (Shanny Neeman and Anees Hussain) present a framework for solving partial differential equations using physics-informed neural networks (PINNs) defined in Chebyshev coefficient space. By representing solutions spectrally, we enable stable differentiation through recurrence-based Chebyshev derivatives and incorporate boundary conditions with minimal pointwise evaluation. To address slow and unstable convergence in PINNs, we investigate preconditioning strategies based on geometric optimisation, with a focus on Natural Gradient Descent. These methods account for the mismatch between parameter-space and function-space geometry, improving optimisation efficiency and training stability. Our results demonstrate that combining spectral representations with geometry-aware optimisation provides a robust approach for learning PDE solutions and offers a foundation for future operator-learning extensions.			

Lucas	**Pigott**	**Bayesian-Optimal Sensor Placement for Latent Extrema Recovery**	**PhD Student**
In nuclear fusion experiments, components fail at localised hotspots of an underlying field. However these fields aren't observed directly, but through heterogeneous sensors, each sensor-type intrinsically biased and acting through a different observation operator that complicates the inverse problem. Recovering the latent extrema requires choosing where to place a whole set of sensors before data is measured. We address this with an information-theoretic acquisition function, quantifying mutual information between fantasised observations and sampled latent maxima, based on a surrogate Gaussian Process. We build on GIBBON (Moss et al., 2021), extending its closed-form batch acquisition function to incorporate indirect, multi-modal observations through linear operator W. This enables a principled batch combining cameras (high information, high bias, blurred) with thermocouples (pointwise and accurate). The GIBBON acquisition function remains a valid lower bound, with the looseness of the underlying approximations yet to be quantified, a key direction for future work.			
Louis	**Clare**	**Analytical and Computational Modelling of Active Stress in Organoids**	**PhD Student**
Skye	**Purchase**	**Physics-aligned interpretable diffusion**	**PhD Student**
In recent years there has been increasing research into using machine learning models to aid in solving physics simulation problems. This has resulted in techniques such as physics informed neural networks and neural ordinary differential equations. Alongside this, progress in generative models has resulted in the denoising diffusion model which has shown impressive generative fidelity. Existing research exploring the use of diffusion models for physics simulation has restricted the use of diffusion models to simple decoders or samplers without physical constraints. In this work we explore how to incorporate the known dynamics of a system into a diffusion model whilst allowing the model to infer the unknown dynamics. In this way the model samples solutions directly whilst being guided by our understanding of the system. This improves the accuracy of the diffusion model whilst reducing hallucinations as well as serving as a powerful method for data-driven physics simulation.			
Lewis	**O'Donnell**	**Foundation Models For Fusion Energy**	**PhD Student**
Ammar	**Ali**	**Using LLM Agents for Chemical Experiment Scheduling and Planning Assistance**	**PhD Student**
Chemical R&D pipelines are sequences of interdependent experiments. Any experiment may fail, cancelling the downstream work that depended on it, while equipment and scientists impose hard resource constraints. Deciding what to run, when, and on which resource, and revising that plan as failures arrive, is a stochastic scheduling problem with conditional task cancellation. We ask whether LLM agents can act as competent schedulers and online re-planners for this class of problem, and how to make their plans trustworthy. We introduce a gym-like simulation environment that models an R&D lab as a dependency graph of experiments with stochastic pass/fail outcomes and resource contention. An LLM advisor observes the lab state, proposes a schedule, observes outcomes, and replans. Because LLM-proposed schedules need not be feasible, we pair the agent with a symbolic validator.			
Charlotte	**Savage**	**Deep Learning approaches for Image Quantification in Brain-focused Positron Emission Tomography**	**PhD Student**
Dedicated brain PET systems like the Positrigo NeuroLF lack an integrated CT, precluding conventional attenuation correction (AC) and biasing quantitative biomarkers (e.g. SUVr) used in Alzheimer's diagnosis. We investigate deep learning for pseudo-CT synthesis, deriving an attenuation			

map (μ-map) from Non-AC-PET, Dixon MRI, and 2D topogram inputs. A U-Net trained on 79 healthy subjects shows that a multi-input model after only 100 epochs approaches a NAC-PET-only baseline on μ-map MAE (0.0089 vs 0.0066 cm^{-1}), despite far less training. Future work will extend training, improve multimodal registration, and implement GAN and diffusion architectures, with evaluation against CT-AC reconstructions for end-to-end biomarker accuracy.

Maciej	**Kaczorek**	**Sound reconstruction through neural activity patterns.**	**PhD Student**

Understanding how the brain encodes sound is central to auditory neuroscience, with direct implications for hearing loss research. A natural approach is to model the forward transformation from acoustic stimuli to neural activity, then invert it to reconstruct sound. Unlike directly learning a neural-to-sound mapping, model inversion imposes structural constraints inherited from auditory processing - though the problem remains ill-posed due to the forward operator's large null space.
We build on ICNet, a convolutional encoder–decoder predicting multi-unit activity in the inferior colliculus of anaesthetised gerbils. Layer-by-layer inversion reveals that strided convolutions introduce aliasing artefacts that degrade invertibility. We replace these with causal blur pooling - a parameter-free low-pass filtering step - yielding a better-conditioned forward operator.
We evaluate 4 reconstruction strategies of increasing complexity. A naive inverse crime analysis confirms the null space as the primary obstacle. Data consistency across all methods, proves that the problem is highly ill-posed even under an inverse crime scenario and classical regularisation methods provide little to no benefit in contrast to the causal blur pooling.

Joe	**Egan**	**Contur: Getting more insight from hard-won physics data**	**PhD Student**

The Large Hadron Collider at CERN is recording the highest-energy collisions ever made in a lab. It is probing new aspects of the Standard Model of particle physics, and searching for extensions to the theory which might address some of the fascinating questions, for example the nature of dark matter, that it leaves unanswered.
The resulting dataset will stand unique for generations. It is essential that it remains reusable and accessible, so that new theories and improved calculations can be confronted with data even if they were not conceived of at the time the measurements were made.
This is the challenge addressed by Contur (Constraints On New Theories Using Rivet). Originating at UCL and leveraging the widely-used Rivet analysis framework, this open-source community-led software allows rapid statistical testing of new models against the huge range of measurements being produced at CERN.

Jocelyn	**Japnanto**	**Listening to the Ocean: Whale Songs Revealed by Earthquake Sensors and Deep Learning**	**PhD Student**

Fin whales are among the largest and loudest animals on Earth, yet their remote offshore habitats make them remarkably difficult to monitor at scale. We repurpose seismic sensor arrays, originally deployed for geophysical research, to detect and catalogue fin whale calls across the eastern North Atlantic. A deep learning model trained on less than 0.1% of the data produces over 6.3 million detections from 380,000 hours of recordings, covering a spatial footprint of 2,000,000 km^2. This is the largest catalogue of fin whale calls ever assembled, at a scale simply unachievable through dedicated monitoring alone. Analysis of this catalogue reveals seasonal and spatial patterns in calling behaviour, offering an unprecedented window into the breeding, migration, and vocal lives of these vulnerable whales, made possible by combining repurposed infrastructure with large-scale automated analysis.

Jiwen	**Jiang**	**Deep Learning of G-Quadruplexes**	**PhD Student**

G-quadruplexes (G4) are polymorphic higher-order nucleic acid structures implicated in diverse regulatory processes. Their conformational heterogeneity, arising from variations in strand orientation and guanine quartet geometry, poses a significant challenge for predicting potential G4-forming sequences and rationalizing their biological roles. Moreover, the intrinsic thermodynamic stability of G4 stem often obscures differences between polymorphic states when analysed by conventional structural and biophysical methods. Here, we employ a deep learning framework based on a Convolutional Variational Autoencoder (CVAE) to investigate and classify G4 structural dynamics. The CVAE captures sequence-, ligand-, and topology-dependent features of G4s and projects them into a low-dimensional latent space, thereby discretizing similarities and differences among distinct polymorphs. This representation enables the unsupervised grouping of G4 structures and provides a quantitative basis for comparing their dynamic landscapes. Our results highlight the utility of deep generative models in decoding subtle structural variations of non-canonical nucleic acid architectures and establish a scalable strategy for integrating dynamics, sequence, and ligand context in the study of G4 polymorphism.

Zijian	**Zhou**	**Hybrid Gated Fusion: A Multimodal Deep Learning Framework for Protein Function Annotation**	**PhD Student**

Protein function prediction is essential for interpreting genomes, understanding cellular systems, and identifying therapeutic targets. However, experimentally validated functional labels lag far behind the rapid growth of protein sequence data. Improving accuracy increasingly requires combining complementary sources of evidence, including amino acid sequence, three-dimensional structure, textual descriptions of protein properties, and protein–protein interaction networks. Yet multimodal prediction remains challenging: existing methods assume all modalities are available at inference, whereas real-world biological data are often incomplete and uneven. Simple fusion strategies may underuse cross-modal information, while more expressive schemes are harder to train robustly. To address these challenges, we present Hybrid Gated Fusion, a multimodal framework for Gene Ontology prediction that combines bilinear gated early fusion with consistency-aware residual late fusion. The design allows the model to assess both each modality's contribution and its agreement with other evidence sources, while remaining robust when inputs are missing. On the CAFA3 benchmark, it achieves state-of-the-art performance in Biological Process and Cellular Component and remains competitive in Molecular Function.

Louie	**Destouches**	**Using machine-learning to explore microbial community cybernetic control for bioproduction applications**	**PhD Student**

Microbial bioproduction offers sustainable alternatives to current methods of production i.e. though biofuels and biopolymer synthesis, or the metabolism of waste streams. Microbial communities can share metabolic load and utilize different sugars simultaneously. They could thus increase bioprocess efficiency and open up new pathways of production. However, due to differences in growth rates or competitive interactions, maintaining stable community composition is challenging, which hinders their widespread application. External environmental control is a promising strategy to reliably control population levels: dynamics can be controlled over time, it requires little to no genetic engineering and is easily integrated with automation. Deep Reinforcement Learning is a type of machine learning that learns what actions to take given a certain goal. I am using DRL to explore environmental control strategies for microbial communities. The agent is trained in silico and in vivo using small open-source bench top bioreactors.

Kewei	**Zhu**	**MOF in Water: An Interpretable Rule-Based Framework for Screening Water-Stable Metal-Organic Frameworks from Systematic Chemical Nomenclature**	**PhD Student**

Hydrolytic stability remains a key limitation for the practical deployment of metal-organic frameworks (MOFs), yet existing screening approaches rely heavily on structural representations that are costly to obtain and difficult to interpret. Here, we present an interpretable rule-based framework that infers stability directly from systematic chemical nomenclature, enabling screening without structural reconstruction. The framework encodes chemically grounded factors, including metal-ligand bonding, coordination environment, and framework connectivity, into a structured scoring scheme applied to a large CSD MOF corpus. The resulting stability assignments reproduce benchmark-level trends observed in the WS24 dataset and identify a chemically enriched subset of water-stable frameworks, which are further organized into confidence tiers for prioritization. The selected candidates exhibit coordination motifs and ligand architectures consistent with established stability principles, while also revealing less conventional stabilization patterns. These results demonstrate that systematic chemical names encode sufficient information for stability inference and provide a scalable, interpretable representation for large-scale materials discovery, with rule-based screening serving as an efficient first-pass filter for downstream modeling and validation.

Yi	**Li**	**SearchMOF: A Literature-to-Structure Retrieval-Augmented Generation (RAG) Pipeline for Metal–Organic Framework**	**PhD Student**

Large language models (LLMs) are increasingly used in metal–organic framework (MOF) research for property prediction, structure generation, and question answering, but they remain prone to hallucinations. Here, we present SearchMOF, a retrieval-augmented generation (RAG) pipeline that combines real-time retrieval from scholarly literature and the Cambridge Structural Database (CSD) MOF subset to generate reference-supported answers without domain-specific fine-tuning. By incorporating current scholarly and structural information at inference time, SearchMOF improves answer grounding and reduces unsupported outputs. On a review-based MOF question–answer dataset, SearchMOF increases the true positive rate from 56.1% to 85.3%, while reducing false positives from 5.5% to 0.6% and true negatives from 21.8% to 6.3%. These results indicate that SearchMOF produces more correct answers while reducing hallucinatory content. Because the framework can be extended beyond MOFs to other structured scientific databases, it offers a practical route toward more reliable LLM-assisted discovery across materials science.

Sotirios	**Andrianos**	**Hydrodynamic and reaction studies in an axially ribbed Taylor vortex reactor**	**PhD Student**

A Taylor vortex reactor with axially ribbed rotor and a 10 mL annulus was characterised for macro-mixing and conversion to construct the reactor model for digital twin development. Macro-mixing was studied via step-input experiments, monitoring Rhodamine B using inline UV–Vis spectroscopy. Residence time distribution (RTD) matched tanks-in-series model predictions, with 3–7 ideal CSTRs. Simulations performed in MATLAB. Space time (1–7 min) had a stronger effect on number of CSTRs than rotor speed (100–300 rpm), supported by partial least squares modelling as well. Conversion was evaluated using n-benzylidenebenzylamine synthesis from benzaldehyde and benzylamine in methanol, a reaction with well-known kinetics. RTD and reaction studies were conducted simultaneously. Conversion was measured by inline Raman spectroscopy monitoring the benzaldehyde peak with adaptive iteratively reweighted penalized least squares baseline correction. Conversions of 55–98% were achieved at 1–7 min residence times and 5–25 °C inlet jacket temperatures.

Yihang	**Fang**	**Systematic Data Generation through automated HTE Acrylamide Coupling Synthesis for future ML reaction optimisation**	**PhD Student**

Acrylamide functionality has gained renewed attention due to their importance as monomer for large-scale polymer production and as covalent drugs. Since machine learning (ML) has emerged as a promising tool to accelerate reaction[LL4.1] condition optimisation for complex chemistry systems, reliable datasets become essential for accurate ML prediction. In this study, we present an automated high-throughput experimentation (HTE) framework designed to generate well-curated, unbiased datasets for acrylamide synthesis that integrates rapid solubility screening as a crucial step to define a feasible concentration window prior to reaction execution. We then use automated protocol control to standardise and record reaction-critical variables which are frequently missing from the literature. Through automated HTE, we build a high-quality acrylamide library synthesis dataset that includes 432 novel reaction combinations with accurate repetition, demonstrating the capacity of HTE to improve reaction reliability and reproducibility. [LL5.1]This comprehensive dataset provides a foundation for subsequent ML-guided reaction condition optimisation.

Casey	**Chen**	**Analysis and prediction of microbial ecological behaviours: predator-prey communities**	**PhD Student**

Natural microbial communities such as our gut microbiome are complex networks of diverse microbial species interacting with one another. To better understand these systems, researchers have built pairwise communities with defined interactions to create simplified abstractions of complex communities using a variety of distinct inter- and intra-cellular mechanisms. It is poorly understood how topologically identical, but mechanistically distinct communities, differ in their fundamental properties and community dynamics.
To address this, we combined in vivo and in silico approaches to characterise and predict community behaviours. Microbial growth data were used to parameterise Ordinary Differential Equation models describing communities with defined interactions. Approximate Bayesian Computation was then applied to explore how different properties of microbes' impact community dynamics. We aim to identify key parameters that determine ecosystem dynamics, and with further research, our models can be applied to understand complex interaction networks in natural communities.

Vaishali Parag	**Waman**	**The Encyclopaedia of protein Domains (TED) from AlphaFold2-predicted structures: expansion of CATH domain space and insights into structural diversity**	**PhD Student**

The Encyclopaedia of protein Domains (TED) from AlphaFold2-predicted structures: expansion of CATH domain space and insights into structural diversity
Lau AM1$, Bordin N2$, Kandathil SM1, Sillitoe I2, Waman VP2&, Wells J2,3, Orengo CO2* and Jones DT1,2*
1 Department of Computer Science, University College London, London, WC1E 6BT, UK
2 Institute of Structural and Molecular Biology, University College London, London, WC1E 6BT, UK
3 Centre for Artificial Intelligence, University College London, London, WC1V 6BH, UK
&Presenting author
$ Authors contributed equally
*Corresponding authors: c.orengo@ucl.ac.uk , dtj@cs.ucl.ac.uk
The AlphaFold Structure Database (AFDB) provides predictions for >214 million three-dimensional structures of full-length proteins. Protein structure is composed of one or more domains. We

present The Encyclopedia of Domains (TED) which is the first structural-based resource to classify domains from the AFDB (https://ted.cathdb.info/).
TED is developed as a collaboration between the Orengo and Jones group (Science, 2024). We used novel deep learning-based methods (Merizo, Chainsaw) and UniDoc, to segment AFDB structures into domains.
TED contains over 370 million protein domains, providing domain structure coverage to over 1 million taxa/species. Notably, 80% of the TED domains exhibit similarities with known superfamilies in the CATH database, expanding the CATH by over 600-fold (https://www.cathdb.info/). We identified unique protein architectures not seen previously (e.g. 11-helix propeller).
TED and CATH annotations are now made available via AFDB. CATH and TED, paves the way to understand biodiversity using protein domains across different taxa and pathogens.

Peng	**Wang**	**Reproducible Collider Physics with Rivet: 20 years of Sustainable Analysis Preservation**	**PhD Student**

Particle physics facilities such as the CERN Large Hadron Collider (LHC) produce vast and complex datasets, requiring sophisticated computational tools to interpret them. A key challenge is ensuring analyses remain reproducible and reusable over time – particularly as theoretical models and simulation techniques evolve.
Rivet (Robust Independent Validation of Experiment and Theory) is an open-source software framework, originating at UCL, Durham and Lund, which addresses this challenge by providing a standardised, experiment-agnostic approach to encoding and preserving particle physics analyses. Widely used around the world, in 2026 Rivet marks its 20-year anniversary. This milestone provides an opportunity to reflect on two decades of sustained community development, and highlights the strong collaboration between UCL and CERN as a strategic partner in advancing digital research infrastructure, sustainable software and data-intensive science, aligning with UCL's ambitions in digital research and innovation.

Speed Postering Session 2

Poster Running Order

Poster Abstracts and Lead Presenter Information

Name	Last Name	Title	Affiliation
Andrew	**Losty**	**An Analysis of Matter IoT Security Against International Standards and Regulatory Framework.**	**PhD Student**

Since its launch in October 2022, the Matter smart home protocol has emerged as a unifying standard for consumer IoT and is supported by leading technology companies.
Prior to Matter, IoT ecosystems were highly fragmented, relying on vendor-specific protocols, control mechanisms, and security models. This study examines how Matter aligns with 18 international standards such as CRA, NIST, and ETSI, initially focusing on six key security domains, with plans to expand the analysis to sixteen domains. We aim to provide practical and

actionable insights for manufacturers, developers, and regulators considering the adoption and implementation of the Matter protocol.

Daniel	**Joinson**	**Active night-time Tweeting is associated with meaningfully lower mental wellbeing in a UK Birth Cohort Study**	**UCL Staff**

It has been suggested that use of social media late at night could lead to worse mental health outcomes. We linked real-world Twitter data to self-reported measures of mental health from the Avon Longitudinal Study of Parents and Children. We aimed to model mixed effect models predicting these measures from the average hour a participant posted Tweets, using data from 310 adults and 18,288 Tweets. Average hour of Tweet posting was modelling using a novel circular statistic methodology. We found strong evidence the average hour participants posted Tweets was associated with depressive symptoms, anxiety symptoms, and mental wellbeing. Average hour explained almost 2% of the variation in mental wellbeing, comparable to reports of the impact of binge drinking and exercise. Participants who, on average, Tweeted through the night (23:00 to 05:00) showed meaningfully worse mental wellbeing than those who Tweeted during the daytime.

Dr. Fjorda	**Kazazi**	**Network Models for Assessing Comorbidity between Stuttering and ADHD**	**UCL Staff**

Overlap of symptoms between people who stutter (PWS) and people with ADHD (PWADHD) is often claimed, but whether this reflects shared mechanisms or distinct cognitive architectures remains unclear. This study tested whether their cognitive architectures differ. Controls (N = 67), PWADHD (N = 79), and PWS (N = 33) were assessed on stuttering, ADHD traits, and phonological working memory (PWM). Network models compared relationships, connectivity, and structure across groups. Controls and PWADHD showed similar architectures, both differing from PWS. PWM was central in all groups, but its connectivity varied: it affected attention in controls and PWADHD, but not in PWS. Stuttering severity affected PWM in PWADHD and PWS, more strongly in PWS. Higher stuttering rates correlated positively with attention issues in PWADHD and negatively in PWS. Findings do not support a shared mechanism, instead suggesting partially distinct cognitive architectures and a differential role of PWM across conditions.

Aygun	**Badalova**	**Digital Neuro-rehabilitation therapy app for treating proper name anomia in dementia**	**PhD Student**

Proper name anomia is a common experience that can become unpleasantly amplified in people with dementia (PWD). The Gotcha! app aims to provide practice-based therapy for PWD enabling them to spontaneously retrieve the names of key people in their lives. It has been developed using the principles of errorless learning and spaced retrieval pioneered by Clare et al, (2000, 2003), but packaged in an app to support self-management.
Methods:
Gotcha! is a digital confrontation naming therapy app. PWD supply images and names of the people they want to be able to name and train on one face per day for six weeks. We employed a single-case experimental design with weekly testing of free-naming in both six-week blocks (pre therapy and during therapy). A novel speech verifier was used to provide real-time feedback (Barbera et al. 2020). PWD also had an MEG scan before and after the therapy block where they attempt to name pictures of familiar (trained) and famous (untrained) faces. We interrogated the behavioural data in two ways: 1) a within-subject non-parametric analysis using Tau-U metric (Parker et al. 2011); 2) a parametric group analysis using an ANOVA. MEG data were analysed in SPM. We measured source localised gamma-band (30-80 Hz) power 0-3400 ms after the onset of a face. We ran a group-based 2x2 factorial analysis on the resultant images (familiar vs. famous; pre- vs. post-therapy) using a repeated-measures ANOVA to look for changes in power.

Results:
Results from the first 38 PWD to complete the trial demonstrate:
1) Within-subjects, non-parametric results 89% (34/38) showed a positive trend with better naming during the training phase with 20/38 reaching statistical significance at the individual subject level (Tau-U)
2) ANOVA showed a significant effect at the group level of training $F(1,32) = 44.79$, $p = 0.01$
Results from the MEG analysis of 26 PWD: We identified a large cluster of 2876 voxels situated in the left anterior temporal lobe (MNI: -54 10 -6, $F=8.37$, $p=0.001$) where gamma reduction was associated with training (pre-post) of familiar faces, but not (untrained) famous faces.
Conclusion:
Gotcha! app-based therapy for proper name anomia works for the majority of PWD in our trial thus far. This is the first study to demonstrate that the left ventral temporal lobe region supports practice-based retrieval of familiar face-name associations in PWD. Being able to freely produce the name of a relative or loved one has a big impact on people's lives.

Zonglun	**Li**	**Computational screening for lung cancer with protein biomarkers**	**PhD Student**

Low-dose computed tomography is effective for early detection of lung cancer, but concerns such as radiation exposure have limited its use primarily to smokers. Although several promising serum protein biomarker candidates have been reported, none are currently accepted for screening. Besides, existing studies evaluating predictive performance of biomarker candidates rely on non-longitudinal machine learning approaches, and the evaluation of longitudinal methods for improving performance remains absent. We analysed 94 proteins in longitudinal serum samples from 98 lung cancer cases and 150 controls from the UKCTOCS trial. We proposed a new computational biomarker selection strategy and evaluated the performance of longitudinal and non-longitudinal models. Our results highlighted the potential benefits of integrating CEACAM5 and MUC-16 into the current screening framework for smokers, facilitated by the fully Bayesian model. We also observed that the logistic regression model incorporating six candidate biomarkers exhibited exceptional performance in the whole population.

Yousuf	**Ali**	**Targeted Closed-Loop Electrophoretic Neuromodulation: Localised Opioid Delivery in Chronic Pain**	**PhD Student**

Chronic pain affects approximately 30–50% of the UK population, yet systemic opioid efficacy is limited by a pharmacological ceiling where doses required for relief trigger fatal respiratory depression. This research develops a localised Brain-Machine Interface (BMI) utilising a microfluidic ion pump (µFIP) for "dry," electrophoretic morphine delivery to the rostral anterior cingulate cortex (rACC)—a critical hub for integrating negative affect and nociception. To enable state-dependent control, we employ high-dimensional behavioral decoding: unsupervised machine learning via B-SOiD segments 3D pose data into sub-second "Pain Syllables," while Autoregressive Hidden Markov Models (AR-HMM) characterise pathological shifts in behavioral grammar. By integrating these data-driven signatures as triggers, we propose a closed-loop system providing relief from affective suffering with sub-second precision. This approach bypasses medullary toxicity and the rapid tolerance associated with traditional midbrain targets, establishing a sustainable, precision-medicine framework for intractable chronic pain.

Joel	**Roca Martinez**	**Identification and characterization of Polyurethane-Degrading Enzymes from MGnify Metagenomes**	**PhD Student**

The discovery of enzymes capable of degrading synthetic polymers remains a major challenge due to the vast size and functional diversity of metagenomic sequence space. Here, we present a

rational pipeline for enzyme discovery, applied to the identification of novel polyurethane-degrading enzymes (PURases). Starting from 2.4 billion protein sequences in MGnify, we performed homology-based filtering against three known PURases followed by conservation analysis of the catalytic triad, yielding ~8,000 candidates. We clustered sequences using an embedding-based classifier (eMMA-FunFamer) and identified function-determining positions (FDPs) conserved within families but variable across them. Physicochemical variation at these FDPs was used to construct a sequence similarity network, revealing distinct functional clusters that guided representative selection. 20 diverse enzymes were selected for experimental characterization. Biochemical assays identified 9 enzymes with activity against carbamate substrates, including 2with activity against polymers. These demonstrate that combining functional with structural analysis enables efficient navigation of metagenomic sequence space and substantially improves hit rates in enzyme discovery.

Jing	**Gu**	**Point Substitution as a Structural Determinant for Molecular Evolution in PER β-lactamases**	**PhD Student**

To gain a deeper understanding of the functional dynamics of PER β-lactamases, we carried out Adaptive Bandit MD simulations on all PER variant enzymes. We hypothesized that Q219E point mutation in PER-2-like β-lactamases (PER-2, PER-6 and PER-14) should alter the dynamics, kinetics and the substrate profile of the variants and make its local stability of the hinge region similar to PER-1.
Q219E point mutation in PER-2-like β-lactamases should alter the dynamics, kinetics and the substrate profile of the variants and make its local stability of the hinge region similar to PER-1 (with E219). PER-2 Q219E mutant behaves as an enhanced version of PER-2 with the stabilized key hydrogen bond network reported to modulate PER-2 activity.

Lin	**Gao**	**Double Deletion Mutation Induced Active Site Remodelling Modulated Substrate Specificity in KPC-14 β-lactamase**	**PhD Student**

KPC-14, a variant containing a double deletion (ΔG242–T243), shows increased resistance to ceftazidime and ceftazidime–avibactam but reduced carbapenemase activity. The structural basis of this shift remains unclear. Here, we combined molecular dynamics with unsupervised deep learning to investigate its conformational effects. Compared with KPC-2, KPC-14 exhibits increased flexibility in the Ω-loop, 240-loop, and 270-loop. A convolutional variational autoencoder revealed distinct conformational landscapes, with three dominant states in KPC-14 and two in KPC-2. The deletion repositions Y241, disrupting hydrophobic packing and reshaping active-site interactions. These changes alter S70 orientation, destabilise the E166 catalytic network, and shift H274 positioning. Overall, the active site becomes enlarged and less catalytically optimal for carbapenems while better accommodating bulky substrates like ceftazidime, explaining KPC-14's altered substrate specificity.

Manming	**Xu**	**Asymmetric Dynamics Between the Protomers of the σ2 Receptor Homodimer**	**PhD Student**

The sigma-2 receptor (σ2R), controlled by the gene TMEM97, is a clinically essential membrane protein that plays a vital role in cholesterol regulation. Although this protein is therapeutically essential, the mechanisms underlying σ2R function and ligand binding remain poorly understood. To fulfill this research gap, we combined adaptive sampling molecular dynamics simulation with quasi-anharmonic analysis and unsupervised machine learning methods, a convolutional variational autoencoder (CVAE), to investigate the functional insights of the receptor. Our results reveal asymmetric dynamics between the two protomers, driven by anticorrelated helical motions and exclusive salt bridge formation. These very first findings provide a plausible explanation for

the fact that the receptor only functions as a dimer, suggesting that ligand binding at one site may allosterically influence the other, modulating receptor function.

Tiejun	**Wei**	**TEDlm: Domain-Centric Protein Language Models with Structure-Informed Pretraining and Attention Supervision**	**PhD Student**

Protein language models are typically pretrained on full-length sequences, which concatenate multiple domains with disordered regions, potentially obscuring fold-specific signals. We present TEDlm and TEDlm3D, masked language models pretrained on structurally defined domain segments from The Encyclopedia of Domains. TEDlm3D additionally incorporates a Cα distance-guided contact prediction loss using structures from the AlphaFold Database. On the CATH S40 homology detection benchmark, TEDlm 650M outperforms the four-fold larger ESM2 3B (AUROC1 0.280 vs 0.223), while TEDlm3D approaches the structure-based method Foldseek (0.497 vs 0.528) using sequence input alone. On Gene Ontology prediction, TEDlm variants substantially outperform ESM2 in the Molecular Function namespace (wFmax 0.336 vs 0.247). Across biophysical property prediction tasks, TEDlm and TEDlm3D maintain competitive performance despite training exclusively on truncated domains, indicating that these properties are largely intrinsic to individual domain sequences. Collectively, these results establish domain-centric pretraining as a principle for building compact pLMs that excel at fold recognition while maintaining broad downstream utility.

Chunling	**Wu**	**NeckScents: Designing a Wearable Olfactory-Thermal Interface for Immersive VR Cultural Heritage**	**PhD Student**

Multisensory design is increasingly explored to enhance immersive cultural heritage experiences in virtual reality (VR). While olfactory and multisensory VR have been extensively studied, fewer works have examined modular neck-worn olfactory–thermal add-ons integrated as event-aligned narrative accents in cultural heritage VR. This study presents a lightweight neck-worn system that delivers short, pre-authored pulses of scented warm airflow near the face to augment audiovisual VR without altering the visual-audio pipeline. In a between-subjects experiment (N = 52), participants experienced either audiovisual VR (AV) or audiovisual VR with olfactory–thermal feedback (AVOT). Compared to the AV condition, AVOT significantly enhanced immersion and emotional-aesthetic appeal, while engagement showed a positive but non-significant trend. Measures of workload and usability indicated no observable cost from adding the wearable cues. As a preliminary study, these findings provide empirical evidence that a modular, event-aligned olfactory–thermal wearable can enhance experiential and affective dimensions of cultural heritage VR, motivating further exploration of event-aligned multisensory authoring and cross-platform deployment in immersive heritage experiences.

Peng	**Wang**	**Reproducible Collider Physics with Rivet: 20 years of Sustainable Analysis Preservation**	**PhD Student**

Particle physics facilities such as the CERN Large Hadron Collider (LHC) produce vast and complex datasets, requiring sophisticated computational tools to interpret them. A key challenge is ensuring analyses remain reproducible and reusable over time – particularly as theoretical models and simulation techniques evolve.
Rivet (Robust Independent Validation of Experiment and Theory) is an open-source software framework, originating at UCL, Durham and Lund, which addresses this challenge by providing a standardised, experiment-agnostic approach to encoding and preserving particle physics analyses. Widely used around the world, in 2026 Rivet marks its 20-year anniversary. This milestone provides an opportunity to reflect on two decades of sustained community development, and highlights the strong collaboration between UCL and CERN as a strategic partner in advancing

digital research infrastructure, sustainable software and data-intensive science, aligning with UCL's ambitions in digital research and innovation.

James	**Legg**	**Contender Novel Machines**	**UCL Staff**

Contender is ARC's collection of novel architecture machines used for evaluation for performance and readiness for adoption, and for users to do similar and also develop and assess their own code before applying for grants or time on systems elsewhere. The poster is advertising for the project and to attract users.

Marlon	**Wijeyasinghe**	**Trusted Research Environment on Multiple Platforms**	**UCL Staff**

This work explores the development of a Trusted Research Environment (TRE) capable of operating on both cloud-based and on-premises infrastructure. Its cloud deployment on AWS provides robust isolation, security and reliability using Amazon's built-in features. This is currently used by both UCL and Oxford as part of our partnership with them. However, costs can scale quickly with the number of users and their activity, particularly for GPU and compute-intensive workloads. Some of the core aspects of the TRE are virtual desktops, data transfer controls and isolation mechanisms.
We present a strategy to recreate the TRE's functionality across platforms, replacing AWS features with terraform-based internal solutions. There causes challenges in resilience, security and regulatory compliance, which require manual engineering to resolve.
Our work so far demonstrates that multi-platform TREs are viable, but they require careful design and operation. Trade-offs between security and usability need to be managed.

Matthew	**Scroggs**	**DefElement: an encyclopedia of finite element definitions**	**UCL Staff**

Solving problems using the finite element method involves defining (usually) polynomial basis functions on each cell in a mesh. A "finite element" is a mathematical object that is used to define these basis functions. For different physical problems, there are a huge range of finite elements and hence basis function types that can be used. To make it easier to browse these, we have created DefElement: an encyclopedia of finite element definitions. This free online encyclopedia currently contains over 60 finite elements, with many useful pieces of information about each one.

Paddy	**Roddy**	**Porting GLASS to the Python Array API: A Clear View to GPU Portability**	**PhD Student**

GLASS is a Python library designed to generate full-universe cosmological simulations for large-scale galaxy surveys. As simulation demands grow and high-performance computing shifts towards GPU-heavy architectures, GLASS must evolve to support these workloads. Historically, its reliance on NumPy — which was not designed for non-CPU devices — has been a significant bottleneck. To overcome this, we are porting GLASS to the Python Array API, decoupling the library from NumPy and enabling native GPU execution. By adopting this standard, we are expanding GLASS to support all Array API-compliant libraries, broadening its ecosystem and utility. This transition involves managing varying levels of Array API maturity across libraries and optimising core scientific models for GPU parallelism. The resulting GPU-enabled GLASS provides a significant performance boost whilst remaining performant on CPUs, facilitating seamless integration with GPU-native workflows, and enabling advanced techniques like auto-differentiable simulations via JAX.

Mike	**Parkes**	**Simple Tools for Better Environmental Metadata in Chemistry**	**UCL Staff**
Every day a large quantity of chemical research data is produced. The reproducibility of this data is strongly influenced by environmental conditions, yet metadata such as temperature, humidity, and atmospheric pressure are often recorded inconsistently or not at all. This poster presents a simplified, data logging system and online guidance designed to help all chemists routinely capture key environmental metadata during experiments. The system integrates inexpensive sensors with automated timestamping and structured data output, reducing reliance on manual note taking while remaining accessible to non-specialist users. Emphasis is placed on usability, minimal disruption to experimental workflows, and compatibility with existing laboratory record keeping practices. By lowering the barrier to systematic environmental data capture, this approach supports more reproducible, FAIR aligned experimental records and enables downstream analysis of environmental effects on chemical outcomes. The project demonstrates how lightweight digital infrastructure can meaningfully enhance data quality in everyday laboratory practice.			
Angharad	**Green**	**Data stewardship in microbial genomics research: the hidden complexity of reference strains**	**UCL Staff**
Bacterial reference strains underpin the reliability of microbial genomics research and are routinely treated as stable, reproducible standards. However, growing evidence shows that strains assumed to be identical can diverge genetically and phenotypically through routine laboratory handling. This work examines the hidden complexity of reference strains, with Streptococcus pneumoniae D39 serving as a case study. Using comparative genomics, we demonstrate how isolates with a recent common origin accumulate fixed mutations, diverge phenotypically, and display markedly different virulence profiles. Population sequencing further shows that standard culture workflows enrich low frequency variants through genetic drift, reshaping population structure over time and impacting experimental outcomes and reproducibility. These findings highlight an often overlooked source of variability in microbial genomics research. Framed through a data stewardship perspective, this work promotes feasible best practices for reference strain management, including routine sequencing, metadata capture, and standardised handling. It also outlines ongoing community driven efforts to develop shared guidance and digital tools to improve transparency, reproducibility, and data reuse in microbial research.			
Preeti	**Matharu**	**Importance of Information Governance in research projects**	**UCL Staff**
The Information Governance (IG) poster highlights the importance of IG in protecting sensitive research data. It raises awareness of IG's role in ensuring data security, privacy compliance, and ethical data use, helping maintain integrity and confidentiality in research projects. Key elements include: The role of Trusted Research Environments (TREs) in enabling secure data access and use. How the Information Security Management System (ISMS) ensures secure information handling across the research lifecycle. An overview of the IG assurance process, including recent updates. Additional requirements for new projects, such as contracts, data management plans, data protection, and research ethics. Overall, the poster serves as a quick reference guide to promote best practices in IG for research.			

Preeti	**Matharu**	**Growing Data Stewardship at UCL**	**UCL Staff**

UCL researchers are generating more data than ever, across clinical trials, wearables, education, qualitative research, and beyond. Making that data FAIR, secure, and genuinely reusable doesn't happen by accident. It takes infrastructure, expertise, and collaboration.
This poster traces what that looks like across ARC's Data Stewardship work: the governance frameworks and technical services that form the foundation, the training that builds lasting capacity across UCL's research community, and the projects where our Data Stewards have worked directly alongside research teams to tackle real data challenges.
Whether you're deep into a complex dataset or planning a new project from scratch, this is a picture of what Data Stewardship at ARC can offer, and how it grows with you.

Tuomas	**Koskela**	**Making Supercomputers Work for Your Research**	**UCL Staff**

This poster introduces you to the work done in the ARC High Performance Computing collaborations subgroup. We are a group of a dozen research software engineers from a variety of scientific backgrounds. We enjoy solving numerical computing problems that go beyond the capabilities of a single workstation, using UCL's research computing platforms and the largest national and international supercomputers. In this poster we will showcase some successful projects with UCL researchers from recent years, in fields ranging from healthcare to mathematics.
We have expertise in parallel programming on both CPUs and GPUs using a range of programming paradigms and languages commonly used in HPC. We advocate reproducibility and sustainability in the software that we develop, and strongly believe in software engineering best practices in all our projects. We are always open for new collaborations, please come for a chat.

Sergio	**Álvarez-Teleña**	**Advances in Agentic AI: Back to the Future**	**Postdoc**

In light of the recent convergence between Agentic AI and our field of Algorithmization, this paper seeks to restore conceptual clarity and provide a structured analytical framework for an increasingly fragmented discourse. First, (a) it examines the contemporary landscape and proposes precise definitions for the key notions involved, ranging from intelligence to Agentic AI. Second, (b) it reviews our prior body of work to contextualize the evolution of methodologies and technological advances developed over the past decade, highlighting their interdependencies and cumulative trajectory. Third, (c) by distinguishing Machine and Learning efforts within the field of Machine Learning (d) it introduces the first Machine in Machine Learning (M1) as the underlying platform enabling today's LLM-based Agentic AI, conceptualized as an extension of B2C information-retrieval user experiences now being repurposed for B2B transformation. Building on this distinction, and understanding transformation as a greenfield built upon Microeconomics, DeepTech and Machine Learning, (e) the paper develops the notion of the second Machine in Machine Learning (M2) as the architectural prerequisite for holistic, production-grade B2B transformation, characterizing it as Strategies-based Agentic AI and grounding its definition in the structural barriers-to-entry that such systems must overcome to be operationally viable. Further, (f) it offers conceptual and technical insight into what appears to be the first fully realized implementation of an M2. Finally, drawing on the demonstrated accuracy of the two previous decades of professional and academic experience in developing the foundational architectures of Algorithmization, (g) it outlines a forward-looking research and transformation agenda for the coming two decades.

Poster Prize Voting

The Festival Poster Prize recognises outstanding digital research presented during the poster sessions. **Voting start and end time:** 09:30 **–** 15:40.

We encourage all attendees to engage with poster presenters throughout the day and vote for the poster they found most interesting, innovative, or impactful.

Important

One vote per attendee.
Multiple submissions from the same attendee will not be counted.
Winners will be announced during the Closing Remarks and Poster Prize Ceremony.

Tell Us What You Think

Your feedback helps shape future ARC community events.
Please scan the QR code below to complete our Festival feedback survey.

Survey length: approximately 3 minutes
Your feedback will help us:

- Improve future Festivals
- Understand the impact of the event
- Identify opportunities for future collaboration and engagement

Thank you for taking part in the Festival of Digital Research, Innovation & Scholarship 2026.