The newsletter of EPCC, the supercomputing centre at the University of Edinburgh

epcc news

Issue 91 SUMMER 2022

Delivering UK National Supercomputing Services since 1994

From our Director

Over the past decade EPCC has grown into an organisation which extends well beyond our initial focus on parallel computing. More and more our work is spanning the domains of computational science and data science, and this is also in evidence in the infrastructure we provide. A key computational science milestone was reached in January 2022 when ARCHER2 opened for general use. Delivering a sustained 19.5 Petaflop/s on the HPL benchmark and placing it 22nd in the TOP500 supercomputers in the world, the system has been an immediate success with very heavy user demand. Clearly justifying the need for more investment in supercomputing in the UK.

In parallel, we've continued to grow our data science infrastructure through the Edinburgh International Data Facility investment. Earlier this year we reached a key milestone when we completed the end-to-end EIDF service roll-out for our main Data Science Cloud. This is now being enhanced with a full upgrade of our Safe Havens environment and the rollout of a rich set of AI-focussed systems including an NVIDIA A100 cluster and Cerebras CS-1 – which will soon also include a Graphcore Pod64 system using the Bristol-based company's world-leading Bow IPU. Having a wide variety of systems allows the academic, public and private sector users we support to explore the most applicable system for their work.

There are a wide variety of articles in this issue of our newsletter which show how we, and the users we support, are using these infrastructures to great effect. I hope you enjoy this issue of EPCC News and please do get in touch with any feedback.

Mark Parsons,
EPCC Director
m.parsons@epcc.ed.ac.uk

Contents

www.epcc.ed.ac.uk
info@epcc.ed.ac.uk

EPCC is a supercomputing centre based at The University of Edinburgh, which is a charitable body registered in Scotland with registration number SC005336.

Meet our new Sysadmin and HPC specialist

Suyash with Tursa, an Extreme Scaling GPU-based DiRAC system hosted by EPCC's Advanced Computing Facility.

Several new colleagues have joined EPCC in the past six months. Here we meet Suyash Janoriya, a new Systems Administrator and HPC specialist at EPCC.

I am from India, Bhopal in Madhya Pradesh, which is a central state with a mixed culture of all the different flavours. I previously worked with AstraZeneca India Pvt Ltd and Novartis Pharma India Pvt Ltd, both located in India.

Joining EPCC was a big career move, and one of the wisest decisions. I have got a chance to work with the UK's most powerful supercomputer ARCHER2 and other very interesting supercomputers which are one of a kind like Tursa, Tesseract, NextGenIO and Cirrus.

I really appreciate the team and working culture here. All the cutting-edge technology in our day-to-day activities have given me a leap in the field of recent HPC developments and changes.

My roles and responsibilities include the HPC System team's day-to-day operations as well as working on different projects at EPCC.

Currently I am working on developing and expanding the automation of standard service operation requests such as access automation, password resets and account creations so they are completed without human interaction.

The team has already done a lot of work on making these changes for ARCHER2, and expanding the rollout of automated service operation to our other services allows the HPC Systems team to focus on supporting more complex queries and providing a better service to our customers.

I am also involved in data science and computing support, getting a chance to work with the National Safe Haven service and Hydra are one-of-a-kind experiences. I am responsible for addressing user queries, problems, and platform support.

Being an HPC professional there is a lot to learn and upskill myself to deliver the best industrial standard.

I look forward to further exploring the work of supercomputers.

Suyash Janoriya, EPCC
s.janoriya@epcc.ed.ac.uk

This is my first time in Edinburgh and I have started to like the surprising weather! People here are very helpful and kind. I like their enthusiasm, and the way I was welcomed and greeted by one and all.

Read about the HPC services hosted at our Advanced Computing Facility: www.epcc.ed.ac.uk/high-performance-computing-services

UK's most powerful supercomputer on show for Prime Minister

In February Prime Minister Boris Johnson visited EPCC's Advanced Computing Facility (ACF) during a tour of the University's world-class data centre facilities.

During the visit Mr Johnson was shown how ARCHER2 – which is the fastest computer in the UK and the second largest of its kind in Europe – is modelling in detail how medicines interact with cells, is revealing how the northern Atlantic sea ice is responding to climate change, and is supporting Rolls Royce in developing more efficient aircraft engines.

World-class technology

ARCHER2 – which is now fully operational – can perform twenty million billion calculations a second. Its peak computing power is equivalent to around 250,000 modern laptops, making it 10 billion times faster than the first Cray supercomputer produced in 1964.

The £79 million system is funded by the Engineering and Physical Sciences Research Council (EPSRC) and Natural Environment Research Council (NERC), both part of UK Research and Innovation (UKRI).

It gives UK researchers world-class computing capabilities to support breakthroughs in areas such as drug development, climate modelling and the design of sustainable technologies and materials.

"It is exciting to see ARCHER2 coming online and the work that it is already supporting to help tackle some of the world's most pressing challenges. Its location within the EPCC is testament to the University's historic and sustained excellence in computer science and the work of Professor Mark Parsons and the whole team there. We have ambitious plans for supercomputing in Edinburgh - ARCHER2 is a significant step towards achieving them."

Professor Peter Mathieson, Principal and Vice-Chancellor, University of Edinburgh

Next generation

As part of the visit Mr Johnson was shown computer rooms built as part of the Edinburgh and South East Scotland City Region Deal and designed to host the next generation of supercomputers, known as Exascale systems.

Edd McCracken,
University of Edinburgh
Edd.McCracken@ed.ac.uk

The Advanced Computing Facility (ACF) is the high-performance computing data centre of EPCC. Operated by EPCC since the turn of the Millennium, the ACF site has had significant investment over the years. At present there are four Computer Rooms. Each hosts specific HPC equipment and is supported by associated plant rooms which provide dedicated power and cooling infrastructure for each room. By utilising water to help with the cooling process of the machines, which is significantly more effective than air alone, the ACF is an extremely efficient data centre.

Prime Minister Boris Johnson and Professor Mark Parsons in front of ARCHER2 at EPCC's Advanced Computing Facility.
Image: Jeff Mitchell/Getty

Exascale computers will be one thousand times faster than ARCHER2 and will be capable of making one billion billion calculations every second. Because of this investment, Edinburgh is one of the few places in Europe able to host a computer of such a scale.

"ARCHER2 is an exciting and important component of the UK's research and innovation system, providing world-leading computing capabilities to deliver the UK's fastest research computing for both our university-based researchers and their business partners. By allowing researchers to perform virtual experiments on this state-of-the-art system, ARCHER2 will catalyse and accelerate discovery-led research and the development of new applications for the benefit of all society."

Dame Professor Lynn Gladden, EPSRC Executive Chair

Innovative potential

"High-performance computing underpins a wide range of significant research and innovation across environmental science, from modelling the effects of climate change to calculating the properties of the Earth's core. ARCHER2 will greatly increase the computational capacity we have at our disposal and will play an important role in supporting new scientific discoveries and innovations."

Professor Sir Duncan Wingham, NERC Executive Chair

"ARCHER2 represents a world-class supercomputing resource for the UK's high performance computing researchers and EPCC is delighted to be continuing to support them. It was a great pleasure to show the Prime Minister the system. It has taken a huge collaborative effort between EPCC, HPE Cray and UKRI to get to this point. We are very pleased that ARCHER2 is now delivering its full potential for innovative computational science in the UK."

Professor Mark Parsons, EPCC Director

ARCHER2, the UK National Supercomputing Service, is a world-class advanced computing resource for UK researchers. The cutting-edge system – which is around 12 times more powerful than its predecessor – is hosted and maintained by EPCC. We are also responsible for the ARCHER2 Service Desk, the first point of contact for all questions relating to the ARCHER2 service.

Read more about ARCHER2, EPCC and supercomputing at the University: www.ed.ac.uk/impact/research/digital-life/rise-and-rise-of-the-machines

ARCHER2 website: www.archer2.ac.uk/

Our current collaborations with industry

This year is proving to be an exciting one for the EPCC Commercial Group.

With the full ARCHER 2 system and EIDF now up and running we have seen a marked increase in industrial activity utilising this infrastructure on a host of activities, ranging from simulation and modelling to data storage and processing.

Our Tier-2 HPC system, Cirrus, has just been upgraded with a new OS and vastly increased disk storage capacity that will support future and ongoing Industrial engagements.

Commercial project work continues with a variety of new collaborations starting across the Financial Services and Space sectors in the areas of data management and high-performance data analytics. We continue to support industry with respect to cloud architecture and optimising product development in this area.

EPCC is working with Alchemy Machines (www.alchemymachines.ai), a London-based SME in the legal tech sector, to optimise its speech-to-text neuro-linguistic programming (NLP) models using the Cerebras CS-1 AI accelerator in the Edinburgh International Data Facility (EIDF). Built and operated by EPCC, EIDF is the set of computational and data services and IT infrastructure that underpins the Data-Driven Innovation Programme.

More efficient waste recycling

EPCC's engagement with Danu Robotics (https://danurobotics.com) continues with the next phase focussed on preparing a working prototype of the Danu Robotics waste recycling solution.

Danu Robotics is working to create a robotic system to automate the sorting of material to be recycled. This system is composed of an AI image recognition component that can detect recyclables on a conveyor belt and a mechanical component that can physically pick up and sort the recyclables.

We are working closely with Danu Robotics to ensure that there is efficient communication between the image-recognition part and the mechanical component of the system.

Thomas Blyth, EPCC
t.blyth@epcc.ed.ac.uk
Julien Sindt, EPCC
j.sindt@epcc.ed.ac.uk

Edinburgh International Data Facility (EIDF)

EIDF is a collection of computational, data management, and safe haven services that underpins the Data-Driven Innovation Programme (DDI). Built and operated by EPCC, EIDF is a place to store, find and work with data of all kinds. It is funded by the UK and Scottish Governments under the DDI Programme of the Edinburgh and South-East Scotland City Region Deal.

www.ed.ac.uk/edinburgh-international-data-facility

The Cerebras CS-1, which is one of the world's fastest AI computers, is part of the Edinburgh International Data Facility and managed by EPCC.

Furthermore Danu Robotics and EPCC are discussing follow-up collaborations. This includes investigating whether the Cerebras CS-1 AI accelerator can be employed to train the image-recognition component of the system in a more efficient and "greener" fashion, making use of the better energy efficiency of Cerebras compared to traditional GPUs.

Quantum computing

EPCC is currently completing a proof-of-concept study that is testing an algorithm used for air traffic optimisation in a Quantum Computing environment as part of an ongoing collaboration with a UK-based SME and large services provider in this sector.

Our research activity in Quantum Computing has picked up significantly with EPCC's founding role in the recent Quantum Computing Applications Cluster here in Scotland (www.qca-cluster.org).

Finally, EPCC is exploring potential collaborations with Level E Research (https://www.leveleresearch.com/) to optimise its autonomous trading portfolio product in cloud and high-performance computing infrastructure. This Edinburgh-based FinTech provides a modular software-as-a-service platform that collates stock market information to offer portfolio and investment recommendations to its clients.

Together, EPCC and Level E Research will be creating a cloud-based replica of the Level E Research computer cluster within the EIDF Virtual Desktop Infrastructure. This replica would enable Level E Research to try out ideas that they have for improving their services without affecting the user experience of their clients. This "virtual test bed" is, by its very nature, expandable and Level E Research hope to use it to test the infrastructure required to improve their platform without needing to commit to purchasing potentially unsuited equipment.

Cirrus

Hosted by EPCC, the Cirrus HPC system is designed to solve computational, simulation, modelling, and data science challenges. It can support a wide range of industry applications and EPCC can provide both access to HPC resources and also to consultancy to ensure HPC resources are effectively exploited.

Read more about our services for industry: www.epcc.ed.ac.uk/industry-solutions

For access to EPCC's HPC systems and expertise please contact Thomas Blyth, EPCC Commercial Manager: t.blyth@epcc.ed.ac.uk.

Biomedical urgent computing in the Exascale era

CompBioMed is a European Commission H2020-funded Centre of Excellence focused on the use and development of computational methods for biomedical applications.

Gavin J. Pringle, EPCC
g.pringle@epcc.ed.ac.uk

Part of CompBioMed's remit is to prepare biomedical applications for future Exascale machines, where these machines will have a very high node count. Individual nodes have a reasonable mean time to failure; however, when you collect hundreds of thousands of nodes together in a single system, the overall mean time to failure is much lower. Moreover, given Exascale applications will employ MPI, and a typical MPI simulation will abort if a single MPI task fails, then a single node failure will cause an entire simulation to crash.

Computational biomedical simulations employing these Exascale platforms may well employ time- and safety-critical simulations, where results are required at the operating table in faster than real-time. Given we intend to run these urgent computations on machines with an increased probability of node failure, one mitigation is to employ what we have called Resilient HPC Workflows.

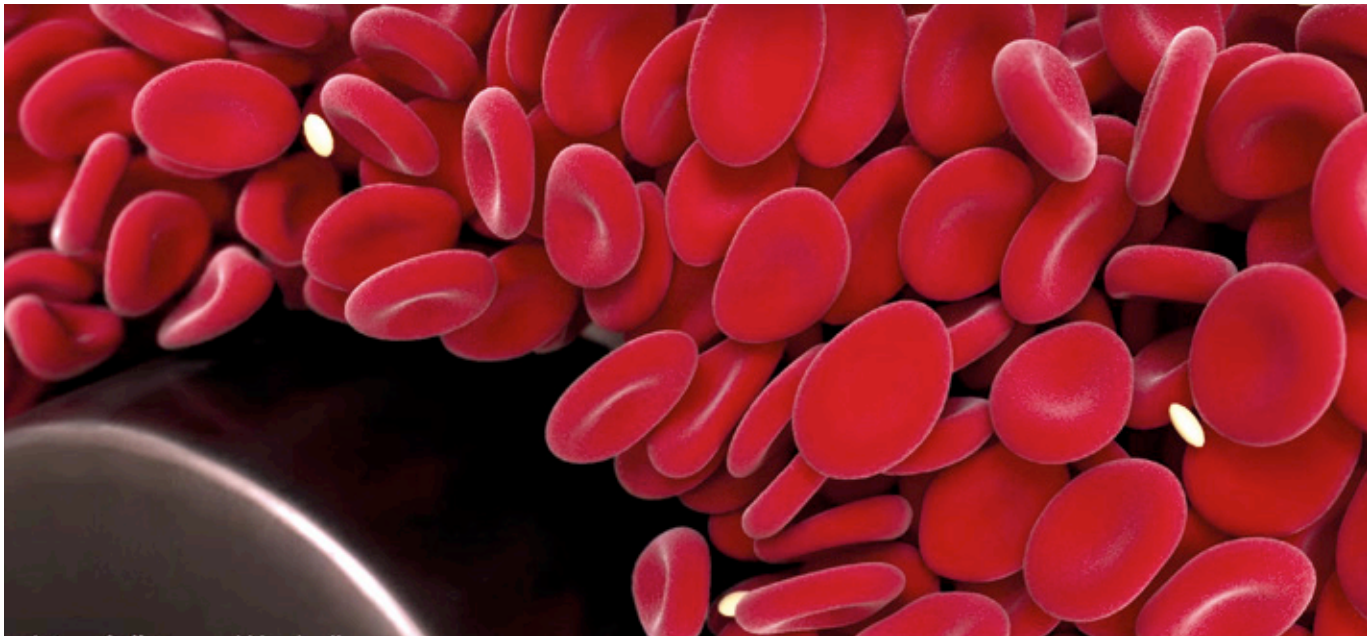
Two classes of such workflows employ replication. The first resilient workflow replicates computation, where the same simulation is launched concurrently on multiple HPC platforms. Here the chances of all the platforms failing is far less than any individual; however, such replicated computation can prove expensive, especially when employing the millions of cores expected in Exascale platforms. The second resilient workflow replicates data, where restart files are shared across a distributed network of data and/or HPC platforms. Then, a simulation that finds its host HPC crash will simply continue from where it left off on another platform.

In both classes, the simulation's results will be available as if no catastrophic failure had occurred; however, the first class, where computation is replicated, will produce the results quicker as the simulation will not have to wait in a second batch system. On the other hand, from personal experience, the time to coordinate multiple HPC platforms to start a simulation concurrently increases as the cube of the number of platforms. As such, the total turn-around time of the second class, ie, where data is replicated, is more likely to be faster. Turn-around might be reduced further via the use of batch reservations.

The LEXIS project [1] is building an advanced engineering platform at the confluence of HPC, Cloud and Big Data, which leverages large-scale geographically-distributed resources from the existing HPC infrastructure, employs Big Data analytics solutions and augments them with Cloud services.

LEXIS already has the first form of reliant workflow in its arsenal. As part of CompBioMed, EPCC is

In the field of personalised medicine, one example of urgent computing is the placement of a stent (or flow diverter) into a vein in the brain: once the stent is inserted it cannot be moved or replaced. Surgeons of the future will use large-scale simulations of stent placement, configured for the individual being treated. These simulations will use live scans to help identify the best stent, along with its location and attitude.



Simulation of the effect on red blood cells of a flow diverter implanted in a brain artery to reduce local pressure.

working with LEXIS and the University of Amsterdam (UvA) to create the second form: resilience via data replication. The simulation employed is the HemoFlow [2] application from UvA. Under normal operating conditions, the HemoFlow application writes frequent snapshots and, less frequently, restarts files.

For the data replication workflow, we have created a data network and an HPC network. The data network includes three data nodes: one at the SARA [3] HPC centre in the Netherlands, one at LRZ [4] HPC centre in Germany, and one at IT4I [5] HPC centre in the Czech Republic. The HPC network includes five HPC systems: Cirrus [6] at EPCC in the UK, two at LRZ (including SuperMUC-NG [7]) and two at IT4I. Both these networks are distributed across different countries to mitigate against a centre-wide failure, eg, power-outage, at the initial HPC centre. The application has been ported to all the HPC systems in advance. The input data resides on the LEXIS Platform and is also replicated across the data nodes. The LEXIS Platform resides at IT4I, and the workflow manager, named the Orchestrator, is run within the LEXIS Platform.

The test workflow is under construction and will progress as follows. The Orchestrator submits the simulation at the initial HPC centre. As soon as the simulation begins and restart files are created, the Orchestrator replicates these restart files across all the data nodes. As the simulation progresses, a node failure is emulated: a single MPI task will abort and, as such, will cause the entire simulation to fail. This failure will trigger the Orchestrator to restart the simulation on one of the remote HPC platforms, using the latest pre-staged restart file. The automated choice of platform ensures the fastest turn-around and will be performed by the LEXIS Platform's broker tool, namely its Dynamic Allocation Module. Once the target HPC platform is known, the latest restart file is staged there from the closest data node.

The staging of data for Exascale simulations must naturally consider both the amount of data that needs to be moved and the bandwidth required; however, given biomedical simulations can contain patient-sensitive data, data staging must also ensure the staging follows FAIR data principles [8]. The amount of data can be addressed by employing the existing GEANT2 network [9], whilst FAIR data principles are addressed by the LEXIS Platform itself, as the underlying mechanism for staging data employs EUDAT [10].

Exascale supercomputers bring new challenges and our Resilient HPC Workflow mitigates the low probability but high-impact risk of node failure for urgent computing.

These supercomputers are on the horizon and present an exciting opportunity to realise personalised medicine via ab initio computational biomedical simulations which, in this case, will provide live, targeted guidance to surgeons during life-saving operations.

[1] LEXIS: lexis-project.eu/web

[2] HemoFlow: zavodszky.com

[3] SARA: surf.nl/en/research-ict/data-storage-and-management

[4] LRZ: lrz.de/forschung/projekte/forschung-e-infra/LEXIS

[5] IT4I: it4i.cz

[6] EPCC Cirrus: epcc.ed.ac.uk/hpc-services/cirrus

[7] LRZ SuperMUC-NG: doku.lrz.de/display/PUBLIC/SuperMUC-NG

[8] FAIR data principles: en.wikipedia.org/wiki/FAIR_data

[9] GEANT2: geant.org

[10] EUDAT: eudat.eu

CompBioMed website
www.compbiomed.eu

A data-driven pandemic response

The Scottish National Safe Haven (NSH) is hosted and operated by EPCC and governed by the electronic Data Research & Innovation Service team (eDRIS), part of Public Health Scotland. At the start of the pandemic a COVID-19 database was set up within the NSH, which includes a subset of Scottish NHS and associated data available for researchers to apply for linked data extracts which are created by eDRIS analysts. Alongside this the ISARIC4C database was set up, which has now grown into the Outbreak Data Analysis Platform.

Lucy Norris, EPCC
l.norris@epcc.ed.ac.uk

The International Severe Acute Respiratory Infection Consortium (ISARIC) was established in 2011 and from this stemmed the ISARIC Clinical Characterisation Protocol (ISARIC4C) led by a UK-wide consortium of doctors and scientists. ISARIC4C had a generic protocol and CRF (Case Report Form, used for data collection in research studies) approved prior to COVID-19, in readiness for the next SARS outbreak. This meant that instead of waiting months, they were able to start recruiting as early as January 2020!

ISARIC4C has recruited 303,251 patients hospitalised with COVID-19 in the UK. During the first wave of the pandemic, three quarters of hospital patients were recruited. Recruitment stopped at the end of February 2022, with final follow up assessment four weeks later. A final check and clean of the database is now underway after which it will be locked. The CRF changed along the way eg addition of vaccination and reinfection data, and additional complications, and also changes to inclusion criteria, so latterly only patients of specific interest eg suspected reinfections, were recruited.

Most of the ISARIC data (Tier 0) is unconsented, with special authorisation given at the start of the pandemic by a COPI (Control of Patient Information) notice and PBPP (Public Benefit and Privacy Panel) permissions. However 2,914 patients have consented to go into Tiers 1 and 2 where additional sample data is collected.

ODAP

In the Outbreak Data Analysis Platform (ODAP) database we have received weekly updates to the ISARIC data and also NHS, ONS and other study data for linkage purposes. These come via the COVID-19 database, and the Scottish NHS data in ODAP is a subset of the COVID-19 data, in line with PBPP agreements. English NHS data has come via NHS Digital. There is a huge amount to decipher, check against approvals and catalogue and there have been amendments to agreements following on from this. Data now arrives monthly and interpretation is ongoing.

NHS data includes hospital inpatient and outpatient data, GP data, mental health, diabetes and cancer datasets. We also have NRS and ONS Deaths data, Variant, Vaccination, and Testing data. Other study data includes PHOSP (Post Hospitalisation COVID-19 Study) data which we are already hosting

ISARIC data

The ISARIC data has 885 columns and 3.25 million rows. It originates from a Redcap database but comes to us in a non-relational format with many variables only completed on some rows. The Surgical Informatics team in the University have developed extensive 'cleaning' scripts for the ISARIC data to update values on the main database and add some derived variables including deprivation indices derived from postcodes which cannot be given to the researchers. The scripts also create several summary tables, including 'oneline', which has only one row per patient but is wider with 1,574 variables.

Paulynn via Getty Images

for PHOSP researchers to access, but will in future be linked to the ISARIC data for broader analysis.

Data extracts process

While eDRIS analysts create the extracts from the COVID-19 database, as they do for Scottish NHS linked data, EPCC does this for ISARIC. We have been creating these extracts since early 2021 – the process is managed in conjunction with eDRIS and Roslin ODAP colleagues, and systems for specification, extraction and governance have been authored and evolved jointly. Researchers specify which variables they wish to have – some are restricted and require additional justification and approvals – and any linked data. We create the extracts and transfer them to eDRIS who have a two-stage checking process before releasing it to the researchers. Researchers conduct their analyses in their project space in the Safe Haven and when they have summary data outputs ready to export, they are subject to a Statistical Disclosure Check with eDRIS and additional approval from the consortium.

Outputs

Researchers for whom we have provided data are spread across the UK, with a variety of specialisms. These include neurology in Liverpool, neuro-psychiatry in Southampton, variants and host factors in Cambridge, cardiovascular and diabetes in Leicester, haematopoeisis in Cambridge, Long COVID in Glasgow and genomics and co-infections in Edinburgh.

The first papers resulting from these data provisions have been published and there are several more now in the pipeline.

ODAP future

As the ODAP platform expands, there are proposed developments to streamline data access through a single/lead data controller, using the Five Safes framework via the HDR UK Innovation Gateway and oversight and strategic direction by the ODAP Steering Group. A new Data Access Committee is being set up as part of the core functions of the ODAP data access activities.

There will be further linkage eg to COG-UK variant data, GenOMICC genome sequence data, UK-CIC phenotype data, and NHS data. We expect to provide an increased amount of linked extracts in future and look forward to seeing more publications resulting from this work.

Flexible compute space

In addition to the Safe Haven part of ODAP, EPCC hosts the Flexible Compute Space (FCS) which holds less sensitive ISARIC data, including the consented samples data. Additional computing capability here, of the new HPE SuperDome Flex large memory system, allows for processing of large datasets including sample and genomic data. This area is managed separately by the Roslin ODAP team. There are now plans to make this into a Trusted Research Environment (TRE) with additional governance in line with the NSH.

Learn more about eDRIS
www.isdscotland.org/Products-and-Services/eDRIS/

ISARIC4C consortium
https://isaric4c.net/

RAMP: Epidemiological modelling of the COVID-19 pandemic

EPCC and the University of Edinburgh's School of Physics and Astronomy have been working on the Royal Society-convened Rapid Assistance in Modelling the Pandemic (RAMP) initiative. The work expanded beyond its original brief to set up a volunteer programme, to also spawn independent teams writing independent epidemic-modelling code. The RAMP volunteer programme has now closed and we are embarking on a project with Microsoft and collaborators in the UK and the USA to quantify uncertainty in epidemiological modelling.

Kostas Kavoussanakis, EPCC
k.kavoussanakis@epcc.ed.ac.uk

On March 25th, 2020, Mike Cates, Lucasian Professor of Mathematics at the University of Cambridge, contacted EPCC on behalf of the Royal Society asking for support in setting up a volunteer programme. The Scientific Pandemic Influenza Modelling Group (SPI-M), the UK academic community of pandemic modellers, was at full stretch and was in need of volunteers with suitable skills in computer modelling. EPCC set up a survey to collect volunteer information and the infrastructure to process the collect data, and by March 27th the Rapid Assistance in Modelling the Pandemic (RAMP) programme was in full flight.

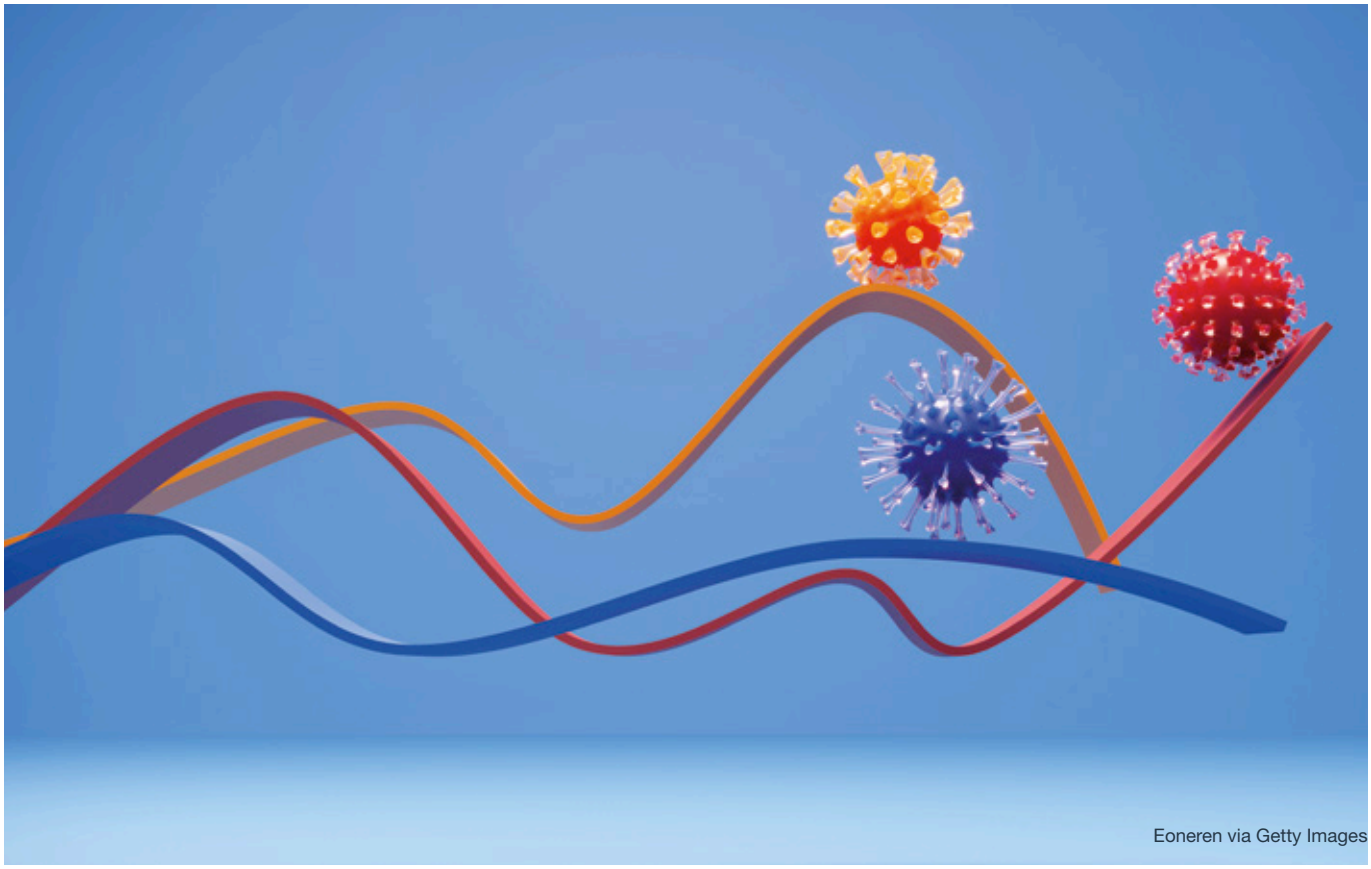
And what a flight that was. When we closed the survey on April 3rd we had received 1,854 responses; as people kept approaching us, we opened more surveys and received in total 2,028 responses from 44 countries. Many of the offers were from groups and the total number of volunteers was never calculated, but the response is humbling.

Led by our SPI-M contact, Prof. Julia Gog, we offered the first batch of volunteers to SPI-M groups on April 7th. By the end of the volunteer programme in August 2020 we had offered 93 responses to 28 SPI-M subgroups. We also created a small community around a forum (we acknowledge the voluntary contribution of thepavilion.io in customising the software for it), which delivered well received rapid review of publication pre-prints, amongst other community discussion of scientific outputs.

In addition to SPI-M, RAMP offered another 118 responses to 11 different organisations and teams. These were scientists with strong computational and mathematical backgrounds, most of whom had not worked on human, or indeed any kind of, epidemic modelling. In addition to fuelling new groups, like the Scottish COVID-19 Response Consortium, a rapid response group at the University of Edinburgh allayed concerns about the Imperial College London code that led to the change in UK and US strategies in addressing the pandemic in March, helping restore faith in science. In addition, Prof. Graeme Ackland of the School of Physics and Astronomy and former EPCC Director Prof. Sir David Wallace designed, wrote and implemented a

So where are all the personal data of the volunteers now? Quite simply, we deleted them. We kept some anonymous information to support simple statistics and some suggestions for the future but the original personal data have been deleted.

The legal teams of the University of Edinburgh and of the Royal Society generated simple Privacy Statement, Data Processor, and Data Controller documentation which governed our use of the volunteer data, and these documents may be useful in a future pandemic.



Eoneren via Getty Images

new epidemic modelling code; the code has now been adopted by the Scottish Government, SPI-M, and the UK Health Security Agency.

The RAMP Steering Group also attracted UKRI funding towards the RAMP Continuity Network (January 2021 till July 2022), which delivers: RAMP-convened scientific discussion meetings, study groups, and workshops; the RAMP Rapid Review Group, which scrutinizes outputs and reports so as to better inform Government; and policy work to ensure the scientific outputs of RAMP and others land correctly within Government to maximise their policy value.

Open-source modelling tools

We now look towards the future. After the volunteer programme of RAMP concluded, we worked with UKRI funding (from June 2020–November 2021) under Prof. Ackland and Microsoft to adopt an open-source modelling tool called COVID-UI that GitHub had developed, and which was aimed at policymakers to run and compare the outputs of various models under user-specified conditions. Through this programme we ported the tool to Microsoft Azure, added WSS to it and made the open-source code more maintainable and sustainable. COVID-UI is now documented in good detail, to support future uptake.

This work was supported through Microsoft's "Studies in Pandemic Preparedness" collaborative research programme. The work is continuing with a follow-up effort towards quantifying and explaining epidemiological model uncertainty, the University of Edinburgh working with the Universities of Exeter, Oxford, Johns Hopkins and Washington, the Wilson Centre, and supported by Microsoft.

The societal goal is to help improve policymaker and public confidence in mathematical models used during the COVID-19 and future pandemics. The research objective is to develop, demonstrate, and deploy state-of-the-art machine learning and data visualisation approaches to measure and better explain uncertainty in models for UK COVID-19 response, US local COVID-19 response, and vector-born epidemic response.

Our work to improve pandemic response is ramping up.

"We founded RAMP to support overburdened epidemic modellers by providing volunteers, model development, literature review, and its current focus of scientific workshops. RAMP volunteers contributed over 15 person-years of effort direct to SPI-M modelling teams, and more than that to independent modelling efforts."
Prof. Mike Cates, Chair of the RAMP Steering Group

The Edinburgh RAMP team linked to UKRI ST/V00221X/1 and UKRI EP/V053507/1: Professor Mark Parsons, Mario Antonioletti, Clair Barrass, Steven Carlysle-Davies, Nick Johnson, Kostas Kavoussanakis (EPCC); Professor Graeme Ackland (School of Physics and Astronomy); Chris Ness (School of Engineering).

Microsoft's "Studies in Pandemic Preparedness" collaborative research programme: https://aka.ms/PandemicPreparedness

New Scottish Medical Imaging Service

Vadzim Kushniarou via Getty Images.

Each year millions of clinical images such as X-rays, CT, MRI, ultrasound, and nuclear medicine images are generated in the NHS in Scotland and stored in the national Picture Archiving and Communication System (PACS).

While such images contain important clinical information they also contain a great deal of potential information about the health of the individual which is currently not made use of in health care.

The Scottish Medical Imaging (SMI) Service is part of the electronic Data Research and Innovation Service (eDRIS) team within Public Health Scotland. Working in partnership with the Health Informatics Centre (HIC) at the University of Dundee, and EPCC on an MRC-funded programme grant, "PICTURES", we have developed and tested the tools to collate, de-identify, and build cohorts from medical images collected for routine healthcare.

This new national resource, which launched in April 2022, will be used to provision pseudonymised images and associated report data to researchers which, if required, may be linked to other available pseudonymised datasets.

The aim of the Scottish Medical Imaging (SMI) service is to provide linkable, population based, "research-ready" real-world medical images for researchers to develop or validate AI algorithms within the Scottish National Safe Haven.

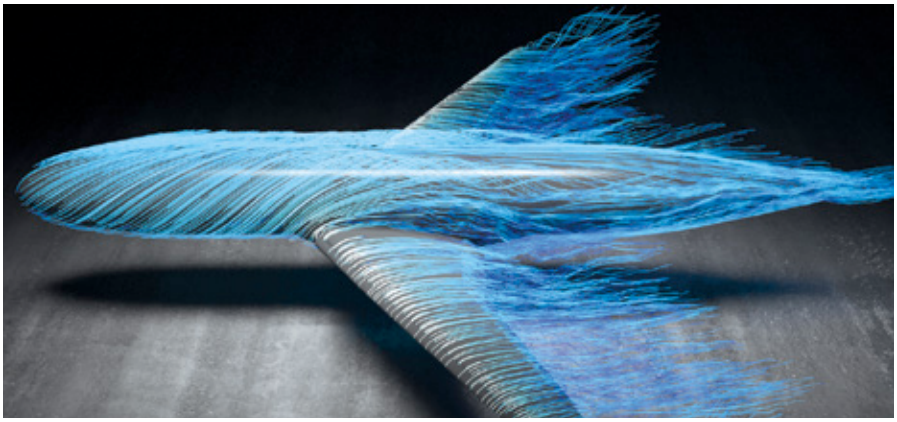
The SMI Service has been created to declutter imaging data access in Scotland by providing a single user journey, regardless of the type of imaging data being requested. It will provide information about the imaging data landscape and how to access the data, what timescales are likely to be encountered, and the information that a researcher needs to provide to gain access to the data sets. With this, SMI is able to guide researchers through the data access journey, offering advice where needed, to make the process as seamless and effective as possible.

Jackie Caldwell,
SMI Strategic Lead
Jackie.Caldwell3@phs.scot

The SMI database has been created from a copy of the national Picture Archiving and Communication System dataset, to be held in the National Safe Haven. This is located at EPCC's Advanced Computing Facility.

To find out more about the Service visit our website or contact the eDRIS team:
phs.edris@phs.scot
www.isdscotland.org/Products-and-Services/eDRIS/Scottish-Medical-Imaging-Service/

The European Centre of Excellence for Engineering Applications



The European Centre of Excellence for Engineering Applications (EXCELLERAT) came to the end of its first phase in May 2022. The Centre was established to be a single point of access for expertise on how engineering workflows can benefit from data management, data analytics, visualisation, simulation-driven design and Co-design with high-performance computing (HPC).

Engineering is one of the key industrial areas that can benefit from the use of HPC, and it remains one the most important sectors in Europe. The European engineering supply chain eco-system consists of over 100,000 companies and more than 10 million jobs. Engineering is seen as one of the industry areas in which Exascale computing can have the most significant impact.

EXCELLERAT brings together Europe's leading HPC centres, application specialists, and supporting partners who have worked with and offered their expertise and knowledge to engineering companies and researchers from across industry and academia for the past three decades.

Potential Exascale applications

The core of EXCELLERAT's work is around six reference applications, which provide a focus for its work. The six applications have been chosen for their potential as Exascale applications and for their industrial relevance. By studying these applications and how they can be adapted for Exascale, the CoE partners have learned valuable lessons and developed new approaches and technologies that can be applied to other software.

The CoE has been running an outreach programme to encourage the wider industrial engineering community to take up the results of the Centre's work.

Service portal

An important activity in EXCELLERAT is the development of a service portal through which external users can benefit from the tools, expertise and training that EXCELLERAT offers. The portal (https://services.excellerat.eu/) is aimed at developers and engineers and is important for making the CoE sustainable.

The CoE plans to continue under the EuroHPC Joint Undertaking (https://eurohpc-ju.europa.eu/). Because of the UK's withdrawal from the European Union, EPCC will be unable to take any further part in EXCELLERAT. However, EPCC remains committed to the development of Exascale computing, and fostering and supporting its take-up by industry.

We extend our best wishes for the future to our many friends and colleagues in EXCELLERAT.

Mark Sawyer, EPCC
m.sawyer@epcc.ed.ac.uk

The goal of EXCELLERAT is to enable the European engineering industry to advance towards Exascale technologies and to create a single entry point to services and knowledge for all stakeholders (industrial end users, ISVs, technology providers, HPC providers, academics, code developers, engineering experts) of HPC for engineering. In order to achieve this goal, EXCELLERAT brings together key players from industry, research and HPC to provide all necessary services.

EXCELLERAT

www.excellerat.eu

Preparing for an unprecedented astronomical data set

The Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) is the most ambitious optical sky survey yet planned. In construction in northern Chile, this US-led project is expected to make a step-change in our understanding of dark energy, the evolution of planetary systems and their capacity to sustain life, and the constituents of matter.

George Beckett, EPCC
g.beckett@epcc.ed.ac.uk

Over the course of a decade, beginning in 2024, the Rubin Observatory will build up a deep, multicolour map of the southern sky, covering 18,000 square degrees. The survey will image each part of the sky more than 800 times over its 10-year course, revealing significant new details of the Universe's dynamics and identifying an unprecedented number of transients, such as supernovae stars, asteroids in the Keiper Belt, and gravitational-wave triggers.

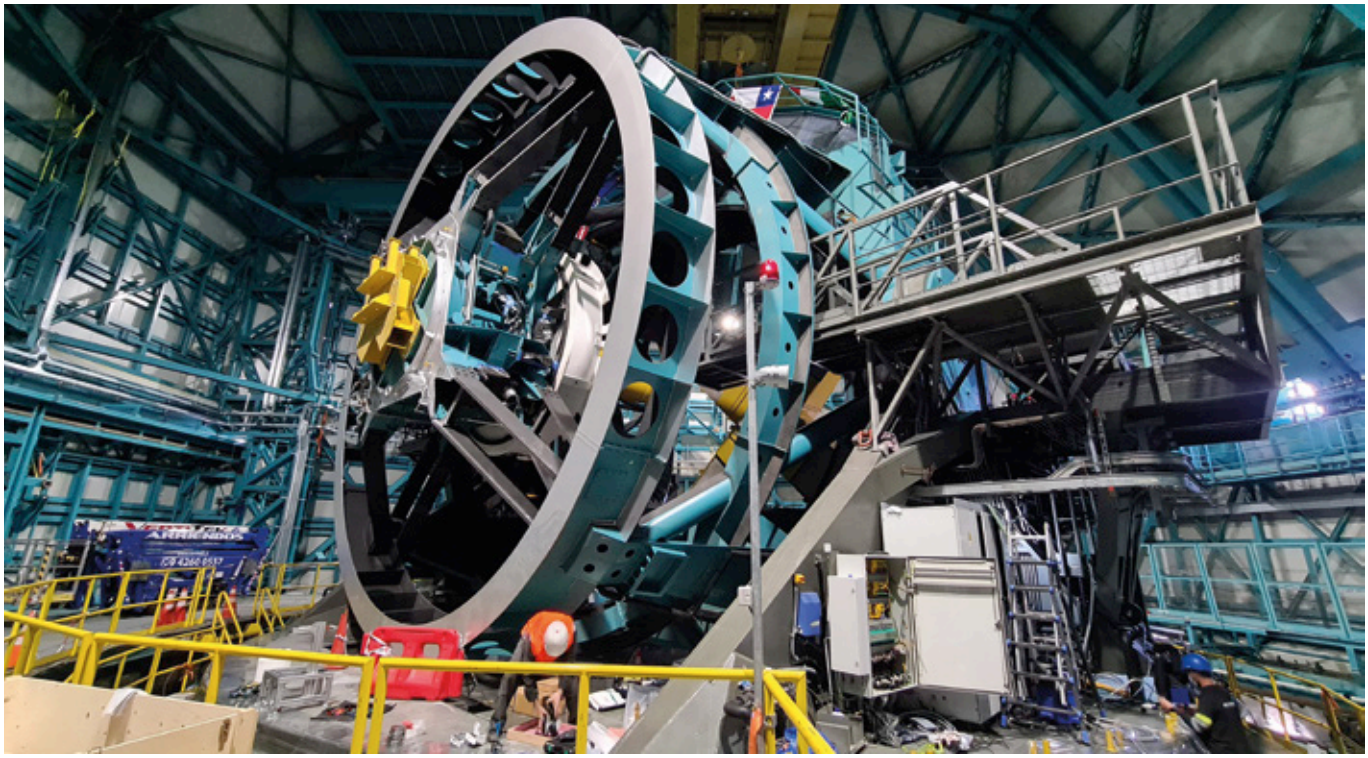
Recognising the significant opportunity of LSST, in 2013 UK astronomy groups formed the LSST:UK Consortium to coordinate UK exploitation of LSST and to develop and secure funding for a substantial 20-year programme of research and development activities. The Consortium has a portfolio of work addressing the processing and curation of survey data products, serving those data products to the community, as well as downstream exploitation for key UK interests in astronomy and cosmology.

The size of the survey makes it impractical for astronomers to download survey data. Therefore they will run their analysis at one of the Data Access Centres (DACs). The University of Edinburgh (in a collaboration between EPCC and the Institute for Astronomy) is a core partner in LSST:UK and will operate one of only three full-scale DACs, holding the complete survey (estimated at 200 Petabytes) and providing cloud-based and HPC analysis platforms to enable world-class astronomy and cosmology.

The applications for LSST are wide and varied. To accommodate this variety, a science portal called the Rubin Science Platform is being developed by the observatory, with three different interfaces: a browser-based query engine called Firefly to address common, interactive astronomer queries; a notebook interface called Nublado, which allows astronomers to develop and part-automate more complex analysis workflows using a scripting language such as Python; and a batch-processing interface, which will allow large collaborations to employ high-performance computing to conduct large-scale batch-processing of survey data.

Alongside the main survey, the Rubin Observatory will monitor night-to-night changes in the sky, searching for rare and scientifically important events such as supernovae explosions, approaching near-Earth objects, and gravitational-wave counterparts. Detected events, expected to number around 10

LSST will deliver a 500 petabyte set of images and data products that will address some of the most pressing questions about the structure and evolution of the Universe and the objects in it. Software is one of the most challenging aspects of Rubin Observatory, as more than 20 terabytes of data must be processed and stored each night.



Inside the observatory, progress continues on the telescope mount assembly.
Image: Rubin Obs/NSF/AURA.

million per night, will be sent to a small number of institutions around the world for further processing. A joint team in Edinburgh and Queen's University Belfast is developing a platform called Lasair to receive this event stream, and it will be one of seven official Rubin brokers world-wide which will receive the stream. Lasair is already running, hosted at EPCC's Advanced Computing Facility (ACF) and processing an alert stream from the Zwicky Transient Facility (a precursor to LSST) which is employed by international research groups.

Alongside the implementation and hosting of a UK Data Access Centre, EPCC is providing overall project management and is also involved in R&D activities focused on UK science priorities. For example as part of the Dark Energy Science Collaboration, we are porting and optimising an important telescope simulation code called GalSim to work with GPUs, to allow it to be run on new HPC services that are coming online in the US and UK. Work to date has more than halved the runtime for LSST-scale simulations and enabled it to run efficiently on many-core systems such as the AMD Epyc processor in ARCHER2.

EPCC staff are also working with a team of astronomers at Exeter University to optimise a program called Macauff that matches LSST objects to equivalent objects from other surveys – for example to allow an astronomer to look up measurements for a galaxy observed by LSST from other surveys. The sheer number of LSST objects (roughly 35 billion) and the density of objects – especially in the galactic plane – make this a computationally challenging task. EPCC staff are creating a parallel implementation of Macauff which could run on a UK Tier-2 HPC service such as Cirrus and be capable of computing a full crossmatch of LSST to another survey in under two weeks.

The LSST:UK programme is a long-term one, due to continue until at least 2035. Observations from the operational telescope are due to arrive from late 2024. Preparations for this are in full swing, with an OpenStack-based DAC platform called Somerville sited at the ACF and hosting pre-cursor surveys and early observations from LSST commissioning activities. These also have scientific value, meaning UK-based astronomers are already beginning to reap the rewards of the Rubin Observatory as well as learning the data-intensive research skills and techniques they will need to up-scale to LSST.

Survey data consists of processed and combined images and catalogues (databases of scientific measurements based on detected objects eg galaxies and stars). LSST is expected to identify and classify around 35 billion objects. The size of these catalogues (up to 20 Petabytes) has prompted the observatory to develop a bespoke database platform called Qserv, a distributed, relational database comprising several hundred SQL databases tied together using tailored workflows that distribute queries across a partitioned view of the sky and reconstruct a result set from the outputs.

LSST:UK
www.lsst.ac.uk

LSST
www.lsst.org

The UK's National HPC Competence Centre

EuroCC is European-funded network of National Competence Centres (NCC) in HPC across Europe. EuroCC@UK is the UK's NCC and acts a central point of contact for HPC and related technologies across the UK.

EuroCC@UK is run by EPCC in collaboration with STFC's Hartree Centre, with both centres utilising complementary expertise to deliver on the goals of EuroCC and the UK's NCC. Now in its second year, the collaboration's two UK partners have focused their efforts on a range of areas including training, industry engagement, facilitating scientific and technical expertise transfer, and competence mapping.

Accessible training

EPCC has taken a lead in investigating innovative training methodologies and accessibility. We carried out a survey of existing training courses and identified topics to be used in an exemplar of accessible training. This has led to the development of an online self-study course (similar to the successful MOOC format) designed to make the course as accessible as possible. The course builds on the web-based delivery format developed by the Carpentries, which has been designed to be accessible and configurable.

Technology transfer

UK technology transfer is our second area of focus, and we have been working to ensure the correct mechanisms are in place to support it. For example, we have initiated an awareness campaign aligned with our normal commercial outreach activities to promote the NCC to UK industry. Under EuroCC we have been assisting companies to prepare bids for further funding by providing technical and business expertise to ensure high-quality proposals.

Pilot studies

EPCC has initiated a pilot study under the UK NCC and supported by the Scottish Funding Council which is investigating novel machine learning techniques applied to recycling packaging materials. A second pilot study developed under the UK NCC has been confirmed and we have a pipeline of interested companies with which we are discussing future pilot studies.

Lorna Smith, EPCC
l.smith@epcc.ed.ac.uk

The National Competence Centres' missions

Develop and publish a comprehensive and transparent map of HPC competencies and institutions in their own country.

Act as a gateway for industry and academia to national and international providers with suitable expertise or relevant projects.

Collect and publish HPC training offers in their own country and international training offers collected by other NCCs.

Foster the industrial uptake of HPC.

Technology reports

We are carrying out technology state-of-the-art investigations, each of which produce a report on a topic of interest to HPC application developers. We have recently initiated two of these investigations.

The first of these will assess the suitability of the Rust programming language for HPC applications. Rust is a relatively new language: like C++ it is object-orientated, but has in-built features that help enforce memory and thread safety. We will be investigating the performance of Rust compared to traditional HPC languages, its portability across CPU architectures, its suitability for both shared memory and distributed memory parallelism, and its ability to integrate with scientific libraries and with batch systems.

The other investigation is focussed on task-aware communication libraries. Programming using tasks with data dependencies is a powerful method for minimising idle time due to excessive synchronisation and load imbalance in parallel applications, and is supported by APIs such as OpenMP and OmpSs. Implementing a full tasking model efficiently on distributed memory is very challenging, so a practical compromise is to use a hybrid of tasks within a node and traditional communication library such as MPI or GASPI between nodes.

Hartree Centre activities

The work described in the three focus areas above is complemented by our partners at the Hartree Centre who, amongst other activities, are working to create a supportive environment in which organisations can explore the latest digital technologies and skills, developing proofs-of-concept and applying them to industry and public sector challenges. They have also taken the lead in exploring different ways of engaging industry in training activities, holding a series of Design Thinking Workshops with the involvement of teaching staff and customers.

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 951732.

Read more about the programme's work and achievements.

www.eurocc-access.eu

Getting to know Smart Data Foundry

We were born out of the Open Banking industry.

Kimberley Mitchell,
Smart Data Foundry
kimberley.mitchell@ed.ac.uk

In 2018 new regulation enabled people to move their consented banking data easily and securely across banks and other providers and between banking apps. One of the main motivating aims of the UK government was to increase the levels of competition in the UK market to enable new Financial Technology (FinTech) start-ups to collaborate or compete with the big high-street banks.

The big promise of Open Banking, and later Open Finance (which extends the technology and regulation into all aspects of your financial life), was that it would enable new services. The benefits would be more convenient banking and the unleashing of the data we create in our daily financial transactions to put it to good use.

This last promise has been harder to realise. Partly because innovating around any data is complex and partly because some of the exciting and interesting things you can do with data carry risks of compromising privacy or enabling financial crime.

About the same time that Open Banking launched, the University of Edinburgh, FinTech Scotland (the organisation supporting the Scottish FinTech cluster), and FDATA (Financial Data and Technology Association: the global association for companies operating in Open Banking) conceived an idea for building a not-for-profit organisation that could enable the private, public, and voluntary sectors to collaborate around unlocking financial data for good. This was Smart Data Foundry's purpose.

In 2020, thanks to early support from the Edinburgh Futures Institute (EFI), part of the Data-Driven Innovation (DDI) City Region Deal programme, the team built a business plan and secured its initial funding from UKRI.

How we keep data safe

We understand the responsibility of working with people's data and strive to do this in the most secure and least intrusive ways available. When we utilise data from a third party, such as a commercial bank or credit agency, we follow a robust information governance model to ensure data is always treated safely. This data can primarily be described as pseudonymised or de-identified – and since we do not hold the pseudonymisation keys, most of the data processed can be considered 'effectively anonymised.

1. It starts with clear, documented agreements between Smart Data Foundry and the provider of data (eg NatWest Group), using Non-Disclosure Agreements (NDAs) and Data Sharing Agreements (DSAs) to agree how data will be shared, stored and used. This includes collecting only the minimum data fields necessary for the research outcomes.

We first launched as the Global Open Finance Centre of Excellence. Since then, the Open Banking and Open Finance world has inspired data sharing movements in many other sectors: energy, telco, and transportation. People are now speaking about smart data to encapsulate the benefits of wider data sharing and consumption.

While there was much value in unlocking financial data, we found that some of today's biggest challenges require bringing together many sorts of data. To reflect our original purpose we relaunched as Smart Data Foundry: an organisation designed to unlock financial data for good – not only to do good but also to create social impact that lasts for good.

2. The supplier of the data works to de-identify the data, making sure that the data contains no Personally Identifiable Information (PII).

3. Smart Data Foundry runs a Data Protection Impact Assessment (DPIA) test to make sure there is limited possibility of re-identification of the de-identified data and agrees further controls where necessary.

4. We then work with the data supplier to make sure they have completed a similar DPIA process to assess risks and mitigations.

5. The data suppliers use an encrypted secure file transfer tool to transfer the data to our EPCC-operated Data Safe Haven.

6. Trained Smart Data Foundry Information Governance staff triage the data to make sure no PII has been transferred accidentally and complete final QA checks.

7. Once in the Safe Haven, only named and vetted individuals can access the data for research and analysis purposes, using vetted data analysis tools. Researchers are unable to access the internet when processing the data.

8. Any published results are checked to ensure the data is sufficiently aggregated to prevent disclosure of individual data points.

Our current missions

Stop the Squeeze. By providing data on financial wellbeing and resilience, we will help Government and industry take real actions to help people on the poverty line survive and thrive.

Countering Climate Change. We are working with academics who understand how satellite data can be better interpreted to reveal patterns of activity on the planet's surface that indicate positive or negative climate change activity. We are also producing enriching sandbox environments supporting innovation in new ESG-related FinTechs by creating synthetic data sources for those businesses to test their business models against.

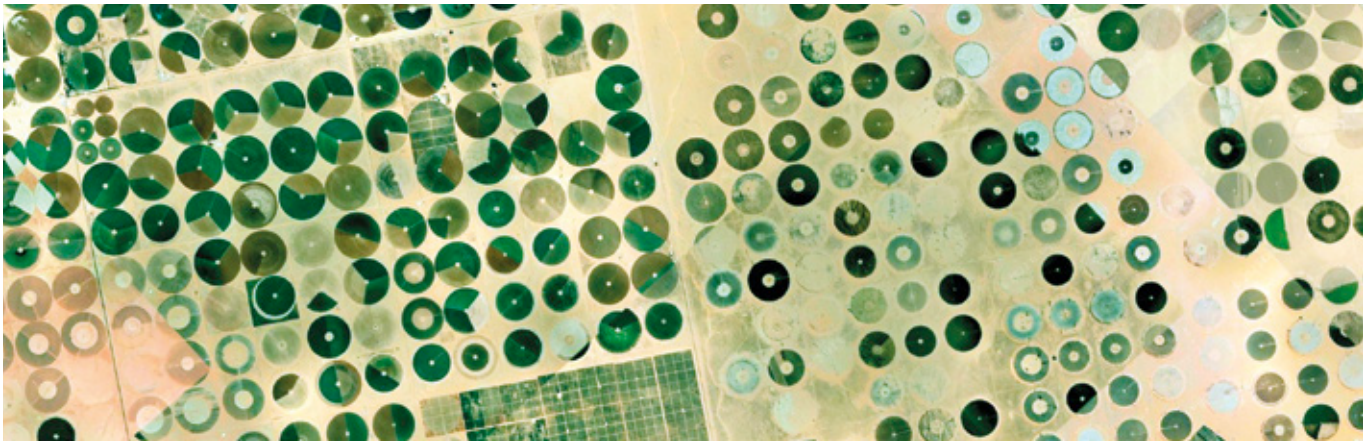
Open Finance for All. We see the future of Open Banking as the blueprint for Open Finance and Smart Data, available to all parts of society (especially traditionally excluded groups), designed to protect against crime, resilient to failure and future shocks.

Strong Small Business. Smart Data Foundry recently partnered with Sage Group, FreeAgent and Equifax to work with their data. We are talking with three of the UK's major High Street banks to contribute banking data to create a unique view of what's really going on. We'll share these insights with Government and others to inform decisions about how to support the sector, meaning better lending options for small businesses, better cash flow, and better productivity accelerating the recovery of the wider UK economy.

As part of our Data Protection Impact Assessment process, we ensure our legal basis of 'legitimate interest' to hold data for research is robustly challenged and recorded to ensure we are handling data appropriately to achieve our purpose of research in the public interest. We also regularly monitor re-identification risks and ensure we have appropriate controls to prevent this, even in the case of a motivated intruder.

SMART DATA FOUNDRY

Smart Data Foundry
www.smartdatafoundry.com



Satellite view of agricultural fields in Saudi Arabia (Google Earth, Maxar Technologies, CNES/Airbus, Landsat/Copernicus, 2021). Taken from Inés Cámara Leret's The Overlay, an early prototype work in part created/inspired by the TNRO platform and which was exhibited at the 2022 Edinburgh International Science Festival.

Art and artificial intelligence

We have been working with The New Real Observatory (TNRO) to develop a platform in which artists can apply artificial intelligence (AI) techniques to produce new digital artworks informed by the climate crisis.

The main technique currently being used in the TNRO platform is the application of Generative Adversarial Networks (GANs) to images, although GANs can also be used in wider application domains and there are plans to incorporate audio and text input/outputs capabilities.

In a GAN a pair of neural networks (a generator and a discriminator) are tuned in a zero-sum game on a supplied input data set. This could be eg a combination of images of Scottish plants and cacti. The training of the neural networks then proceeds in a double loop: in the inner loop the generator attempts to create images that look like Scottish plants and/or cacti; in the outer loop the discriminator is trained, using the same data set, to distinguish those images that look like Scottish plants/cacti from those that do not.

The training of a GAN can be very computationally expensive so TNRO employs pretrained GANs, with artist-supplied images used for final tuning – even this can take from a couple of hours to a day on a GPU. Once trained the GAN operates relatively quickly and can transform an image in seconds.

All the supplied images can be mapped to a multi-dimensional abstract space called the latent space. The artist will supply two classes of images – in the latent space a line can be set from the centroid of one of the image's classes to the centroid of the other class so moving along this line can make a new 'visual dimension', as imagined by the artist, that spans Scottish flowers to cacti.

Another aspect of the TNRO platform now comes into play. Using data from the ECMWF Copernicus Climate Data Store that has been incorporated into the TNRO platform, the artist can project the change in temperature, wind or rainfall at a given location on Earth within the next 70 or so years using one of three projection models – from pessimistic to optimistic – to determine whether their image should eg look more like a cactus or a Scottish flower.

This is only one possible use case, and the artists are using the platform in more imaginative and abstract ways.

Mario Antonioletti, EPCC
m.antonioletti@epcc.ed.ac.uk

The New Real Observatory (TNRO) is a sub project of The New Real. The TNRO platform is being developed by Informatics, EPCC, the School of Engineering and the Edinburgh College of Arts all at the University of Edinburgh and is funded by funded by the EPSRC via the Alan Turing Institute. The New Real is based at the Edinburgh Futures Institute.

The New Real Observatory
https://newreal.cc/

The Overlay by Inés Cámara Leret
https://newreal.cc/artwork/the-overlay

HPC-Europa3 visitors in the Bayes Centre: (left to right) Rodrigo Arias, Jorge Baeza, Damla Serper, and David Albandea.

End of an era

The HPC-Europa3 programme finished in April 2022 after five years. Although EPCC has been involved in similar Transnational Access programmes since 1993, COVID-19 meant HPC-Europa3 was like no other programme before it.

In March 2020 visits ended abruptly as restrictions came in and programme participants scrambled to return home early before international borders closed.

For months no visits took place, and by autumn 2021 our target number of visitors looked far out of reach, with only seven new visits having started since the COVID-19 pandemic began. However, as travel restrictions eased, visitors were finally able to plan their long-postponed visits, and application numbers rose sharply again for the last two closing dates.

EPCC finally supported 171 visits over the five years, with 60 of these taking place in the final four months of the programme, which was an extremely hectic period. This was despite some 27 visits being cancelled, mainly due to the pandemic.

Adapting to the circumstances, HPC-Europa3 introduced "virtual visits" for those who were unable to reschedule their visit in the available timeframe. Virtual visits had been offered under the previous HPC-Europa programme, but there was little interest in them and their value was not appreciated compared to in-person visits. However the pandemic has changed the way people work and it has been great to see that the majority of the 27 recent "virtual visits" have been highly successful, with the participants extremely enthusiastic about their experiences.

In total visitors came from 30 different countries, with most from EU countries and Associated States such as Norway and Switzerland, and a few from further afield, including Mexico, Iran, South Africa, Singapore and New Zealand.

The majority of visitors still come from chemistry and physics, where HPC has long been used, but life sciences are increasingly represented, with many projects focused on drug design and cancer research in particular and, since 2020, on COVID-19 research.

Catherine Inglis, EPCC
c.inglis@epcc.ed.ac.uk

As yet there is no news of an HPC-Europa4 programme, but we are actively seeking an appropriate opportunity to apply for Horizon Europe funding to continue this immensely successful programme, which many past visitors have cited as a key moment in their research career.

Infrastructure on High Performance Computing
Pan-European Research
HPC Europa3

www.hpc-europa.eu

Updates from the Software Sustainability Institute

From research to workshops, the Software Sustainability Institute has been busy promoting the vital role software plays in research.

Research Software Camp: Next Steps in Coding

Our latest Research Software Camp ran in May and focused on improving computational and training skills. There was a series of workshops (including developing intermediate research software skills), a mentorship programme and software surgery, as well as resources published on our website. All events were online and free to attend. Read a review on our website: www.software.ac.uk

Ten years of RSE

In March we celebrated the tenth anniversary of the coining of the term Research Software Engineer (RSE) at the SSI's Collaboration Workshop 2012, reflecting on how far the RSE movement has come over the last decade. People shared how the movement has impacted them using #RSE10years on Twitter.

New SSI Fellows

Earlier this year we welcomed a new cohort of 26 Fellows - including our first four international Fellows! We received many outstanding applications, and our new research software ambassadors represent some of the best people working in – and advocating for – better research software. Find out about the new Fellows at: www.software.ac.uk/news/announcing-2022-software-sustainability-institute-fellows

Code for Thought podcast

We've joined forces with the Code for Thought podcast to bring listeners episodes from the SSI on topics ranging from 'software horror stories' to meeting the new SSI Fellows and hearing about their research. Listen to the podcast: https://codeforthought.buzzsprout.com/

Collaborations Workshop 2022

At the start of April we ran our Collaborations Workshop, which focused on the themes of Code Review, Ethics, Hybrid Working, and Software Sustainability. Our Collaborations Workshops bring together researchers, developers, innovators, managers, funders, publishers, policy makers, leaders and educators to explore best practices and the future of research software. Find videos from the event at: www.youtube.com/c/SoftwareSaved

Jacalyn Laird, EPCC
j.laird@epcc.ed.ac.uk

The Software Sustainability Institute cultivates better, more sustainable research software to enable world-class research.

We help people build better software, and we work with researchers, developers, funders and infrastructure providers to identify key issues and best practice in scientific software.

#RSE10Years

UKRI community studies

The Software Sustainability Institute has been delivering a series of studies of different UKRI communities to better understand their research software needs and requirements.

ESRC study

The Economics and Social Sciences Research Council (ESRC) awarded the SSI funding to conduct a study that improves the understanding of software and data use in projects within the social sciences research areas. The survey of digital methods and software in the social sciences closed in March and we received over 165 responses and also conducted a series of interviews. The data will be analysed and collated into a report that will be submitted to the ESRC and be made available to the public.

AHRC study

Last year we worked with the Arts and Humanities Research Council (AHRC) to deliver a survey of the digital and software requirements of the AHRC research community. The survey asked about views on digital tools/software, experiences of developing these, and practices and preferences for recruiting help with digital tool/software development. The results will inform AHRC's digital infrastructure funding to better align with the communities' needs.

UKRI Digital Research Infrastructure EPSRC study

We are currently running a study funded by the Engineering and Physical Sciences Research Council (EPSRC) and the Digital Research Infrastructure Phase 1 Fund to better understand the software and skills required for large-scale research computing. We are gathering data about the software and skill needs in High Performance Computing, High-Throughput Computing, Cloud Computing, Artificial Intelligence/Machine Learning and Data Science.

Find out more: https://edinburgh.onlinesurveys.ac.uk/survey-on-software-skills-and-infrastructure-for-research-6

A summary of all research studies carried out by the SSI can be found at: www.software.ac.uk/research-studies.

The Software Sustainability Institute represents the needs of software users and developers in the research community. We lobby for the recognition of the role of software in research, better software education for the research community, recognition of the role of people who develop research software and better engineering of software to provide confidence in the results that software generates.

Read more about the Institute's work: www.software.ac.uk

EPCC MSc programmes: supporting students

Marharyta Marko via Getty Images

Our Masters programmes are about more than just extensions and coursework. We work hard to ensure our students feel part of the team.

Supporting students in their studies can sometimes feel like the individual gets lost in the 'process', especially in a large organisation. Discussion or assistance regarding a practical exercise leads to work towards an assessment, anonymously marking an assessment leads to the exam board, conducted anonymously; a struggling student applies for an extension or special circumstances, which goes to the exam board via a special circumstances committee.

"I'm really surprised how much you guys care about us. Honestly it's quite touching." Anwar Orabi, first year part-time (online) MSc High Performance Computing with Data Science student.

As EPCC scales up its online MSc programmes (10 entrants in 2020/21, 25 in 2021/22) it is wonderful to see feedback like this and credit especially should go to Weronika Filinger in her role as Online Programmes' Director.

Being able to break that cycle of impersonality when supporting students can be important, especially in this post-lockdown world where many interactions (by design in the case of our online programmes) may still take place through a computer screen with or without a camera.

Building those interpersonal connections can be just as important to staff enjoyment of teaching on and supporting the MSc as it can be a hugely important part of the process for students: both in terms of developing graduate attributes and in being comfortable asking for help.

For on-campus students a year is such a short time to achieve this, while for online part-time students who may take up to six years to complete the programme maintaining at least some synchronous discussion-based time as accessible as possible to students in multiple timezones is absolutely critical.

This connection applies to both EPCC's administrative staff (Ben Morse and Jemma Auns) in the Postgraduate Programmes team in their roles as both Course/ Programme Administrators and Student Support Officers, but also our technical and academic colleagues in their roles as Course Organisers, Dissertation Supervisors, Lecturers, and Personal Tutors.

Ben Morse, EPCC
b.morse@epcc.ed.ac.uk

"The best thing about the MSc was being able to take a break and grab your lunch or a coffee next to an expert like Mark Bull and just talk about OpenMP, even months after that course."

Jonas Faßbender, MSc High Performance Computing with Data Science 2020 graduate

EPCC Masters programmes
www.epcc.ed.ac.uk/education-and-training/masters-programmes

Get in touch
On-campus MSc
msc@epcc.ed.ac.uk

Online programmes
online.learning@epcc.ed.ac.uk

ISC 2022 Student Cluster Competition

TeamEPCC 2022: (left to right) Alexander Menegas, Worapol Boontanonda, Alex Woods, Anukrat Bhansali.

We are TeamEPCC, a team of four MSc students from EPCC who will participate in the ISC22 on-site Student Cluster Competition (SCC) in Hamburg, Germany in June. The ISC22 SCC includes both the digital competition and the on-site competition. The competition will form the basis of each team member's MSc dissertation project.

TeamEPCC was one of the five student teams selected to compete on-site this year – we will run a list of benchmarks and scientific applications on our own cluster with the aim of delivering the highest performance within the limit of 3 kilowatts of power. We look forward to the competition challenges and meeting the academic and industrial experts in the HPC world as well as the other SCC teams at the ISC22.

In preparation for the competition, our self-designed cluster has been set up with the support of the EPCC system team. The announced benchmarks and applications for the ISC22 on-site SCC include HPCC, HPCG, HPL, ICON, Xcompact3D and NWChem. Due to the nature of the applications and benchmarks, we chose a GPU-based cluster consisting of two nodes with each node containing eight NVIDIA A100 GPUs, two AMD 7713 64-core CPUs, and 1TB RAM. A node is based on the HPE XL675d Gen10 product server. It uses the HPE Apollo d6500 Gen10 chassis and is cooled by fifteen 80mm fans.

We have tried to port, run and optimise all the codes with various configurations in order to find those which perform the best on our cluster. Of particular concern is the power limitation; the cluster has the potential to run at well over the 3 kilowatt limit, so an important task for many applications is to discover the optimal way to reduce the cluster's power consumption below this limit while still achieving the best performance.

As a part of our preparations for the competition, our team visited EPCC's Advanced Computing Facility (ACF), where several UK national HPC systems, such as ARCHER2, Cirrus, and Tesseract, are located. Our own designed cluster is also in the ACF.

During our training visit there we tried to learn how to perform basic maintenance on the hardware as well as seeing how its constituent components fit together and, importantly, were cooled. In addition we had the chance to see the ARCHER2 and Cirrus supercomputers, both of which had been used extensively by all team members as part of the High Performance Computing/with Data Science MSc. Overall, it was a very enlightening and enjoyable experience.

Xu Guo, EPCC
x.guo@epcc.ed.ac.uk
Spyro Nita, EPCC
s.nita@epcc.ed.ac.uk

Hewlett Packard Enterprise

We would like to thank HPE for being our team's sponsor.

Read more about this year's Student Cluster Competition at ISC:

www.isc-hpc.com/student-cluster-competition.html

Image: Paul Dodds

Study HPC with us

Master's degrees in High Performance Computing (HPC) and in HPC with Data Science

EPCC is the UK's leading supercomputing centre. We are a major provider of HPC training in Europe, and have an international reputation for excellence in HPC education and research.

Our MSc programmes in High Performance Computing (HPC) and HPC with Data Science have a strong practical focus and provide access to leading edge systems such as ARCHER2 (the UK's National HPC Service), and Cirrus (an EPSRC Tier-2 National HPC facility including over 150 GPUs).

MSc students have the opportunity to undertake their dissertations as an industrial project, building on EPCC's strong business links. Recent project partners range from start-ups to multinationals.

Optional course choices include modules from the School of Informatics and the wider College of Science and Engineering.

Our graduates are in high demand in both academia and industry in the UK and abroad.

The University of Edinburgh is ranked in the top 30 universities in the world by both Times Higher Education World University Rankings 2021 and QS World University Rankings 2021.

"The MSc in High Performance Computing course had a really good practical focus which allowed me to develop my HPC programming skills and get hands-on experience working on applications using the skills I learned during the MSc on a daily basis."
Holly Judge
MSc student 2019
Now an Applications Developer at EPCC.

"With one of the most experienced teaching and research staff, combined with its cutting-edge supercomputers, EPCC enabled me to establish the fundamentals of HPC and Data Science and to build a solid base for my further academic pursuit."
Liang Liang
MSc student 2020
Now a PhD student at Imperial College London.

www.epcc.ed.ac.uk/msc