SCIENTIFIC COMPUTING WORLD

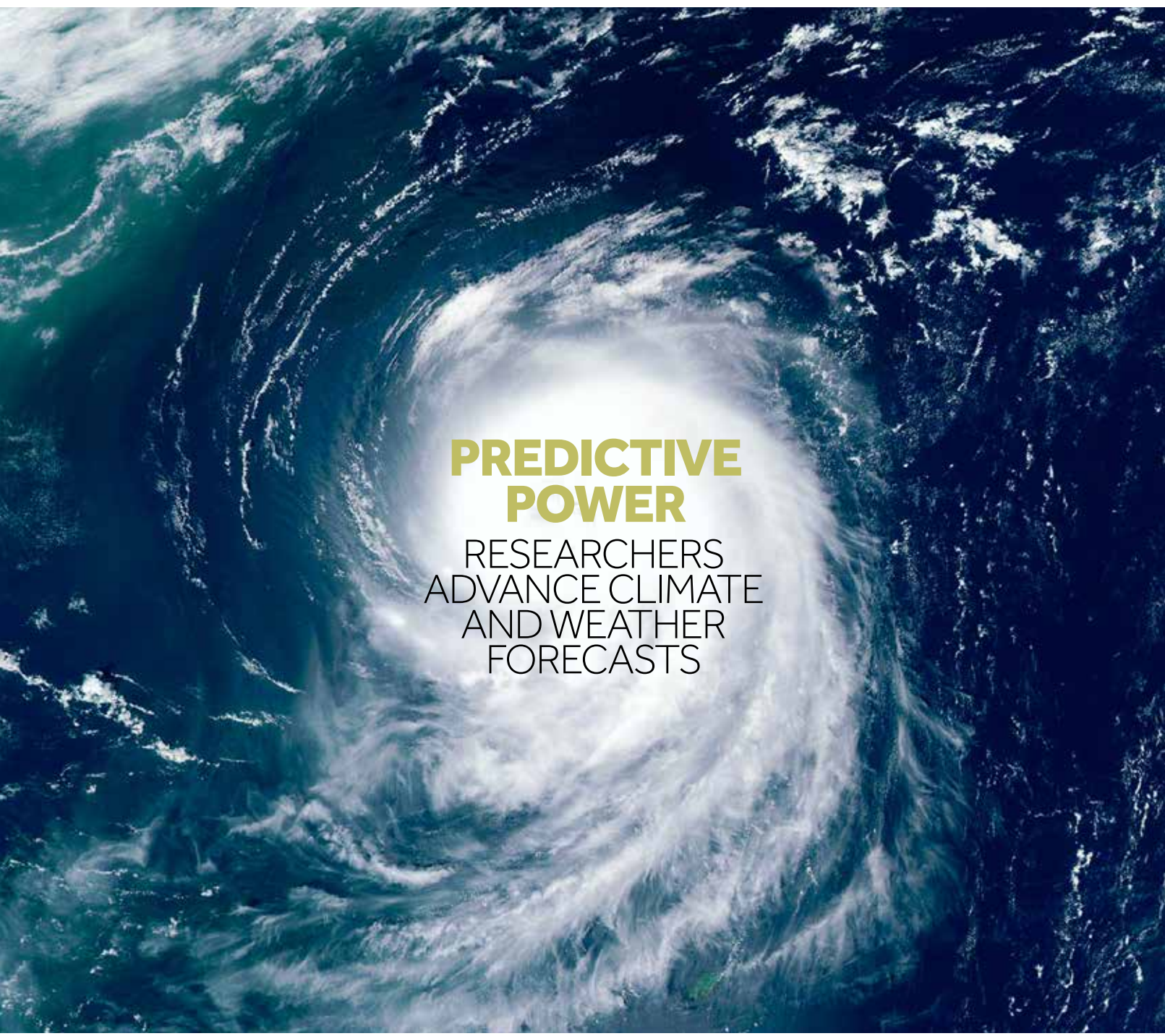
Computing solutions for scientists and engineers

Spring 2021
Issue #175

High performance computing
Accelerating AI

Laboratory informatics
Laboratory Integration

Modelling and simulation
Enhancing race safety



PREDICTIVE POWER

RESEARCHERS ADVANCE CLIMATE AND WEATHER FORECASTS

www.scientific-computing.com

AMD RYZEN™ THREADRIPPER™ PRO PROCESSORS

Up to 64 cores with professional security and manageability features

AMD THREADRIPPER PRO | BOSTON Servers I Storage I Solutions

SUPERMICRO 5014A-TT WORKSTATION

IDEAL FOR: **ARTISTS, ARCHITECTS, ENGINEERS, AND DEMANDING PROFESSIONALS**

KEY FEATURES

- AMD RYZEN™ THREADRIPPER™ PRO
- UP TO 64 CORES/128 THREADS
- UP TO 4.3GHz MAX BOOST
- 8 DIMM SLOTS; UP TO 2TB
- 4x DUAL WIDTH GRAPHICS CARDS
- PCIe® GEN 4 SUPPORT AND M.2 STORAGE
- REGISTERED ECC DDR4 3200MHz SDRAM

TEST DRIVE THIS SOLUTION AND MORE TODAY IN BOSTON LABS

SCAN ME

WEB: WWW.BOSTON.CO.UK | **EMAIL:** SALES@BOSTON.CO.UK | **PHONE:** +44 (0) 1727 876 100

Contents

Spring 2021 ● Issue 175

High performance computing

Laboratory informatics

Modelling and simulation

Resources

Editorial and administrative team
Editor: Robert Roe
robert.roe@europascience.com
Managing editor: Tim Gillett
editor.scw@europascience.com
Specialist reporters: Sophia Ktori, Gemma Church
Design: David Houghton, Zöe Andrews
Partnership & events executive: Charlie Mitchell
charlie.mitchell@europascience.com

Advertising team
Advertising manager: Mike Nelson
mike.nelson@europascience.com +44 (0) 1223 221039
Production manager: David Houghton
david.houghton@europascience.com +44 (0) 1223 221034

Corporate team
Managing director: Warren Clark

Scientific Computing World is published by Europa Science Ltd, 4 Signet Court, Cambridge, CB5 8LA ● ISSN 1744-8026
Tel: +44 (0) 1223 211170 ● Fax: +44 (0) 1223 213385
Web: www.researchinformation.info

Subscriptions: Free registration is available to qualifying individuals. Register online at www.scientific-computing.com Subscriptions £180 a year for six issues to readers outside registration requirements. Single issue £20. Orders to ESL, SCW Circulation, 4 Signet Court, Swann Road, Cambridge, CB5 8LA, UK. Tel: +44 (0)1223 221 030. Fax: +44 (0)1223 213 385.

US copies: Scientific Computing World (ISSN 1356-7853/USPS No 018-753) is published bi-monthly for £180 per year by Europa Science Ltd, and distributed in the USA by DSW, 75 Aberdeen Rd, Emigsville PA 17318-0437. Periodicals postage paid at Emigsville PA. Postmaster: Send address corrections to Scientific Computing World PO Box 437, Emigsville, PA 17318-0437.

Cover: Zenobillis/Shutterstock.com

Subscribe for free at
www.scientific-computing.com/subscribe

Robert Roe
Editor

Feel the power

In this issue of *Scientific Computing World* there are several pieces focused on the ways that scientists are benefitting from technologies enabled by the increasing availability of powerful computing platforms – from drug discovery to genomics, weather simulation and even deep learning in hospitals.

While advanced computing is typical in high performance computing, there are growing numbers of scientists that need to use supercomputing resources and so training and skills development is the focus of the first feature in this issue, found on page 4. Next are the two tech focus articles on page 8 and 10 respectively. These articles take a broad look at the technologies available for HPC users in both storage and cooling.

Page 17 kicks off the laboratory informatics content for this issue with the second part of Sophia Ktori's feature on software integration. On page 20, there is an article exploring the use of genomics and software that can help researchers make better use of their research data.

The focus switches to drug discovery on page 24 and the use of outsourcing and managed services to deliver efficiency and performance increases in both software and hardware. On page 28 is a story detailing the UK National Health Service (NHS) rollout of a deep learning-powered 3-D imaging technology to diagnose and treat patients with suspected heart disease. Page 30 takes a look at FAIR research and the importance it has on life science innovation.

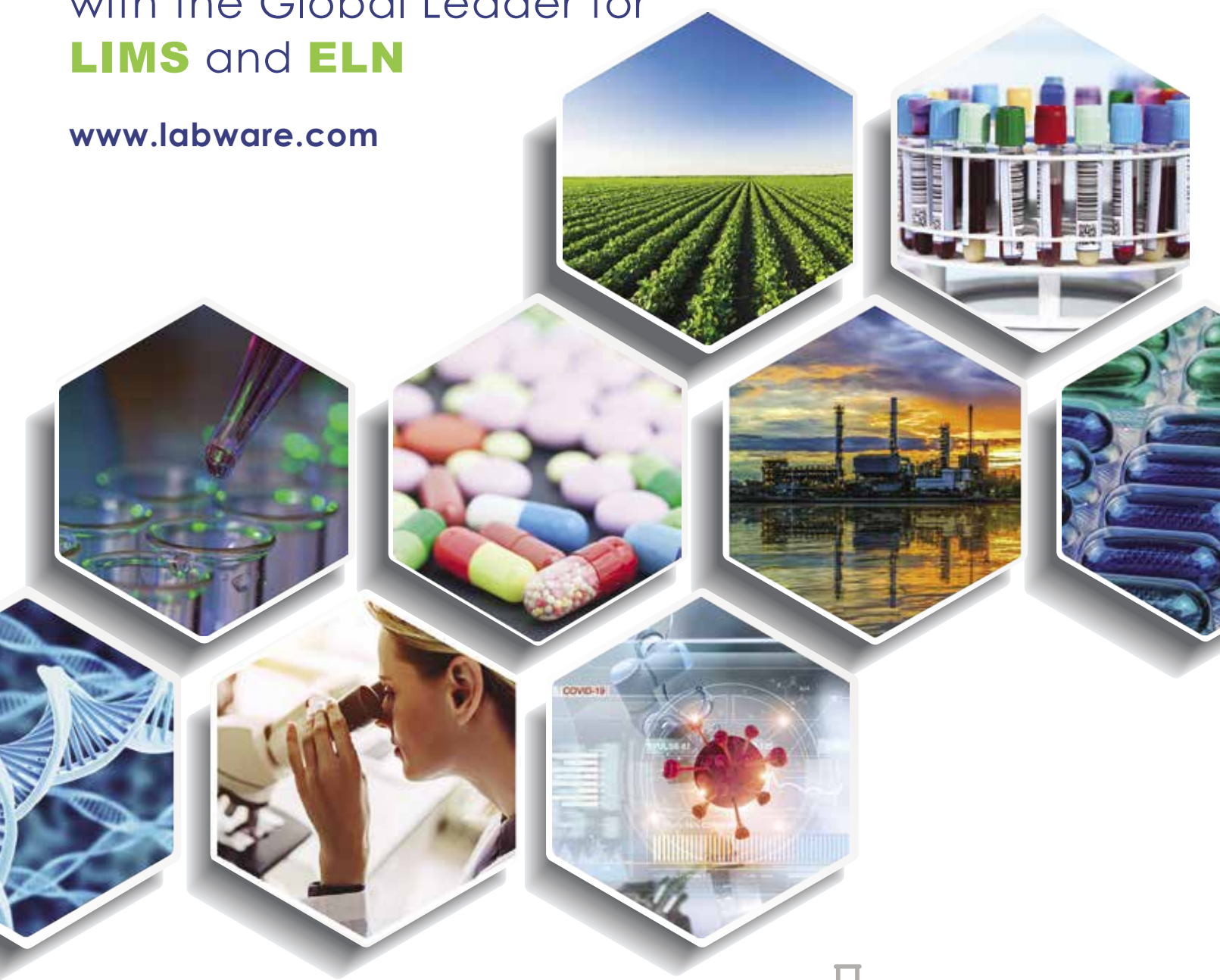
Gemma Church takes a look at weather simulation on page 32. Detailing how simulation and modelling is helping researchers unlock long-term, accurate weather and climate predictions.

@scwmagazine

LABWARE® 8

Automate Your Laboratory
with the Global Leader for
LIMS and **ELN**

www.labware.com



LABWARE®
Results Count


MERCK
Dive into your data
BSSN™ SciDiver lets you access and analyze your laboratory data with anyone, anywhere, on any device.
Scientific breakthroughs aren't confined to the bench. Collaborate with remote colleagues and connect your instruments with SciDiver, a secure cloud solution for organizing, harmonizing, and visualizing all your analytical and biological data.
Learn more at bssn-software.com
BSSN Software is now part of MilliporeSigma
The life science business of Merck operates as MilliporeSigma in the U.S. and Canada.
© 2020 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved. Merck, the vibrant M and BSSN Software are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources. 34066 11/2020

Developing skills for HPC

ROBERT ROE TAKES A LOOK AT THE TRAINING AND SKILLS DEVELOPMENT OPPORTUNITIES FOR SCIENTISTS USING HPC IN THEIR RESEARCH

Training and the development of specific skills to use HPC are becoming increasingly important as the number of users and potential applications continues to rise. Scientists and engineers in many disciplines can make use of HPC, it is not just a technology reserved for climate science and large scale astronomy simulations.

For many years most HPC users worked in national labs and were a relatively consistent set of employees involved in long-term research programmes. They were well supported by their research centres and Independent Software Vendors (ISVs) that continuously contributed time to training users and maintaining knowledge on the available tools. In contrast, many of the HPC users today have very different characteristics – they use HPC facilities intermittently and for short bursts of time. Today many more people need access to HPC facilities but do not necessarily have the experience and skills needed.

While these new users may be familiar with traditional cloud computing, which provides them with a bare-bones infrastructure to scale applications, they are less familiar with the HPC framework, which offers remote systems with pre-installed environments that are configured and targeted toward scientific computation.

Even though some uniformity does exist in the HPC world in terms of software stacks, the environments remain largely heterogeneous and therefore the users will be exposed to toolsets that they have never used before – or have very little experience with. As a result, these new users are not well-versed in the tools or systems, nor do they have a desire to gain in-depth, working knowledge of the infrastructure.

Furthermore, these users will rarely want to participate in full-scale training about a system they will only be using sporadically. This creates new challenges for developing successful training procedures for these HPC users.

"HPC training and education is a hugely important topic in the context of the EuroHPC Joint Undertaking"

In the typical HPC world, the most common type of training requests are either for an introductory course at the original purchase and delivery of the system, or for a set of advanced training packages for fine tuning the skillsets of seasoned users. There has been an increase in the number of on-demand users and this in turn changes the type of training requests. There is a noticeable shift from users looking to learn all the bells and whistles to users wanting to maximise the use of tools for a specific problem. This is because the goal of HPC for these users is to solve a particular problem, to run a set of applications and receive a result, rather than to become experts in the use of HPC.

As HPC users become more common, and the use cases become more ubiquitous, it stands to reason that some users just need HPC on-demand rather than the traditional model. How can HPC facilities support these sporadic or part time users alongside the more traditional users that demand frequent access to HPC systems.

HPC facilities are being challenged to provide more frequent and accessible training. In essence, they will have to take on a consumer-oriented model, in which the advice is narrowly focused on the task at hand, versus the model of a technical college that provides a full curriculum for developing an expert.

Entry level and regional HPC training and support

There are several opportunities for HPC training and skills development at various levels from introductory courses to application development and preparatory access to large scale tier-0 facilities. In this feature, we will discuss options available predominantly in the UK and Europe but this will by no means be an exhaustive list of the training options available to HPC users.

In the UK, training and skills development is predominantly provided by the Edinburgh Parallel Computing Centre (EPCC) and universities that have their own HPC resources. The EPCC, for example, is one of the major providers of training in high performance computing (HPC) in Europe, offering a range of courses for users of HPC throughout the UK and Europe.

ARCHER Training: The UK national supercomputing service, ARCHER, provides large amounts of HPC training that is free for all UK academics. Courses cover a range of abilities from beginner to advanced and are run at a variety of locations around the UK.

University of Bristol ACRC Training: The Advanced Computing Research Centre (ACRC) at the University of Bristol runs a number of HPC training courses and the majority of their material is freely-available online for people to study remotely.

Bristol offers training and support across several primary areas including Linux and specific training for the various clusters at the university's disposal. Before users are allowed to access the HPC systems they need to demonstrate some understanding of Linux and the University offers short courses on Linux and an intro to HPC training, access to the BlueCrystal HPC systems is also available for more advanced users or those with specific application requirements.

PRACE: The Partnership for Advanced Computing in Europe (PRACE) offers a number of different training opportunities including the PRACE Advanced Training Centre (PATC), PRACE training portal and access to advanced skills development PRACE training centres. These facilities deliver regular programmes of courses in many aspects of HPC and advanced computing. Bristol is an example of a university that offers several courses to support HPC users developing their skills. This includes introductory courses and more advanced offerings for experienced users. →

slas 2022

INTERNATIONAL CONFERENCE & EXHIBITION

5-9 FEBRUARY

Boston Convention and Exhibition Center | MA

GET HYPED for research

Call for Abstracts for Podium and Poster Presentations

- Tony B Academic Travel Award submissions due: **September 13** (podium); **October 18** (poster)
- Student Poster Awards submissions due: **October 18**

Nine Scientific Tracks

- Advances in Bioanalytics and Biomarkers
- Assay Development and Screening
- Automation Technologies
- Cellular Technologies
- Data Science & AI
- Micro- and Nano Technologies
- New Modalities
- 'Omics
- Precision Medicine and Diagnostics

Learn more at **SLAS.ORG/2022**

→ **PRACE and European training**

PRACE makes it possible for researchers from public and private institutions from across Europe and the world to apply for resources on high-end Tier-0 HPC systems via a centralised peer review process.

PRACE is a huge provider of training and support for new and existing HPC users. While many may think of PRACE as delivering tier-0 facilities to European researchers it also offers new users and industrial users opportunities to get access to HPC systems through the various programmes such as SHAPE and PRACE Preparatory Access. PRACE operates 14 PRACE Training Centres (PTCs) and they have established a state-of-the-art curriculum for training in HPC and scientific computing.

PRACE training courses are open to participants from all European countries. PTCs carry out and coordinate training and education activities that enable both European academic researchers and European industry to utilise the computational infrastructure available through PRACE and provide top-class education and training opportunities for computational scientists in Europe.

In addition, PRACE seasonal schools complement the PTC training program with three such events usually held throughout the year. One is usually held in autumn, one in winter and one in spring. Each of these is held in a non-PTC country (see next section) and at different geographical locations. Their curriculum is also different and usually varies to that of the PTC events. Registration for all PRACE Training courses is free and open to all. Specific courses can be found on the PRACE training portal.

In addition to computing time, support from a high-level support team (HLST) may be assigned to selected research projects. HLSTs will help projects of outstanding scientific value to further utilise the capabilities of PRACE Tier-0 systems through code optimisation.

HLSTs are available in combination to Tier-0 systems of the following PRACE hosting members: Grand Équipement National de Calcul Intensif GENCI, France; GAUSS Centre for Supercomputing GCS, Germany; CINECA – Consorzio Interuniversitario, Italy; Barcelona Supercomputing Center BSC, Spain; Swiss National Supercomputing Centre CSCS at the Swiss Federal Institute of Technology in Zurich (ETH Zurich), Switzerland.

Call for preparatory access

The objective of PRACE Preparatory Access is to allow PRACE users to optimise, scale and test codes on PRACE Tier-0 systems before applying to PRACE calls for Project Access. The next PRACE call for proposals for Project Access will most likely open in Autumn 2021. Production runs are not allowed as part of PRACE Preparatory Access. Currently, PRACE offers four different schemes for Preparatory Access based on the type of application and the maturity of the project.

"One of the biggest hurdles ... is providing educational access to HPC resources in a consistent way at the required scale"

LearnHPC

LearnHPC is a website set up to ensure that HPC is an accessible technology for the widest possible community of scientific researchers. The site acts as a gateway providing materials, resources and tools that will lower or remove barriers.

EU-wide requirements for HPC training are increasing as the adoption of HPC in the wider scientific community gathers pace. However, the number of topics that can be thoroughly addressed without providing access to actual HPC resources is very limited, even at the introductory level. In cases where such access is available, security concerns and the overhead of the process of provisioning accounts make the scalability of this approach questionable.

EU-wide access to HPC resources on the scale required to meet the training needs of all countries is an objective that we attempt to address with this project. The proposed solution essentially provisions virtual HPC systems in a public cloud. This infrastructure will allow us to dynamically create temporary event-specific HPC clusters for training purposes, including a scientific software stack. The scientific software stack will be provided by the European Environment for Scientific Software Installations (EESSI) which uses a software distribution system developed at CERN, CernVM-FS, and makes a research-grade scalable software stack available for a wide set of HPC systems, as well as servers, desktops and laptops.

Through the FENIX Research Infrastructure and AWS, LearnHPC offers the use of moderately sized clusters configured specifically for your training events. At present, there is no specific mechanism to request access to LearnHPC resources.

In a recent interview with FENIX Research Infrastructure, Dr Alan O'Cais, software manager for E-CAM Centre of Excellence at Forschungszentrum Jülich, discussed the role of LearnHPC and the drive to develop HPC skills across Europe. 'Through my involvement in the E-CAM Centre of Excellence and FocusCoE, I am aware that HPC training and education is a hugely important topic in the context of the EuroHPC Joint Undertaking. There is, however, an enormous logistical challenge in extending HPC training of a consistent standard to an ever-growing pool of researchers in 32 countries.'

'One of the biggest hurdles that I foresee is providing educational access to HPC resources in a consistent way at the required scale,' added O'Cais. 'In the context of HPC training, I wouldn't immediately draw a distinction between "students, researchers, and users", I would see them all as learners.'

'What LearnHPC will hopefully do for all learners is make the mechanics of accessing HPC training uniform, well documented and as easy as possible. We want to remove, hide or simplify the technical barriers that tend to increase the slope of the learning curve when it comes to HPC.

'Learners may still ultimately need to know about ip-restricted ssh keys or how to compile the latest GCC compiler from source, but these can be introduced at a more appropriate time in their learning journey.'

Additional resources

ARCHER » Training
www.archer.ac.uk/training

ACRC training, Advanced Computing Research Centre, University of Bristol
www.bristol.ac.uk/acrc/acrc-training

Training portal (prace-ri.eu)
https://training.prace-ri.eu

PRACE preparatory access guide
https://prace-ri.eu/hpc-access/preparatory-access/preparatory-access-open-calls

BlueCrystal Phase 4 user guide
https://www.acrc.bris.ac.uk/protected/bc4-docs

BC4 User documentation
https://sso.bris.ac.uk/sso/login

LearnHPC - Scalable HPC Training
http://www.learnhpc.eu

White Papers
now available online

VIEW FOR FREE*

The essential guide to managing laboratory samples
TITIAN SOFTWARE
This white paper draws on Titian Software's long running partnerships with multiple vendors of laboratory automation and its extensive experience of implementing sample management systems for organisations of all sizes, plus the author's 40 years' experience in the pharmaceutical and software industries.

Integration Strategies for Digitising your Lab - Part 1: Liquid Handling
TITIAN SOFTWARE
This white paper looks at the different approaches to integration and questions to consider when integrating liquid handling equipment in your lab

Integration Strategies for Digitising your Lab: Part 2: Automated Stores and Other Processes
TITIAN SOFTWARE
This section of the white paper explores how these different levels of integration apply to both automated stores and to manual processes.

Time is Money: The Hidden Cost of Inefficient Laboratory Practices
MERCK
This whitepaper presents an overview of the main inventory management challenges faced by fast-paced biotech and pharmaceutical laboratories, highlighting how this directly impacts research outcomes, costs and regulatory compliance.

Achieving Data Integrity Through Starlims Standalone Scientific Data Management System
ABBOTT INFORMATICS
Learn more about how STARLIMS Standalone SDMS can support organizations to efficiently capture and parse documents and data, reducing the tedious, time consuming and error-prone manual data entry.

Boosting productivity of monoclonal antibody development using Chromeleon CDS: A collaboration to move biopharma forward
THERMO FISHER SCIENTIFIC
Using the right tools is of paramount importance in the effort to advance tomorrow's therapies.

Getting started with AI software
NVIDIA
Learn about the developer tools and optimized algorithms your teams can use to expedite workflows and turn data into insights faster than ever.

Implementing Sample Management Software Your 7-Step Guide to data migration
INTERACTIVE SOFTWARE LIMITED
Moving from spreadsheets to a LIMS? Interactive Software Limited's 7-Step Guide helps you standardise and improve the quality of your existing data.

AnIML: Concepts & Applications
MERCK
This whitepaper presents an overview of AnIML, the Analytical Information Markup Language. It is intended to assist decision makers who are evaluating the use of AnIML as a data format.

*Registration required

www.scientific-computing.com/white-papers

SCIENTIFIC COMPUTING WORLD

Tech Focus: Storage

A ROUND-UP OF THE LATEST STORAGE PRODUCTS AND TECHNOLOGY AVAILABLE FOR SCIENTISTS USING HPC AS A TOOL FOR THEIR RESEARCH

The storage market is in a unique position, in that there is demand from both the traditional HPC and enterprise storage markets, such as media and entertainment, alongside new growing markets for AI and machine learning. This has created huge potential to increase market share for storage vendors, as long as they can deliver the necessary performance required by HPC and AI users.

While storage volumes continue to increase dramatically, storage providers are trying to meet demand by increasing storage performance, while introducing more efficient methods of managing data across large multi-petabyte storage platforms.

Choosing the right system for a particular workflow is critical to getting the most out of storage technology. The traditional products associated with parallel file systems still persist, but now there is increasing competition from cloud and all-flash storage arrays, which are becoming increasingly attractive to users at opposite ends of the hardware spectrum.

More in store

Atipa's Capella storage solutions are scalable, affordable, reliable and high performance. The Capella LZE series storage solution is its flagship Lustre appliance, including high-availability and advanced monitoring features. The starting configuration consists of one Lustre metadata/management module and one Lustre object storage module.

BullSequana Xstor is a single family of storage appliances from **Atos**. With machine learning and data analytics creating many of the most challenging HPC workloads in the move towards the exascale era, Atos participates in initiatives to strengthen industrial control of digital simulation and big data technologies.

BeeGFS is a parallel file system, developed and optimised for high-performance computing. It includes a distributed metadata architecture for scalability and flexibility, and is available free for end-users. For enterprise systems, professional commercial support is available, usually in co-operation with international turn-key solution partners.

Caringo's Swarm software-defined object storage provides a complete platform for data protection, management, organisation and search at massive scale for all unstructured data (e.g., photos, videos and files). Users no longer need to migrate data to separate solutions for long-term preservation, delivery and analysis. Consolidate all files in Swarm, then easily find the data.

ClusterVision has 16 years of knowledge and expertise working with different storage solutions, with straight-forward to very complex installations. It can offer classic NAS systems, (hybrid) cloud solutions, parallel HPC storage installations with BeeGFS, IBM Spectrum Scale or Lustre, and many more configurations.

DDN is one of the premier providers of high-performance storage solutions, voted top in high-performance storage for AI and HPC computing in 2020 (HPCwire Readers' and Editors' Choice Awards). With DDN A3I and EXAScaler data management solutions, organisations can analyse petascale datasets, with capacities up to 500+ PB, and throughput up to 2TB/second.

Excelero's NVMesh software-defined storage deploys distributed data protection over public cloud instances with ephemeral NVMe drives, creating a unique, virtual, high-performance and low-latency storage pool. Excelero delivers low-latency distributed block storage for web-scale applications. NVMesh enables shared NVMe across any network and supports any local or distributed file system.

HPC using **Fungible** solutions eliminates lengthy delays and lost output typically related to legacy HPC clusters. Elastic, high-performance configurations that support practically unrestricted scalability allow you to grow and contract infrastructure as applications require.

HPE provides HPC storage solutions that span the whole storage hierarchy to accelerate time-to-insights, while managing and protecting valuable data in parallel file systems in a cost-effective way. Its Cray ClusterStor E1000 storage system embeds Lustre and is ideal for attachment with HPE Slingshot, InfiniBand HDR and 100/200 GbE to HPC

Cray EX supercomputers, and large clusters of HPE Apollo systems.

IBM Spectrum Scale is a parallel high-performance solution with global file and object data access for managing data at scale, and can perform archive and analytics in place.

MemoScale specialises at developing compression and erasure coding software for data storage solutions. It has provided technical solutions and has extensive experience in software and hardware optimisations.

MooseFS allows the combination of data storage and data processing in a single unit using commodity hardware, providing an extremely high ROI. Through this approach, it provides professional services and expert advisory for storage solutions.

NEC's Storage HS Series platform enables long-term data retention through scalability of performance and maximised capacity without the complexity and inherent limitations of legacy storage solutions.

NetApp has a proven track record when it comes to modernising and simplifying storage environment. From simple, smart, trusted storage for shared NAS and SAN environments, to arrays built for dedicated, high-bandwidth applications like data analytics and disk-based backup, it can design the perfect solution.

Open-E DSS V7 is a fully featured NAS and SAN software platform. NAS offers support for multiple file protocols which enable you to build cross-platform solutions and SAN offers iSCSI Target and Fibre Channel Target functionalities.

The flexible design and excellent scalability of Open-E DSS V7 enables companies and institutions of any size to create effective storage solutions to meet and adapt to the simplest or most complex data management needs.

Panasas supports industry and research innovation around the world with PanFS Dynamic Data Acceleration that automatically adapts to evolving workloads to deliver a consistently fast, total-performance HPC storage solution. PanFS is delivered on the ActiveStor Ultra turnkey appliance, to maximise simplicity, boost reliability and provide the lowest TCO.

Pure Storage aims to empower innovators by simplifying how people consume and interact with data. FlashArray//C delivers NVMe performance, hyper-consolidation, and simplified management. FlashArray//X is designed for entry-level to enterprise applications.

QCT HPC/DL solution integrates with open source software to offer a software stack that includes resource management for a fully end-to-end integrated HPC/DL solution. It also features a modularised architecture that can be easily tailored to meet customer demands, simplifying the implementation journey and accelerating time-to-market.

QNAP NAS is a reliable, accessible and fast storage solution for post-production workflows. TVS-473e transfers high-volume file and provides large storage for high-resolution 3D scanning.

Qualstar magnetic tape and data storage solutions are known for their high quality, ease of use and reliability. Tape has long been used as a digital data storage medium, but it is the rapid expansion of big data in recent years that has really made its value apparent.

Qumulo Core is a high-performance file data platform designed to help you store, manage and build workflows and applications with data in its native file form, at massive scale, across on-prem and cloud environments.

Quobyte is a software storage system that delivers unlimited performance and consistent low latency to applications and users through scale-out, without limits. Quobyte is designed around the core principle of unconditional simplicity, so anyone can easily manage hundreds of petabytes with many applications and users.

Scality helps scientists manage file and object data from cloud-to-core-to-edge. Scality Ring is a scale-out file and object storage, while Scality Artesca provides lightweight, cloud-native object storage.

Consolidate and manage massive datasets with ease using high-throughput, scalable storage architecture for mass capacity. Built for rapid throughput, **Seagate** storage arrays affordably scale out petabytes of storage to meet data-intensive HPC workloads, while maximising capacity over time.

SoftIron's HyperDrive Performance+ leverages the AMD EPYC 3000 processor, the Performance+ family debuts the firm's first x86 CPU and NVMe SSD-based designs in its HyperDrive series.

Spectra Logic deliver disk, object storage, tape and data management software for archive, backup and cloud. Their storage solutions give cost-effective storage to meet performance, growth and environmental needs.

Suse offers an adaptable Linux operating system and the only open Kubernetes management platform.

Weka offers a modern storage architecture that can handle demanding I/O intensive workloads and latency-sensitive applications at Exabyte scale to help firms improve business outcomes wherever their data resides – on-premises, the cloud or across both.

Tech focus: cooling

A ROUND-UP OF THE LATEST COOLING TECHNOLOGIES FOR SCIENTISTS USING HPC TO SUPPORT THEIR RESEARCH

As computing challenges increase, so too must the size of high performance computing systems, and this drives constraints to power usage and heat generation. Maintaining power efficiency requires innovative data centre and cooling design which can support these large HPC systems without compromising performance.

For computing applications such as complex weather models, simulations in automotive or mechanical engineering, animations in films, research analysis, or AI-based high-frequency trading of securities, HPC users require high performance cooling which can support their computing requirements.

In addition, companies are also increasingly using artificial intelligence applications to minimise the maintenance time of machines via machine learning in production environments.

Several different cooling technologies are available to HPC users, and these vary in performance in the amount of heat they can dissipate, cost and ease of maintenance, or upkeep costs. Rear Door heat exchangers (RDHX) rely on passive or active cooling in the server door, which can help to reduce cabinet temperatures and better control the overall temperature of the server rack.

Liquid cooling is in theory very similar to using air but instead liquid is pumped in a closed system. Water has a better thermal conductivity, so this is potentially a higher performance system but requires additional infrastructure. Water-cooled data centres may also require additional costs from raised flooring or other infrastructure to support the water pipes and other equipment. Immersion cooling solutions use a non-conductive liquid that can sit directly alongside components. As the liquid heats up, it circulates and moves away from the hot components, creating a flow that keeps the component cool.

With the global race to deliver the HPC industry's next supercomputing milestone gathering pace, there is possibly little surprise that innovation within the cooling market for HPCs/data centres has likewise stepped up a gear, as the efficiency and performance of cooling technology becomes ever more significant in preventing the operational costs of HPC resources from reaching prohibitive levels.

Further information

By transitioning data centres from traditional cooling

Sponsored by

CoolIT systems™

methods to immersion cooling with **3M** fluids, businesses can better prepare for the unprecedented performance requirements of the future, while managing costs and the impact on our natural resources.

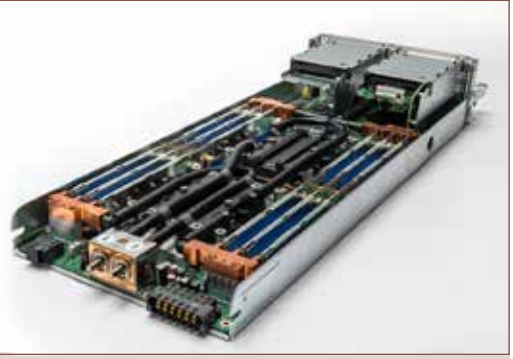
Aqua Cooling aims to build a life-long relationship with its customers, becoming essentially a 'one-stop shop' for all their temperature control requirements.

For HPC cooling, this heat needs to be extracted and removed as quickly and efficiently as possible. The simple tried and tested – and the most energy efficient and greenest – way to achieve HPC cooling is to use a Rear Door cooler as used in many of its data centre cooling solutions.

CPUs and server systems are running hotter than ever. Many industry experts are challenged with the problem of dwindling floor space, no electrical reserves and a need for more server racks to support the expanding user and applications base. The industry is looking for solutions, and **Aquila** can provide liquid cooling.

Asperitas provides cutting-edge immersion cooling for data centres globally and

CoolIT Systems FEATURED PRODUCT



Reinvent your data centre capabilities with Direct Liquid Cooling. Direct Liquid Cooling (DLC) uses the exceptional thermal conductivity of liquid to provide dense, concentrated cooling to targeted areas, resulting in optimised system downtimes and reduced component failures.

By using DLC and warm water, the dependence on fans and expensive air handling systems is drastically reduced, lowering OPEX. CoolIT products are equipped to support any server, rack configuration and data centre environment to enable server technology to perform at its peak.

With CoolIT's modular approach, customers work directly with CoolIT's engineering experts to select components specific to their needs. It's as easy as 1, 2, 3. Our team of industry experts can help identify the ideal custom solution and technology for your specific computing desires. Contact us today to get started with the future of data centre technology.

https://www.coolitsystems.com/

Nortek Data Center Cooling FEATURED PRODUCT

Data centre professionals looking to build the next big thing, need a cooling system that can keep up. Nortek Data Center Cooling StatePoint Indirect Cooling Technology brings reliability and optimum performance to any data centre, regardless of location or climate.

This revolutionary technology delivers a solution that is better for business and the environment. StatePoint uses an indirect evaporative cooling method with a unique semi-permeable membrane exchanger that completely separates water from the air stream to prevent cross contamination. The result – significant power and water savings (as much as 30 per cent) and minimised risk for biological growths, such as Legionella.

StatePoint easily integrates with a variety of systems to offer a complete cooling solution. The unit can use recycled water, interface with waste heat recovery systems and operate off wind and solar power to help data centres take the next step toward net-zero operation.
Learn more about StatePoint:
https://www.nortekdatacenter.com/statepoint/

CPC's Everis

FEATURED PRODUCT

Getting top liquid cooling performance with CPC's Everis purpose-built quick disconnects

As heat generated by HPC increases, use of liquid cooling becomes necessary to dissipate the heat and help ensure system performance. Use of quick disconnects (QDs) or quick-release couplings that are specifically designed and manufactured for liquid cooling of electronics applications, is the way to achieve cooling targets and reliable optimal ongoing operation. CPC's Everis line of QDs are built for liquid cooling.

All Everis QDs are non-spill and they feature superior flow to size ratios and redundant multi-lobed seals. Everis quick connects are available in a wide array of styles, materials, terminations, and sizes to match cooling load demands, coolant selection and system configuration. Additionally, CPC liquid cooling application engineers are available to answer questions about the wetted loop and QD impact to flow and pressure. Ultimately, CPC understands the holistic nature of liquid cooling. Contact us.

Request product samples for your next project
https://products.cpcworldwide.com/en_US/ProductsCat/LiquidCooling/

Download Chemical Compatibility Tech Guide
https://www.cpcworldwide.com/Thermal-Campaigns/Liquid-Cooling-and-Chemical-Compatibility

sustainably, specifically with their AIC24 solution. The IT equipment within the module is immersed in a synthetic, single-phase fluid developed by research and development partner Shell. Key benefits include efficiency, no water consumption, free cooling and heat reuse.

Calyos are leaders in the design, development and manufacturing of passive loop heat pipes. Their solutions enable the freedom to design without thermal limitations, unlocking the best performance from components and enabling products to be the most competitive.

ColdLogik's CL20 Proactive Rear Door Coolers replace the traditional approach to data centre cooling, allowing loads of up to 92kW per cabinet, with the added benefit of removing real estate inherent with hot aisle/cold aisle, in row cooling, CRAC cooling and aisle containment designs. Achieving all of this, while also enabling remote monitoring and full control of the data centre environment.

CoolIT Systems is a liquid cooling solution provider for HPC, cloud and enterprise data centre markets worldwide. The experienced innovator has more than 60 patents, and more than four million liquid-cooling units deployed around the globe.

Ecocooling's evaporative cooling and ventilation systems are designed to be used across the UK, Europe and internationally. An EcoCooling direct evaporative cooling system can reduce your IT or data centre cooling costs by up to 90 per cent, comply with ASHRAE 9.9 and produce a PUE (Power Usage Effectiveness) of under 1.1.

Green Revolution Cooling's (GRC) patented immersion cooling technology helps create favourable budget economics by reducing server energy a full 10 to 20 per cent. GRC's immersion cooling data centre solutions can future-proof your operation by reliably cooling up to 100kW/rack. The support for high rack densities allows you to put more compute in a rack, while the absence of hot/cold aisles helps you put racks back-to-back, saving valuable space in data centres.

Iceotope helps users distribute AI and HPC workloads with confidence to any location. Providing a highly resilient and quietly efficient micro data centre solution for deploying accelerated AI/HPC workloads to any edge location – no matter how extreme.

Motivair provides cooling solutions for data centres that are faced with ever-changing thermal challenges as big data, artificial intelligence, robotics and the exascale era continue to evolve.

HPC applications have become a driver of innovation in many industries, so companies have to deal with the energy-efficient expansion of their data centres.

nVent provides solutions that support the highest performance classes and deliver the very high energy efficiency.

Submer are specialists in liquid submersion cooling, a routine method of cooling large-power distribution components such as transformers. Still rarely used for the cooling of IT hardware, this method is slowly becoming popular with innovative data centres worldwide.

The **TMGcore** OTTO Platform has been designed from the ground up to provide ground-breaking improvements in equipment densities, ease of use, reliability and resiliency. Offering substantial savings in both CapEx and OpEx, OTTO provides tools to simplify, centralise and automate the administration of all aspects of the platform.

Asperitas

FEATURED PRODUCT

Dutch clean and high-tech scale-up Asperitas provides cutting-edge immersion cooling for data centres globally, specifically with their AIC24 solution. The IT equipment within the module is immersed in a synthetic, single-phase fluid developed by research and development partner Shell. Key benefits include efficiency, no water consumption, free cooling and heat reuse. The innovative company developed and marketed the Immersed Computing concept, a technology based on immersion liquid cooling to enable data centres to function in a sustainable manner, even for the most demanding users and environments.

In 2018, Asperitas won the international New Energy Challenge competition organised by Shell, Rockstart and YES!Delft. In May 2020, the World Economic Forum published a paper on Transformational Energy Innovations of the last 10 years, selecting Asperitas technology as the only data centre energy efficiency solution in their ground-breaking paper.

To book your online demo, please contact info@asperitas.com or visit www.asperitas.com

Evolving AI

WITH THE GROWTH OF AI AND DL COMES NEW OPPORTUNITIES FOR EMERGING APPLICATIONS, FINDS ROBERT ROE

As artificial intelligence (AI) and deep learning (DL) technologies mature, there are increasing numbers of applications available to scientists and researchers who are adopting these methodologies to increase research output.

In addition to emerging applications in AI, the accelerator technologies developed for AI or machine learning are now finding new applications in more traditional HPC and scientific computing use cases. Nvidia has recently announced a collaboration with biopharmaceutical company AstraZeneca and the University of Florida's academic health centre, UF Health, on new AI research projects using transformer neural networks.

Transformer-based neural network architectures – which have become available only in the last several years – allow researchers to leverage massive datasets using self-supervised training methods, avoiding the need for manually labeled examples during pre-training. These models, equally adept at learning the syntactic rules to describe chemistry as they are at learning the grammar of languages, are finding applications across research domains and modalities.

Nvidia is collaborating with AstraZeneca on a transformer-based generative AI model for chemical structures used in drug discovery that will be among the very first projects to run on Cambridge-1, which is soon to go online as the UK's largest supercomputer.

The model will be open-sourced, available to researchers and developers in the Nvidia NGC software catalogue, and deployable in the Nvidia Clara Discovery platform for computational drug discovery.

Separately, UF Health is harnessing Nvidia's state-of-the-art Megatron framework and BioMegatron pre-trained model – available on NGC – to develop GatorTron, the largest clinical language model to date.

New NGC applications include AtacWorks, a deep learning model that identifies accessible regions of DNA, and MELD, a tool for inferring the structure of biomolecules from sparse, ambiguous or noisy data.

This is just one example that highlights the success of Nvidia's drive to capture the AI and DL markets. So far they have been incredibly successful but there is mounting pressure from other accelerator technology providers. One such example is Graphcore, the UK-based company developing its own brand of general purpose accelerators known as intelligence processing units (IPU).

Graphcore released the second generation of IPU products in 2020 and is quickly gaining pace, with exciting benchmarks in both ML and scientific computing.

There are several examples on the Graphcore website, for example in the areas of drug discovery and life sciences where the IPU has already been deployed for several different applications.

In the example of BERT-BASE training, the IPU achieved 25 per cent faster training time at 20 per cent lower power, meaning the algorithm will run faster at a lower cost. BERT-BASE inference training against a V100 GPU showed that IPU provides 2x higher throughput, making it possible to use BERT in an interactive setting where scalability is a priority. There are also several examples with EfficientNet-B0 and Markov Chain Monte Carlo (MCMC) simulations.

Matt Fyles, SVP Software at Graphcore, explains that while the first generation proved the use case for the technology and got it into the hands of developers, →

"We have built the software and hardware together in a way that allows us to very quickly realise performance"

→ the second generation provides a boost in performance and scalability. 'The first generation of IPU products was really to prove the IPU as a computing technology that could be applied to problems in the machine learning (ML) space primarily, but also as a general programming platform,' said Fyles.

'We started out with a view that we are making a general purpose compute accelerator. Our initial focus is on ML and that is where a lot of the new capabilities that we have added can be deployed. But IPU is also equally applicable to problems in scientific computing and HPC. And while that was secondary to our plans to begin with, it was more about building out the ecosystem to support both aspects,' states Fyles.

Fyles notes that getting IPUs into datacentres and the hands of scientists as being the primary goal for the first generation. This then allows Graphcore to begin engaging and developing an ecosystem of developers and applications.

'A lot of it was about bringing IPU to the world and getting people to understand how it worked and how we could move forward with it,' staytes Fyles.

The second generation of IPU technology drives performance and allows users to scale to much higher levels than was previously possible. Scalability is something that Graphcore have spent a lot of time on to ensure that both software and hardware can make use of the additional resources.

'We have now got this scalable system that is built around our m2000 platform, which is a 1U server blade containing four IPUs supported by the Poplar software stack. We can build a POD16, as we call it, which is four of our M2000 connected together – but we can also scale it up a lot further to have 64IPUs or 128 IPUs and allow the software stack to programme it in a similar way,' added Fyles. But driving adoption is not just about hardware. Fyles stressed that the software and hardware have been designed from the ground up to ensure ease of use, scalability and performance: 'The common theme is the software development kit and all the great work that the team has done on it to deliver a mature product has continued across both generations and will continue in the next generation that follows.'

> **"People are starting to solve challenges and use IPU themselves. We see a lot of great feedback which helps to push the software forward"**

'We had the opportunity to redesign the software stack alongside the hardware from the beginning. To augment and try to steer the hardware and software from where it came about out of the HPC space,' added Fyles. 'We have had the opportunity for a clean slate approach to both software and hardware that has meant that we can potentially find performance where others cannot and we have a much simpler legacy software to maintain over time. We have built the software and hardware together in a way that allows us to very quickly realise performance.'

Fyles stressed that huge amounts of time and resources have been spent on the Graphcore developer portal and populating that with examples, videos and worked examples of how to use and get the most out of IPU products.

'We are trying to build an ecosystem of IPU developers that starts small and grows over time. We do that in a number of ways, one is relatively organically, we have interesting people come to us with interesting problems for the IPU to solve and that is more of a traditional sales engagement,' states Fyles. 'Then there is the academic programme that we have introduced recently which starts to work with key academic institutes around the world. The IPU is a platform for research and development that is different from something that they have now but it is also easy to use, well supported and documented. People can take applications now and they can do things without asking for our help. There are a number of great projects I have seen, for example, the University of Bristol.'

The University of Bristol's High-Performance Computing group, led by Professor Simon McIntosh-Smith, have been investigating how Graphcore's IPUs can be used at the convergence of AI and HPC compute for scientific applications in computational fluid dynamics, electromechanics and particle simulations.

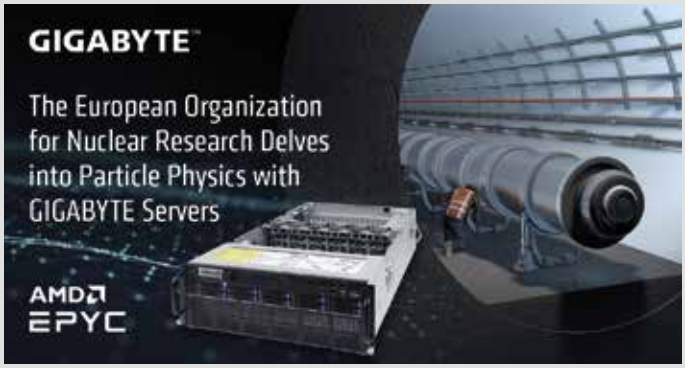
'Traditionally that seems like something that we would have had to give a lot of help with and that is the good thing. People are starting to solve challenges and use IPU themselves. We see a lot of great feedback which helps to push the software forward,' explains Fyles.

There are many different technologies available to AI and ML researchers including GPUs, IPUs and FPGAs to name a few. However, Graphcore's Fyles believes that only technologies that have been designed around ease of use for software engineers can succeed. As such, technologies such as FPGAs will most likely never see wide adoption in the AI and ML space as they require skills that are not suited to HPC or AI.

'People have spent a lot of time trying to turn it [FPGAs] into a platform that software engineers can use but the best platform for a software engineer is a processor that uses a standard software stack with software libraries that they are familiar with,' states Fyles.

'You can abstract a lot of it away but that is why it is hard to get wide adoption on such a platform because the people using the product are software engineers. Now it is all about Python and high level libraries and ultimately people do not always care about the hardware underneath, they want their application to run and they want it to go fast,' added Fyles. 'There is still the low level, close to bare metal developers that have always existed in HPC but the wider audience is using Python and high level machine learning frameworks. That is where a lot of the new developers that come out of university have been trained. They do not expect to go down to the hardware level.'

Case study: the European Organisation for Nuclear Research delves into particle physics with Gigabyte servers



The European Organisation for Nuclear Research (CERN) uses Gigabyte's high-density GPU Servers with 2nd Gen AMD EPYC processors. Their purpose: to crunch the massive amount of data produced by subatomic particle experiments conducted with the Large Hadron Collider (LHC).

The impressive processing power of the GPU Servers' multi-core design has propelled the study of high energy physics to new heights.

The biggest and most energy-intensive CERN project is its particle accelerator: the Large Hadron Collider (LHC). To detect the subatomic particle known as the beauty (or bottom) quark more quickly, CERN has decided to invest in additional computing equipment to analyse the massive quantities of raw data produced by the LHC.

When CERN looked for ways to expand their data processing equipment, the top priority was to acquire HPC capabilities. They specifically wanted servers equipped with 2nd Gen AMD EPYC processors and multiple graphics accelerators. The servers should also support PCIe Gen 4.0 computing cards. Gigabyte was the only company with the solution to match their demand. CERN selected the Gigabyte G482-s51, a model that supports up to eight PCIe Gen 4.0 GPGPU cards in a 4U chassis.

Gigabyte has plenty of experience in HPC applications. When AMD began offering PCIe Gen 4.0 technology on their x86 platforms, Gigabyte immediately responded by leveraging technological know-how and expertise to design the first GPU Servers capable of supporting PCIe Gen 4.0. Gigabyte optimised the server's integrated hardware design, from the electronic components and PCB, all the way down to the high-performance power delivery system.

Signal integrity was maximised by minimising signal loss in high-speed transmissions between CPU and GPU, and between GPUs. This resulted in a GPU server that features slower latency, higher bandwidth and unsurpassed reliability.

> "Gigabyte has plenty of experience in HPC applications"

To effectively process huge quantities of data with HPC technology, more than the CPU and GPU's combined computing power comes into play. High-speed transmission is crucial; whether it relates to the computing and storage of data between multiple server clusters, or the accelerated processing and communication of data between devices linked through the internet. The Gigabyte G482-s51 overcome this challenge with its PCIe Gen 4.0 interface, which supports high performance network cards. The increased bandwidth enables high-speed data transmission, which in turn enhances the performance of the entire HPC system, making it possible to process the 40 terabytes of raw data that the particle accelerator generates every second.

Custom-designed servers provide CERN with cutting-edge computing power

To quickly analyse the massive quantities of data generated by experiments conducted with the LHC, CERN independently developed its own powerful computer cards to handle all the calculations. They paired these cards with graphics cards designed for image processing. The combined might of these specialised tools became the last word in cutting-edge computing.

Gigabyte customised the G482-s51 to meet the client's specific requirements, including specially-designed expansion slots and minute adjustments to the BIOS, and an advanced heat dissipation solution.

Specially designed expansion slots and minute adjustments to the BIOS: To account for CERN's self-developed computer cards, Gigabyte customised the expansion slots of the G482-s51. Gigabyte also ran data simulations and made minute adjustments to the BIOS to better link all the computer cards to the motherboard, so that each card could achieve the maximum PCIe Gen 4.0 speed.

Advanced heat dissipation solution: CERN had special specifications for just about everything, from the power supply to the network interface cards to the arrangement of the eight GPGPU cards. The heat consumption of these interwoven I/O devices is not all the same. Gigabyte leveraged its expertise in heat dissipation designs and integration techniques to successfully channel airflow inside the servers, so excessive heat would not be a problem.

Gigabyte worked closely with AMD to expand the horizon for HPC applications. One of the most noteworthy advantages of the AMD CPU is its multi-core design. In the push to solidify the AMD EPYC processor's position in the server market, Gigabyte has an important part to play. By creating an AMD EPYC Server that showcases top performance, system stability and steadfast quality, Gigabyte was able to satisfy CERN's need for a solution capable of analysing large amounts of data and completing HPC workloads.

Gigabyte's responsive customisation services and in-depth experience in research and development were just what the client needed to meet their specific requirements. By pushing computing power to the limit with state-of-the-art technological prowess, Gigabyte has taken an impressive stride forward in the application of HPC solutions to academic research and scientific discovery.

To learn more about Gigabyte server solutions, please visit *https://www.Gigabyte.com/Enterprise*
For direct contact, email *server.grp@Gigabyte.com*

FREE ON-DEMAND WEBCAST*
AVAILABLE NOW

Enabling AI in Pharma with FAIR Data

SPEAKERS

Dr. Haydn Boehm,
Head of Commercial Marketing,
Connected Lab, Merck Group

Haydn Boehm is the head of commercial marketing for MilliporeSigma's Connected Lab digital program and has over 20 years' experience in marketing and business development within the life scientific sector. Haydn has a PhD in organic chemistry from the University of Nottingham, UK

Artificial intelligence (AI) has long been heralded as a game changer for the pharmaceutical industry, so why are there so few drugs on the market today that resulted from AI?

In this webinar, we'll explore the logistical barriers that are impeding data-driven drug discovery programs, and discuss how open standards and the guiding principles of FAIR (findable, accessible, interoperable, and reusable) data management can accelerate speed to market and enable enterprise-wise digital transformation. In addition, we will consider pragmatic steps your lab can take today to harmonise your data inputs and automate data capture.

SCIENTIFIC COMPUTING WORLD

MERCK

www.scientific-computing.com/webcasts

*This webcast is FREE to view, though registration is required

Intelligent integration

SOPHIA KTORI CONTINUES THE DISCUSSION ON SOFTWARE INTEGRATION IN THE LABORATORY

One of the most obvious, and immediate challenges when bringing new software into a lab environment is the likely 'spaghetti soup' of existing platforms – possibly from multiple vendors – that are already installed, suggests Richard Milne, vice president and general manager, Digital Science, at Thermo Fisher Scientific. 'Each of these will offer different levels of integration'.

The situation is compounded because even in the same organisation, there may be different suites of instruments and software in separate labs and across departments. While there is an ambition to integrate instrumentation and software tools across a business and geographic sites, the reality may be what Milne describes as an 'unstructured legacy of decisions'. Each of which represented a theoretically attractive investment at the time, but which in practice offered a point solution that ultimately 'confuses' the whole environment.

This means many organisations will have some level of legacy investment, in instrumentation, in SOPs, in working practices, and in pieces of software that people use every day. It's likely that most integration projects will be very much in a 'brownfield' setting, rather than setting up 'greenfield' labs, Milne said. And that brownfield environment will almost undoubtedly be very fragmented in terms of its legacy systems. So, connectivity needs to happen at the enterprise level, not just at the lab level, and encompass that existing ecosystem of digital technologies, Milne believes.

'The overarching aim is to generate an environment that allows you to unravel that spaghetti soup of existing platforms, and make sure that there's coherent organisation, and the ability to use it all,' Milne said. 'Ultimately, this will allow organisations to purchase tools based on the capabilities and features of those tools, rather than on whether they will talk to the lab's existing equipment.' And this means that wherever R&D teams, service providers or partners are located and whatever technology they use, they should be able to collaborate and share data →

"The overarching aim is to generate an environment that allows you to unravel that spaghetti soup of existing platforms, and make sure that there's coherent organisation, and the ability to use it all"

→ across a cloud-based environment.

'As well as solving the integration challenge for individual scientists, we want to make sure it can also be achieved at scale,' he continued. 'We are trying to square the circle a little bit, by creating a platform environment' – and this will be a cloud-based environment, Milne noted – 'that will give people the freedom to make those decisions at the level of their workgroup, lab or building, but expand the value of that integration at scale.'

This also means that the concept of integration doesn't stop at the level of lab instrumentation and software, he believes. 'Lab function will increasingly be married to asset performance management, inventory or resource management, and quality management. Again, businesses will have many options to choose from, and then the issue becomes, for example, how do I integrate my resource management tool into my existing laboratory information management system (LIMS).'

Leveraging advanced tools

The concept of holistic lab orchestration hinges on addressing three basic problems, says Trish Meek, director of marketing at Thermo Fisher Scientific. 'Firstly, getting all of the data together so that you can use it for analysis, visualisation, and increasingly, for leveraging AI and advanced machine learning tools. Then there's the human experience in the lab. How do you optimise the scientific experience for scientists day-to-day. Third is the need to improve and facilitate process optimisation.'

Instrument and software vendors are already making moves to facilitate easier integration, Milne acknowledged. 'From a lab connectivity perspective, then when we look at instrumentation and software used for everyday lab work, such as sequencers, qPCR, flow cytometry, etc., the vendors of these types of instrumentation are already starting to think about connectivity when they develop their new instruments.' Thermo Fisher Scientific, for example, is building connectivity into all of its new instrumentation, he continued. 'But then there will also need to be some sort of 'retrofittability,' and that will be part of our initial offering. This will be achieved through the creation of a gateway that will make it possible to connect instruments and software in the lab, into a cohesive environment.'

While there are possibly multiple aspects to the issue of achieving seamless connectivity in the lab, the ultimate aim is to make laboratory systems more effective at what they do, every day, Dave Dorsett, principal software architect at information technology consultancy Astrix Technology, suggests. 'That's a foundational concept; how to improve usage of systems – such as a LIMS or ELN platform – from the perspective of everyday use, and how to get these systems to work together to support the labs on a day-to-day basis.'

Consider the software and hardware tools that a lab ecosystem relies on, and much of the interruption in integration will commonly be due to the diverse nature of instrument architecture, Dorsett noted. An organisation may have LIMS systems from multiple vendors in use across different departments, for example, he said, mirroring Milne's sentiments. 'Some of these systems, whether LIMS platforms or other hardware or software, are more challenging to integrate than others. And this makes it costly for individual companies to set up and maintain them from an integration perspective.'

What this means at the most basic level, is that many labs may still rely on manual data transcription or 'scientist-facilitated integration', Dorsett continued. '"Sneakernet" [physically transferring data from one PC to another using portable drives and devices] remains just part of everyday lab life. And no matter how careful you are with manual transcription and data input, or how effective your data review processes, the ultimate quality of that data is always going to be at risk.'

There are two challenges, in fact, Dorsett suggested. Sometimes the issues are not so much with getting systems to talk to each other, as they are with aligning and harmonising the data that comes out: getting data out of point systems and enabling the flow to the next stage represents another stumbling block to seamless lab integration, Dorsett suggested. If it's hard to get data from a LIMS, ELN or other key piece of software back out as accessible and meaningful, then it may not be possible to use that tool or platform to maximum effectiveness and efficiency.

Dorsett continued: 'One approach to addressing such issues is to bring data from multiple systems into data lakes,

where it can feasibly be compared, but again, you have to ensure that your data are equivalent, particularly where your labs may be running multiple LIMS or ELNs, for example. You may have one LIMS for stability testing, and another for batch release, plus method data in an ELN.'

A typical problem organisations face is how to compare all of that data once you have technically integrated your systems. 'For any laboratory organisation, one of the biggest challenges to using the systems that they want to integrate, is how to ensure both data quality and data comparability/equivalence across systems, even once they are interconnected. Are your experimental methods equivalent, for example, or does a sample ID from your LIMS match a sample ID from a CRO?' Dorsett said.

It's important to try to understand what tools are used at the level of the lab, facility and enterprise, as the basis for working out how to maximise use of that collective investment, identify key gaps, and define a longer-term roadmap that recognises the importance of sustainability and total cost of ownership. 'You want to try to find ways of using integration technologies that are already there more effectively, as well as to be able to bring in new technologies,' Dorsett said. For instrument integration, there are middleware companies who are positioned to offer specific software to facilitate instrument integration, Dorsett suggested, citing SmartLine Data Cockpit, TetraScience and BioBright – the latter having been acquired by Dotmatics in 2019. 'These companies are focused on providing tools that can address how people gather all their data from the different instrumentation,' said Dorsett.

Consolidating data

Any rounded conversation on LIMS/ELN and software integration and management will at some point likely come around to the concept of data lakes, noted Robert D Brown, vice president, product marketing at Dotmatics. 'The initial concept was that data lakes could house all the lab-derived data, and that people could then dip in and retrieve what they needed, when they needed to.' But in a real-world setting there are two types of data, he explained. 'You have the structured data, such as data in your ELN, but then you will also have the vast numbers of unstructured data files that are being generated by all of this automated instrumentation that labs now use. Typically these files may be output in proprietary, non-standardised formats, and the data contained will first need to be parsed out before being put into a LIMS, or an ELN.'

But with the right tools in place, we can now have the best of both worlds. 'We can have hybrid systems where the unstructured data lives in the lake, and the structured part of that data can then go into the ELN. As long as both sides have a good API, and you have a way of parsing the data, then its possible to overcome most technical hurdles. The trick is to link the two types of data appropriately.'

> "As well as solving the integration challenge for individual scientists, we want to make sure it can also be achieved at scale"

From a software perspective, the ability to work with both biologics and small molecules using the same overall platform is founded on the use of software components that can be slotted together in multiple ways to establish the right workflow for the right outcome. A chemist may do things in one order, but a molecular biologist might do them in another. 'That's the real trick,' he continued, 'to be able to put the different pieces of the same overall solution together so that they match the workflow for the different scientists.'

And the next stage in software evolution will at least in part – and perhaps inevitably, Brown noted – focus on integrating AI into the everyday lab function. 'First, you have to add AI and machine learning into your software stream,' and this is more procedural, he indicated. 'But critically – and this is perhaps the biggest problem – it's imperative that you are getting absolutely clean data into those ML models in the first place. If you don't put clean data in, you will get garbage out.'

This brings us back to the concept of data automation, so that you don't have to use humans to move data around, which will at some stage run the risk of human errors in data manipulation, Brown said. Automating data generation, management and transfer will also facilitate that ability to pull complete datasets across from any system, into the ML learning pool.

'And here is where we have the advantage of the BioBright lab automation solution, which automates the complete process of getting data off instruments into the lake, parsing it, and putting it into the notebook. Compare this with the requirement for human transfer of files between systems, and the manual inputting of data, and sharing spreadsheets, which is inherently error prone.'

With this goal of complete round-trip automation in mind, Dotmatics announced a partnership with HighRes Biosolutions, which designs and builds robotic systems and laboratory devices, in January. The collaboration focuses on marrying high-throughput laboratory automation capability of the Dotmatics ELN with the HighRes instrument control software Cellario. This combination frees scientists to plan experiments, run individual instruments and publish and analyse data in a single software interface.

UK drives genomics research

ROBERT ROE TAKES A LOOK AT GENOMICS RESEARCH UNCOVERING INSIGHTS INTO COVID-19, AND PAVING THE WAY FOR BETTER CANCER TREATMENTS

Patients across the UK will benefit from better healthcare, treatments and faster diagnosis as the government sets out how it will continue to deliver world-leading genomic healthcare.

Genomics is the study of genetic information and can help diagnose diseases earlier and more accurately, reduce some invasive procedures and enable tailored treatments. Building on the success of the 100,000 Genomes Project, the UK government has committed to sequence one million whole genomes – 500,000 genomes in the NHS and 500,000 in UK Biobank – which will transform healthcare in the UK and create jobs.

In addition, genomics has also been used to better understand Covid-19 and the variants that have increasingly become one of the biggest concerns of the pandemic.

Each variant is made up of a collection of mutations. The majority of mutations don't change how the virus behaves. However, some mutations can change the properties of the virus, and potentially give rise to a new variant. Many of these 'mutations of interest' occur in the spike protein, which is what gives the virus its ability to target, latch onto and enter the cells that it infects.

Working with key partners across the genomics community, the bold new Genome UK implementation plan 2021 to 2022, published in May, sets out 27 commitments to deliver over the next year, including five high-priority actions: faster diagnosis; whole genome sequencing for patients with rare diseases; engagement closely with different communities to ensure diverse datasets; recruitment of up to five million people representative of the UK population; to develop global standards and policies for sharing genomic and related health data.

Faster diagnosis and treatment of cancer using genomics through a partnership between Genomics England and NHS England will help researchers and healthcare professionals identify technologies that could be used to provide faster and more comprehensive genomic testing for cancer.

Whole genome sequencing for patients with rare diseases and cancer, as part of the NHS Genomic Medicine Service, will build on the success of the 100,000 Genomes Project, and aims to increase the amount of genomic data available to researchers.

The drive for larger and more diverse datasets from different communities aims to ensure that everyone across the UK can benefit from genomic healthcare and genomic databases that are representative of such a diverse population. This is essential for equitable access to new techniques, such as polygenic risk scores (PRS), which compares a person's risk to others with a different genetic makeup, and pharmacogenomics, which examines the role of the genome in the body's response to drugs.

Developing global standards and policies for sharing genomic and related health data ensures accurate and quick sharing of research data, which will help to benefit the entire genomics community.

The National Institute for Health Research, Medical Research Council and Wellcome Trust will, over the next five years, provide £4.5m of funding to the Global Alliance for Genomics and Health, ensuring standards are easily accessible and usable by global genomic

programmes and data-sharing initiatives, placing the UK at the forefront of secure sharing of international genomic and health-related data.

Matt Hancock, the UK's Health Secretary, said: 'We will transform the UK into a life sciences superpower. We'll build on the success story of our life sciences during the pandemic, which has led the world in everything from vaccine development, to finding effective treatments that work, to genomic sequencing.

'Today we've published our Genome UK implementation plan for how we can build on this even further, including our commitment to sequence one million whole genomes. Because genomics saves lives, and I'm determined the UK stays at the forefront of this vital new technology,' Hancock continued. 'If we draw on ingenuity like this, we can keep up the fight against Covid-19, and also tackle the other things that stop us living healthier lives like cancer, dementia and heart disease.

'So, we're increasing UK investment in research and development, bringing much more of the supply chain onshore, sparing no effort to attract the brightest innovators and the best manufacturers,' he concluded.

Minister for Innovation Lord Bethell said: 'The UK has a proud history in developing genetic and genomic technologies which improve the lives of patients across the country and globally.

'This implementation plan demonstrates the great strides we have already made since the launch of Genome UK, and outlines the actions we are taking to progress key commitments over the next year.

'It is vital that we continue to maintain and develop our global leadership in this field, to realise the full potential offered by genomics,' Lord Bethell added.

> "Genomics saves lives, and I'm determined the UK stays at the forefront of this vital new technology"

This first phase implementation plan follows on from *Genome UK: the Future of Healthcare* published in 2020, which set out a vision to create the most advanced genomic healthcare system in the world, to deliver better healthcare at lower cost.

Genomics is just one example of the government's commitment to driving forward health innovation in the UK, which will be central to the future health resilience, the growth of the UK's life sciences sector and improving patient care.

Chris Wigley, Genomics England CEO, said: 'Since the days of Darwin, Rosalind Franklin, Crick and Watson, and Fred Sanger, the UK has been at the forefront of genomic science. With this publication it's exciting to see the next chapter of that story coming to life. Our ecosystem has come together as never before through the difficult times of the pandemic – and this implementation plan will allow us to build on this collaboration between all of the world-leading genomics institutions in the UK.'

Professor Dame Sue Hill, NHS England's Chief Scientific Officer, said: 'The NHS is already a global leader in genomics and has introduced a range of new cutting-edge tests for people with rare diseases and cancer over the last year, despite the pandemic.

'Genomics can truly transform the way patient care is delivered, helping to predict and prevent disease, personalise treatments and ultimately save lives.'

Understanding Covid-19

In February 2021, the UK's Covid-19 Genomics UK Consortium (COG UK) launched the COG-UK Mutation Explorer (COG-UK-ME) – an interface that provides access to data on Sars-CoV-2 mutations and variants of interest in the COG-UK genome sequence dataset. COG-UK-ME allows anyone to view information about important changes in the Sars-CoV-2 genome over time.

The tool is updated twice weekly, and largely focuses on spike gene mutations →

→ of potential or known importance; providing information on cumulative frequency and data for the last 28 days, to give an approximate assessment of recent changes.

COG-UK-ME draws UK genome data from the MRC-CLIMB database. This data visualisation tool allows anyone to follow information over time on important changes in the Sars-CoV-2 genome.

How does COG-UK-ME work?

Selecting the 'Mutational Explorer' tab takes you to three tables. Table 1 lists mutations in the spike gene that have led to an amino acid change (called a substitution, which is concentrated on because it may change the way that the virus interacts with humans).

Mutations are ranked by frequency in the MRC-CLIMB database (the most common mutations first). A search function allows individual mutations to be selected, and a file downloaded containing a list of COG-UK identifiers, dates and lineages. For example, selection of E484K provides links to information for each genome that carries this mutation, the date of the sample, and the lineage the isolate belongs to.

Data can also be visualised for each mutation in a graph by clicking the visualiser tab. This shows the number of times the selected mutation has been detected over time.

COG-UK-ME also displays mutations that could affect the way that the virus interacts with the human immune response based on laboratory studies (Antigenic information tab).

Scientific evidence is graded. 'High confidence' is applied when a mutation is found by multiple independent studies using multiple different approaches, including studies using polyclonal (convalescent or post-vaccine) antisera; 'medium confidence' means this has been found by multiple independent studies; and 'lower confidence' indicates this has been found by a single study only. Mutations with an antigenic role can also be filtered by domains of the spike protein.

The Explorer will be updated with new functions over time, based on scientific observations and ways of describing and thinking about variants. The current Covid-19 pandemic, caused by Sars-CoV-2, represents a major threat to health. The Covid-19 Genomics UK (COG-UK) consortium has been created to deliver large-scale and rapid whole-genome virus sequencing to local NHS centres and the UK government.

Led by Professor Sharon Peacock of Cambridge University, COG-UK is made up of an innovative partnership of NHS organisations, the four Public Health Agencies of the UK, the Wellcome Sanger Institute and 12 academic partners providing sequencing and analysis capacity. Professor Peacock is also on a part-time secondment to PHE as director of science, where she focuses on the development of pathogen sequencing through COG-UK.

COG-UK was established in April 2020 supported by £20m funding from the Covid-19 rapid-research-response 'fighting fund' from the UK government, and administered by the National Institute for Health Research, UK Research and Innovation and the Wellcome Sanger Institute.

The consortium was also backed by the Department of Health and Social Care's Testing Innovation Fund in November last year to facilitate the genome sequencing capacity needed to meet the increasing number of Covid-19 cases in the UK over the winter.

Subscribe for free*

SCIENTIFIC COMPUTING WORLD

Do you compute?

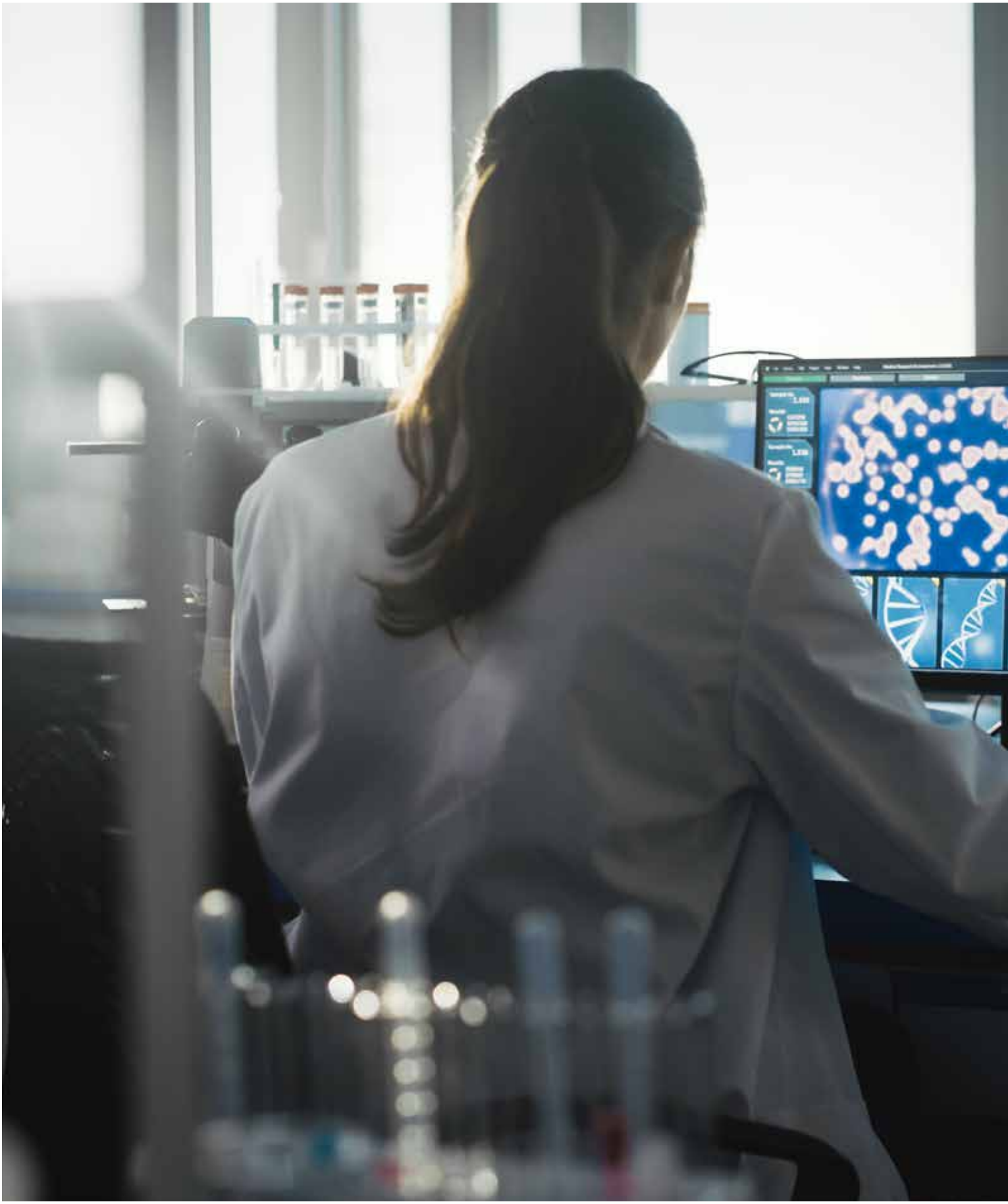
The only global publication for scientists and engineers using computing and software in their daily work

Do you subscribe? Register for free now!

scientific-computing.com/subscribe

*Registration required

Also published by Europa Science



Streamlining drug discovery

ROBERT ROE EXPLORES THE SERVICES AND SOFTWARE AVAILABLE TO SUPPORT ORGANISATIONS DRUG DISCOVERY RESEARCH

Drug discovery is a complex process that requires increasing resources and time to generate new molecules. As the processes have matured there is an increasing reliance on computational methods to accelerate the discovery process.

The use of computational chemistry software such as computer aided drug discovery (CADD), analytics, and large scale simulations are being increasingly deployed by drug discovery companies. This is also compounded by emerging technology such as artificial intelligence (AI) and machine learning (ML) methods which drive the requirements for computing infrastructure to support research.

These changing requirements force organisations to look at their infrastructure and staffing, as these advanced computing platforms require IT expertise alongside domain experts in a number of scientific disciplines. There are also knock-on effects from these changing requirements that are driving the creation of new software and services that can support these companies and allow them to focus on science rather than supporting software and developing IT infrastructure.

Michael Riener, president of RCH Solutions, explains that much has changed in the last 20 years of drug development: 'There have been a lot of changes but one thing stays the same - the constant advancing speed of science. That makes it difficult to keep pace and it makes it difficult for the people that support scientists.'

'It is a challenge but there are solutions that help you do that. 20 years ago, many companies would not have needed RCH. A scientist sitting in his lab had their whole environment to himself, operating in a silo with all the resources available to him from a scientific computing standpoint. As things evolved, the business changed due to cost challenges and the like,' added Rainer.

Although the exact timeline varies for each organisation 'at some point a decision was made' to start sharing those resources with other people. 'Inevitably service gets diluted and goes down, and all the good people leave,' exclaimed Rainer. 'Then it went a step further lets outsource compute and IT services. And while that can save money, the outcome is not always what you want.'

This shift in the way that resources are managed – largely to outsource IT as a cost saving measure – meant that many companies struggled to deliver the kind of computing environments that scientists require. Over time, this has led companies like RCH to deliver a bespoke research computing environment (RCE) that can →

> "There have been a lot of changes but one thing stays the same - the constant advancing speed of science"

→ support drug discovery and biotech organisations scientific computing needs.

'That has opened up opportunities for companies to come in and provide services that meet the demands of customers that traditional IT could not meet,' stated Rainer. 'IT has a different role that fits the rest of the business, which, for the most part, compared to science and research is fairly standard.'

Scientists might need new resources or to adopt a new technology quickly and that is not necessarily supported through a traditional IT model. 'We found a niche that has evolved to support that scientific computing realm. We fit between IT and we fit between science and we have experience and expertise on both sides,' said Rainer.

Phil Eschallier, chief technology officer at RCH Solutions, added: 'Life science companies whether it is pharma biotech etc are going to higher PhD scientists to do science and what we do is do something very well, which allows them to do what they are good at instead of forcing scientists to do both the science and the IT.'

Streamlining molecular discovery

Optimising software and the underlying technology can be a hugely important step in ensuring organisational efficiency. Tools such as computer aided drug discovery (CADD) can help organisations make better use of the structural knowledge of either the target (structure-based) or known ligands with bioactivity (ligand-based) molecules. This can be used to highlight potential drug candidates for further study.

Cresset has several tools available for small molecules drug discovery including protein-ligand analysis, molecule design, ligand-based virtual screening and many other tools. For example, the newly released 'Flare V5' builds on Cresset's established structure-based design platform - integrating ligand-based methods.

The latest version embeds functionality previously available in Cresset's Forge software, such as qualitative and quantitative SAR models, pharmacophore building capabilities and expanded QSAR functionality powered by the Flare Python API.

> "That has opened up opportunities for companies to come in and provide services that meet the demands of customers that traditional IT could not meet"

In a recent blog post, Martin Slater, director for Cresset Discovery Services, discussed the potential to outsource computational chemistry to complement internal research. 'Our CADD scientists apply the best ligand and structure-based solutions for each project, and supplement our own suite of software with select third-party tools.

'Cresset software centers around our proprietary XED force field to describe molecules as they behave in a biological context,' Slater continued. 'Working with Cresset's field technology gives a rich, informative view of each individual molecule that allows us to perform experiments such as scaffold hopping and fragment replacement.'

'We find that this view resonates with synthetic chemists who tend to think about molecules in terms of their electronic characteristics, such as electron-rich or electron-poor, when assembling them. The result is a method that is both cutting-edge but also intuitive to the scientists who will apply the results.'

Cresset has always provided consultancy alongside its software, but over the past few years there has been a steady growth in demand for consultancy services. Just as there is growing demand for hardware support for drug discovery, there is also a need to support for the software

'Maintaining an in-house team is a luxury, and outsourcing offers a way to benefit from the advantages that computational methods deliver without committing to a significant investment,' notes Slater.

Cresset has considerable experience in applying computational methods to any type of molecular discovery. Primarily this means pharma and biotech organisations, but the company also collaborate with teams from the agrochemical, and flavour and fragrances, industries.

'As the Cresset technology can work with or without the structure of a target protein, we are able to work on the widest range of target classes,' said Slater. 'Having an unknown target protein structure can simplify matters when engaging in a discovery project. For example, if we're trying to modify an active compound that is unusable, either due to off-target effects or patent conflicts, but keep the biology the same. We can characterise the molecule according to its field activity and look for compounds with new chemistry that have the same activity, which are often from a different structural class. If a company identifies a problem or bottleneck that they would like support with, we'll set up a free initial discussion with our modelling experts to evaluate whether it is a project we believe we can help with.'

'When a customer chooses to collaborate with Cresset, they get access to the entire discovery services team, not just an individual,' stressed Slater. 'Each project employs our expert modelers, application scientists and medicinal chemists to provide specific chemistry knowledge.'

Developing a research computing environment

As the complexity of these services continues to increase with drug discovery companies now regularly delving into the worlds of high performance computing, advanced analytics and AI and ML the requirements to support these systems continues to increase. This is leading more and more companies in biotech and drug discovery to adopt managed services which free the organisation to focus on its scientific output.

'Typically we would meet with a customer, they would outline the requirements for a project or a set of projects. We would deliver with them a proposal in the form of a statement of work which outlines their requirements, our deliverables, pricing and finite times scales,' said Rainer. 'The key is with a managed services model you effectively have one person who is the key point of contact. Behind the scenes there are a lot of very experienced individuals doing the work.'

'Scientists are paid to do science and research – they are not paid to manage scientific computing resources. We focus on what we do really well and that allows them to focus on what they do really well,' Rainer added.

Changing requirements

While there are similarities in the goals and challenges that these organisations face the technologies, software and implementation models they choose can be vastly different, which means that RCH is constantly adapting to meet new challenges and driver RCE's, which can support scientific innovation.

'The interesting thing with this job is there really isn't too much that is typical,' notes Eschallier.

'Everyday brings a new set of challenges and the type of people that we hire really embrace. They are not the kind of people that use standard operating procedures, that like to rinse and repeat problems.'

'We will get involved with anything based on: computing at scale, specalised applications, HPC, leveraging public cloud, curating data, analytics, mining, driving ML/DL,' Eschallier continued. 'Each customer has a similar problem, "I need to find a target, synthesise it and move it forward and look for things like toxicity and efficacy".

'But how they do it and the technologies they want to use and the processes they want to follow are all different,' he stressed.

'Sometimes they want a holistic, cradle to grave approach and other times customers will say I just need you to do these three niche items.

'It is about standing up software, making data available and performance tuning is a growing requirement,' Eschallier concluded.

> "Maintaining an in-house team is a luxury, and outsourcing offers a way to benefit from the advantages that computational methods deliver without committing to a significant investment"

Fixing a research computing environment

Rainer also gave a practical example based on the company CelGene which has now been acquired by Bristol Myers Squibb. The company had already developed its own computing environment for scientific research but found that it did not meet the requirements of its researchers upon completion.

'CelGene was growing rather quickly and they were in a state of change. They were acquiring a lot of companies and leadership decided that it was going too fast and they needed to determine some parameters and take control,' stated Rainer. 'There were many demands placed upon them from the business but the one that stood out was the need for a scientific computing environment.'

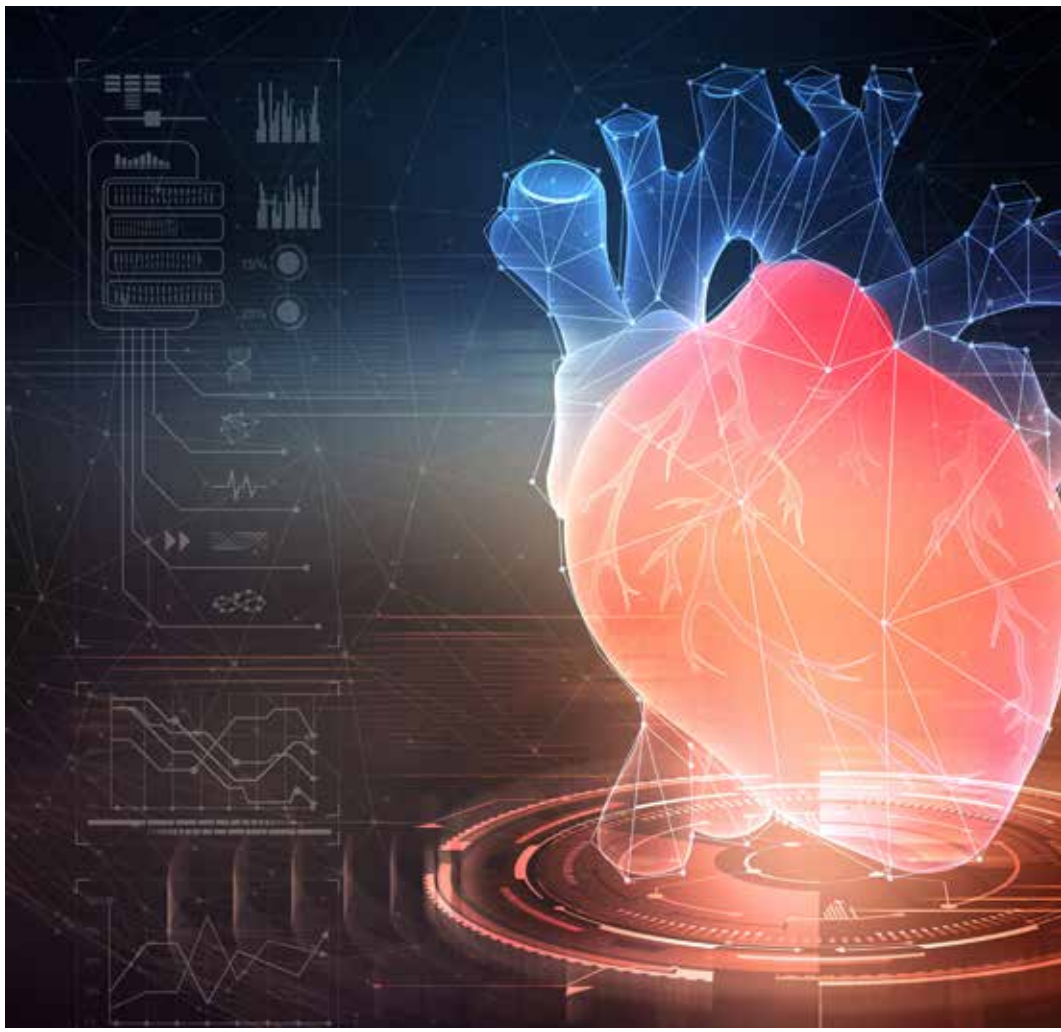
'Corporate realised that if they were going to succeed collectively they required a platform that would allow them to scientific computing, share data and collaborate,' he continued. 'Effectively they built this RCE, opened up shop and nobody came. When they investigated why, the feedback that they got was not very positive. It was largely based on the fact that this RCE was not built with researchers in mind. It was designed with good intentions but it was designed around IT.

At this point RCH was allowed to come and investigate and recommend how this might be fixed. After working with CelGene and investigating the challenges and solutions they helped the company redefine the RCE to better suit the researcher and scientists.

'We happened to be very fortunate at the time that we arrived,' notes Rainer. 'We were asked by the chief architect of this platform to come in and evaluate it. We respectfully and constructively pointed out some things that needed to be changed and they embraced most of what we had to say and redesigned this platform with them. We worked with them and brought the teams together to redesign this RCE and deployed it,' Rainer concluded.

3D heart scans to speed up NHS diagnosis

THE UK NATIONAL HEALTH SERVICE AIMS TO ACCELERATE THE DIAGNOSIS OF HEART DISEASE USING 3D IMAGING TECHNOLOGY AND DEEP LEARNING



A five-times faster than normal way to diagnose and treat patients with suspected heart disease – by using a deep learning-powered 3D imaging technology – is being introduced by the UK's NHS.

Known as HeartFlow, the latest innovation delivered as part of the National Health Service's Long Term Plan, turns a regular CT scan of the heart into a 3D image, allowing doctors to diagnose life-threatening coronary heart disease (CHD) in just 20 minutes. The programme aims to adopt the AI-powered HeartFlow fractional flow reserve – computed tomography (FFRCT) Analysis to fight CHD.

HeartFlow Analysis takes data from a coronary CT angiography (CTA) scan and uses deep learning technology and highly trained analysts to create a personalised, digital 3D model of the patient's coronary arteries. Its algorithms solve millions of equations to simulate blood flow in a patient's arteries, to help clinicians assess the functional impact of any blockages.

The National Institute of Health and Care Excellence (Nice) advocates a CTA-first approach to diagnosing CHD.

HeartFlow Analysis has been declared a best practice non-invasive option for patients in recent updates to Nice's guidance, stating that the technology should be considered as an option for patients with stable, recent onset chest pain who are offered CTA. HeartFlow plays a central role in providing clinicians greater confidence in their diagnoses, while reducing the time to diagnose patients.

> "By rapidly improving the rate we diagnose and treat those with a heart condition, we will save thousands of lives"

Once patients are diagnosed using the 3D image, treatments include surgery, medication or having a stent fitted. For less serious cases patients will be given tips on healthy lifestyle changes or cholesterol-lowering medication – meaning the risk is quickly resolved before it becomes life-threatening.

NHS medical director Stephen Powis said: 'The NHS Long Term Plan committed to cutting strokes, heart attacks and other →

major killers, as well as ensuring patients would benefit from cutting edge therapies and techniques, and HeartFlow is just the latest example of that.

'By rapidly improving the rate we diagnose and treat those with a heart condition, we will save thousands of lives and ensure, as well as delivering the most successful vaccination programme in health service history, the NHS is able to deliver routine services even quicker than before the pandemic.'

Around 100,000 people are eligible to use HeartFlow over the next three years, with more than 35,000 people set to benefit each year.

Matt Whitty, director of innovation and life sciences for NHS England, said: 'This latest innovation will help patients and will contribute to helping the NHS to recover from the pandemic as we continue to deliver on our ambitious Long Term Plan commitments to provide patients across the country with the most up to date tech, as quickly as possible.'

'HeartFlow has been a huge success in clinical trials, and will now help tens of thousands of people a year receive quick diagnosis and treatment, and ultimately save lives,' Whitty continued.

Dr Derek Connolly, consultant interventional cardiologist at Sandwell & West Birmingham Hospitals NHS Trust said: 'For every five patients who have a cardiac CT and a HeartFlow Analysis, four patients go home knowing they don't need anything else. Half of those patients will be on cholesterol tablets because they have early disease, and the other half will have normal coronary arteries.

'Incorporating the HeartFlow Analysis has had a meaningful impact at our hospitals, improving the diagnosis and treatment of the leading cause of death,' added Connolly.

The NHS Long Term Plan promised that patients would benefit from faster adoption of cutting-edge technology and treatments.

This latest technology has been rolled out across the NHS as part of the MedTech Funding Mandate policy, an NHS Long Term Plan commitment that supports the implementation of proven medical devices, diagnostics and digital products.

Since the NHS Long Term Plan was published, NHS England has rolled out the headache-busting device GammaCore, as well as Covid-friendly cancer drugs.

In January HeartFlow announced that the National Health Service England (NHSE) and NHS Improvement have mandated that English hospitals adopt the AI-powered HeartFlow FFRCT Analysis.

The HeartFlow Analysis has been selected as one of the innovations supported by NHSE's new MedTech Funding Mandate. The mandate, which began on 1 April, aims to provide innovative medical devices and digital products to NHSE patients faster, and is a key policy in helping to improve patient care and reduce costs for the public health service.

The mandate includes the option to extend funding for up to an additional three years through 31 March 2024.

> "Incorporating the HeartFlow Analysis has had a meaningful impact at our hospitals, improving the diagnosis and treatment of the leading cause of death"

At the January announcement Dr Anna Beattie, consultant cardiothoracic radiologist at Newcastle Hospitals, commented on the new technology: 'We have used the HeartFlow Analysis in Newcastle Hospitals since August 2018. It has contributed to significant changes in the way we diagnose patients and use hospital resources.

'Prior to adopting the technology, we used a coronary CT-first approach for 28 per cent of all patients referred to our rapid access chest pain clinic. Now that has risen to more than 45 per cent and the number of diagnostic-only invasive angiograms we perform has reduced. The HeartFlow Analysis is a great tool to supplement a CT-first approach by providing functional information, on top of the anatomical information from a CT scan. This is great for patients, as it reduces unnecessary risks, and means they spend less time at the hospital.'

Dr Timothy Fairbairn, consultant cardiologist at Liverpool Heart and Chest Hospital, said: 'In our practice, four out of five patients who have a HeartFlow Analysis avoid the need for further invasive testing. This enables us to treat patients more quickly, leading to an improved patient experience, and has had a positive impact on our waiting times for other non-invasive stress tests.

'The MedTech Funding Mandate allowed us to have a positive conversation with our commissioners, emphasising the real benefits our practice has experienced, and secure funding for continued use of the HeartFlow Analysis going forward. The MedTech Funding Mandate and funding commitment from commissioners means NHS hospitals will be able to continue to offer the best available technologies in our approach to the diagnosis and management of CHD for patients.'

HeartFlow's technology is already used in 60 NHSE hospitals. The extension of the ITP and introduction of the MedTech Funding Mandate will allow hospitals to continue providing cutting-edge care to patients and accelerate the adoption of the technology in hospitals throughout the country.

Lack of FAIR data reduces life science innovation

Siloed research data can limit life science organisations' impact, which negatively affects scientific innovation, writes Robert Roe

As the volume of data and the number of use cases in life sciences continue to grow, there is growing concern that a lack of reusability will impede innovation.

To solve this issue companies are exploring FAIR (Findable, Accessible, Interoperable and Reusable) data principles to enable them to make better use of data stored in an organisation.

In May 2018, the EU published a report estimating that not having FAIR research data costs the European economy at least €10.2bn every year. In addition, the report also draws a rough parallel with the European open data economy. It concluded that the downstream inefficiencies arising from not implementing FAIR data could account for a further €16bn in losses annually. Similarly, in the US, according to recent Gartner research, the average financial impact of poor data quality on organisations is significant.

The cumulative effect of unproductive time spent by every scientist across multiple departments or sites can negatively impact business outcomes. This could mean losing competitive advantage in the pharmaceutical industry, due to extended time-to-market. Similarly, in contract research organisations (CROs), motion waste can gradually diminish business potential with slower project turnarounds limiting the number of clients served.

This time-sink caused by everyday motion waste directly impacts data findability. Plus, the lack of a reliable system makes it difficult to access records for annual stock reviews or audits. During instances of data intervention by either a regulatory authority or a laboratory manager, additional time is then spent on re-capturing old data records into more preferred formats. Frequent episodes of data cleansing can result in substantial motion waste, as scientists pause research projects to instead get inventory records in order.

As new and emerging technologies come into play, and secondary use cases for existing research data continue to grow, it is becoming increasingly important for data to be accessible, rather than siloed, within specific departments or disparate data systems. An example of this is AI and machine learning, as these approaches become more popular and more widely adopted.

These techniques require researchers to take large sets of historic data and apply them to solve more problems and ask new questions – again leading to new data uses and requiring fresh data management techniques that can support these large data sets.

> "The cumulative effects of unproductive time spent by every scientist across multiple departments or sites can negatively impact business outcomes"

This means that organisations have the potential to benefit greatly from a shift in the way they manage data and data sharing. Rather than individual scientists' data only being used by a specific person or team for one purpose, today data can be used by the entire company and even the wider industry, to advance innovation. However, the industry is still playing catch-up to make this data sharing a reality.

Covid-19 has highlighted the urgency of addressing these problems and has provided a wake-up call for many organisations. To ensure that organisations can respond quickly to such events in the future, scientists need to be able to access the right data in a functional form as quickly as possible. This data may need to be shared with other life science and biotech companies, and potentially integrate with large-scale real-world evidence (for example, data from self-reporting mobile apps like ZOE in the UK). FAIR data is essential to bring global solutions to such a huge public health crisis and others that may follow.

AI requires new ways of managing data

In recent years, the life sciences industry has suffered an unignorable decline in innovation efficiency, but AI has the power to change this. Drug developers are looking to bring together everything that is known about a problem, to build a more accurate and nuanced picture of patients, diseases and medicines. As such, we need new ways to capture and manage these varied data. FAIRification is one such way.

Employing data to build more realistic, multi-dimensional analyses will help researchers better understand diseases and assess how chemical entities behave in biological systems. Data that are structured in line with FAIR principles, and so are interoperable and reusable, will make this approach possible. It's also an approach that promises to slash drug development times and vastly reduce late-stage failures.

To build such in-depth patient and product profiles, life science companies need access to greater volumes of data external to their organisation, including: public domain sources (such as PubMed, ClinicalTrials.gov, FDA); commercial intelligence (such as Sitetrove, Pharmaprojects, Pharmapremia); data provided by CROs; and real-world evidence (such as electronic health records (EHR) and patient self-reporting).

Since 2016, the FAIR Data Principles have been adopted by the European Union (EU), together with a growing number of pharmaceutical companies, research organisations and universities.

To accelerate innovation and productivity, more organisations and public bodies will need to follow in their footsteps.

Common challenges to FAIR implementation

While the ideas behind the FAIR principles have been around for some time, implementation in the life sciences has been slow, because the path to adoption is neither finite nor predetermined.

FAIRification is the long-term overhaul of how data are created and used in an organisation, and this process is continuously influenced by an ever-changing knowledge landscape. When organisations begin the FAIRification journey, they face some common challenges, including:

- Unstructured legacy data – often data are not tagged, contain haphazard names or identifiers and lack common terminology;
- Data silos and trapped historical data – technologies used in previous research are likely obsolete or no longer supported; often personnel responsible for creating original datasets have moved on, leading to data becoming inaccessible or uninterpretable;
- Scientific complexity – machine-readable representations of biological information can quickly become extremely complex;
- Ontology management – there are multiple competing ontologies and vocabularies, often even in a single organisation, with little standardisation across the industry; and
- Cultural barriers – changing the culture of an organisation can be one of the most challenging tasks; researchers and organisations are typically very protective of even non-proprietary data. Incentivising all parties to do their parts in generating high-quality FAIR data will require valuing efforts to that end, as much as the marketable output of a drug development pipeline.

Path to implementation

FAIRification does not happen immediately and comprehensively. Making data FAIR is an evolving and progressive process.

This is especially true for the pharmaceutical industry, where data production is continuous and new knowledge is always reshaping the information landscape for research questions.

However complex FAIRification may seem, it is critical to start the process and allow for an agile, test-and-learn adoption. Helpfully, companies do not need to go it alone. There is a large and growing network of organisations offering assistance, expertise and tools to help FAIRify data.

This includes the Analytical Information Markup Language (AnIML), which is an emerging ASTM XML standard for analytical chemistry and biological data. AnIML was instigated to solve problems with existing data standards and take advantage of the recent development of the eXtensible Markup Language (XML). The goal of the ASTM working group is to develop an analytical data standard that can be used to store data from any analytical instrument.

The Pistoia Alliance, a non-profit group advocating for better data sharing in life sciences, also offers a free FAIR toolkit for implementation.

The FAIR movement is not the first attempt at merging data from disparate sources. But it is gathering pace, due to the combination of increased computing power, data availability and data generation that are making AI a reality for many organisations. To make effective use of this shift to advanced analytics, life science organisations should take the first steps towards FAIR data.

Weathering the storm

GEMMA CHURCH FINDS OUT HOW SIMULATION AND MODELLING IS UNLOCKING LONG-TERM, ACCURATE WEATHER AND CLIMATE PREDICTIONS

Things are hotting up in the world of weather and climate forecasting thanks to today's highly ambitious projects. Some initiatives are building world-class supercomputing facilities. Others are creating digital twins of the Earth. But what impact does this work have on the world of simulation and modelling?

Jon Petch, associate director of weather science at the Met Office, explained: 'Climate and weather predictions, produced using supercomputing, are ever-increasing in complexity and size as atmospheric physics develops and new data is captured from earth monitoring systems.'

This is where new supercomputing facilities and modelling techniques can help. In April, the Met Office signed a multimillion-pound agreement with Microsoft for the provision of a supercomputer to accelerate weather and forecasting in the UK. The supercomputer will be twice as powerful as any other machine in the country.

It's a massive step forward for both UK research and the world of weather and climate prediction. Petch explained: 'This investment is for a 10-year service delivery and includes two substantial increases in supercomputing capacity through two generations of supercomputing implementations, with a return on investment of around 9:1 - resulting in financial benefits totalling up to £13bn for the UK over its ten-year lifespan.'

'Increased capacity will permit an increase in the detail of both ocean and atmospheric models, allowing a more realistic representation of the large-scale weather systems that drive UK weather,' according to Petch. 'It will enable ever more localised climate predictions, ensuring infrastructure, housing, transport networks etc, built today, will be safe from the weather impacts of the future.'

Microsoft Azure's Supercomputing-as-a-Service will also be used during the project, allowing the Met Office to 'leverage the best blend of dedicated and public cloud services to provide more accurate predictions to help the UK population and businesses plan daily activities, better prepare for extreme weather, and address the challenges associated with climate change,' according to a Microsoft spokesperson.

'We are delighted to be working with the Met Office to deliver what will become the world's leading climate and weather science supercomputing service. Combining the Met Office's expertise, data gathering capability and historical archive with the sheer scale and power of supercomputing on Azure will improve forecasting, help monitor and tackle climate change and ensure the UK remains at the forefront of scientific and technological research over the next decade,' the spokesperson added.

The new supercomputing facility will extend the Met Office's longer-range predictions, enhancing accuracy and supporting medium-term decision-making for business and industry, and improve its understanding and analysis of climate change, while driving 'technological innovation by UK business and industry,' Petch noted.

The supercomputer itself will pack a powerful prediction punch. The Microsoft spokesperson said: 'The first generation of the supercomputer solution will have a combined total of more than 1.5 million processor cores and over 60 Pflops, otherwise known as 60 quadrillion (60,000,000,000,000,000) calculations per second of aggregate peak computing capacity. Microsoft will also deliver further upgrades in computing capability over the ten years.'

Petch added: 'Additionally, the increased supercomputing power will allow us to increase the number of model runs we undertake, which will improve assessments of current risks and predictions, undertake rapid attribution of severe weather in relation to our changing climate, and allow the characterisation of future worst cases – all of which is very compute hungry.'

> "Increased capacity will permit an increase in the detail of ocean and atmospheric models, allowing a more realistic representation of large-scale weather systems"

Updating the dynamical core

This is an important point. As prediction ambitions and supercomputers scale up, the number of computational resource and data management challenges also increases.

This issue is something that the Met Office is more than aware of. A project to redesign the Met Office's dynamical

core (the numerical algorithm at the heart of its atmospheric model) for the next generation of supercomputers has just marked its 10th anniversary.

'This is part of a larger programme to reformulate and redesign our complete weather and climate research and operational/production systems to allow the Met Office and its partners to fully exploit future generations of supercomputers for the benefits of society,' Petch explained.

'It covers the atmosphere, land, marine and Earth system modelling capabilities and ranges from observation processing and assimilation, through the modelling components, to verification and visualisation,' he added.

One of the key parts of a weather or climate model is the dynamical core – the numerical algorithm that solves the equations governing fluid motion. The Met Office's current dynamical core is known as ENDGame, and it describes the Earth using a latitude-longitude grid.

'However, it has been known since the late 2000s that this grid will cause problems on future supercomputers, which will rely on spreading the calculations involved in the simulation over ever-increasing numbers of computer processors,' Petch explained.

'A programme to reformulate and redesign the dynamical core at the heart of our weather and climate model is underway (GungHo) together with a programme to design, develop and implement a new model infrastructure with the specific aim of being as agnostic as possible about the supercomputer architectures (LFRic).'

With these solutions and increased supercomputing capability, the Met Office can 'make a step change in the level of precision with which it can forecast the impact of severe weather, with city-scale predictions of rainfall, winds and air quality, helping to protect life and property,' Petch added.

By replacing and increasing its supercomputing capability, the Met Office also hopes to expand its localised climate projections to better inform future climate risk. This includes city-scale projections to 'enable better investment in infrastructure and adaptation measures to keep people safe,' said Petch.

Digital Earth

Destination Earth (DestinE) is another key initiative, from the European Union, which is tasked with developing a high-precision model of the Earth to monitor and simulate both natural and human activity.

> "This investment is for a 10-year service delivery and includes two substantial increases in supercomputing capacity through two generations of supercomputing implementations"

To create a digital twin of the Earth, an approximately one to three kilometre global grid spacing is required between neighbouring simulation (grid) points. These points represent as many physical processes as possible from first principles to 'simulate as observed' and make the digital twin seamlessly interact with other applications and users.

To achieve this, vast amounts of natural and socio-economic information are required to continuously monitor the health of the planet and support EU policy-making and implementation.

From a data management perspective, the challenges are extensive. Nils Wedi, head of the European Centre for Medium-Range Weather Forecasts' (ECMWF) Earth Modelling Section, said: 'To put it in perspective, a single simulation will produce 100 to 200TB per day, similar to todays' entire volume of daily production at the ECMWF.'

Wedi added: 'We anticipate using the latest developments on federated data access, such as Polytope datacube access of weather data, and federated data lakes, combined with unsupervised learning and data reduction. It is not anticipated to be able to archive native resolution data for longer periods and beyond certain cut-off times raw data will have to be deleted.'

The Polytope datacube is one example of the new processes and technologies being put in place to help manage this data. It stores meteorological datasets in n-dimensional arrays (or datacubes) so data is returned in an accessible format.

The ECMWF is also following a four-strategy approach to adapt its HPC architectures and develop 'an accelerator-enabled multi-architecture prediction model,' according to Wedi.

First, ECMWF is introducing the use of single (instead of double) precision accuracy in its forecast algorithms. Second, it uses platform-specific →

→ accelerated libraries for computationally intensive parts of the model. Third, it is using separate data-layout, memory placement and science-driven code developments, enabling asynchronous and data-driven programming models, and the use of source-to-source translators (using DSL toolchains).

Finally, the ECMWF is also developing alternative and novel algorithms, for example, part replacing time-critical code with machine learned equivalents, and/or the use of alternative discretisations that are potentially better suited for emerging HPC.

'How we can best use a quantum computer is still to be answered, but researched,' Wedi said.

Streamlined simulations

There are many efforts to streamline weather and climate change prediction systems. The National Oceanic and Atmospheric Administration (NOAA), for example, is part of a broader community modelling effort called the Unified Forecast System (UFS).

Dr Vijay Tallapragada, chief of the Modelling and Data Assimilation Branch in NOAA's Environmental Modeling Center, explained: 'UFS is integrating numerous environmental models into a unified Earth modelling system that will be used to predict weather from local to global domains at time scales from minutes to seasons.

'This unified system allows better collaboration between NOAA and the extramural science community, and will accelerate the development and integration of innovation into NOAA's operational weather forecast systems.'

NOAA is migrating towards simplifying the operational production suite by adopting the community-based UFS for all operational applications in the next five years. Both its Global Forecast System (GFS) and Global Ensemble Forecast System (GEFS) have already been migrated to the UFS framework, and the rest of its applications are currently being developed and merged into the same framework to streamline its research and operations.

The Met Office also uses a Unified Model of the atmosphere for both its weather and climate applications. 'Although we have several modelling systems available to us, the Unified Model is key for our weather forecasts and climate predictions,' Petch added.

To help manage the resulting, demanding workloads, commercial tools are also available. Altair, for example, provides workload management solutions for the world's weather sites, such as the National Center for Atmospheric Research (NCAR) in the United States, Australia's Bureau of Meteorology, and the US Naval Research Laboratory.

Sam Mahalingam, Altair's CTO, said: 'Sophisticated and well-supported HPC workload management and optimisation are a must for these sites, where HPC downtime, productivity loss and inefficient resource utilisation can threaten critical real-world research.'

At NCAR, Altair's PBS Professional is already used for workload orchestration on the organisation's current supercomputer, Cheyenne. PBS Professional and Altair Accelerator Plus will also be used on its new system, Derecho, which is predicted to be one of the world's Top 25 HPC systems.

'Features like high-throughput hierarchical scheduling with Accelerator Plus offer six to ten times HPC throughput improvements, as well as better license and resource utilisation, and more flexible scheduler usage models,' Mahalingam explained. 'At NCAR, this will help develop and test the Weather Research and Forecasting model for atmospheric research and operational forecasting applications.

'Other features, such as cloud bursting, which provides massive scalability and flexibility, and key access portals and alerting mechanisms, are also critical to weather research and climate modelling,' according to Mahalingam.

At Australia's Bureau of Meteorology, staff members at many sites are required to constantly monitor the environment. There, an Altair solution that leverages the Cylc workflow engine provides detailed information to staff, allowing them to monitor the supercomputer hardware, Cylc suites, and PBS Professional jobs, while reporting status clearly and concisely.

The solution is designed to be modular and general-purpose, so any site can deploy it out of the box, or substitute components they're more familiar with.

Seamless integration is a sign of things to come in the world of weather and climate prediction, as Mahalingam explained: 'We expect the use of multi-dimensional HPC such as storage-aware scheduling and hierarchical scheduling, cloud bursting and automated cloud migration, and workload simulation, as well as the use of HPC to propel machine learning applications, will continue to gain traction in the coming months and years.'

Such developments are key to not only futureproof the world of weather and climate forecasting, but also protect our planet. Petch concluded: 'Predicting the weather and climate has become one of the most important areas of scientific endeavour, and increasing our computing capability is essential if we are to continue to improve our climate predictions and climate change simulations.'

"Features like high-throughput hierarchical scheduling with Accelerator Plus offer six to ten times HPC throughput improvements"

The latest news stories in scientific computing

MODELLING AND SIMULATION

Supernovae twins open possibilities for precision cosmology

Cosmologists have found a way to double the accuracy of measuring distances to supernova explosions – one of their tried-and-true tools for studying the mysterious dark energy that is making the universe expand faster and faster.

The findings could help researchers enhance dark energy experiments at major telescopes.

The results from the Nearby Supernova Factory (SNfactory) collaboration, led by Greg Aldering, of the Department of Energy's Lawrence Berkeley National Laboratory (Berkeley Lab), will enable scientists to study dark energy with greatly-improved precision and accuracy, and provide a powerful crosscheck of the technique across vast distances and time. The findings will also be central to major upcoming cosmology experiments that will use new ground and space telescopes to test alternative explanations of dark energy.

Two papers published in *The Astrophysical Journal* report these findings, with Kyle Boone as lead author. Currently a postdoctoral fellow at the University of Washington, Boone is a former graduate student of Nobel laureate Saul Perlmutter, the Berkeley Lab senior scientist and UC Berkeley professor who led one of the teams that originally discovered dark energy. Perlmutter was also a co-author on both studies.

'We've long had this idea that if the physics of the explosion of two supernovae were the same, their maximum brightnesses would be the same. Using the Nearby Supernova Factory spectra as a kind of CAT scan through the supernova explosion, we could test this idea,' said Perlmutter.

Supernovae were studied in 1998 to make the discovery that the expansion of the universe is speeding up, not slowing down as had been expected. This acceleration – attributed to the dark energy that makes up two-thirds of all energy in the universe – has since been confirmed by a variety of independent techniques, as well as with more detailed studies of supernovae.

The results by the SNfactory come from a multi-year study devoted to increasing the precision of cosmological measurements made with supernovae.

Measurement of dark energy requires comparisons of the maximum brightnesses of distant supernovae billions of light-years away with those of nearby supernovae 300 million light-years away. The team studied hundreds of such nearby supernovae in exquisite detail. Each supernova was measured a number of times, at intervals of a few days. Each measurement examined the spectrum of the supernova, recording its intensity across the wavelength range of visible light.

Supernovae twins open possibilities for precision cosmology

An instrument custom-made for this investigation, the SuperNova Integral Field Spectrometer, installed at the University of Hawaii 2.2-metre telescope at Maunakea, was used to measure the spectra.

The SNfactory collaboration includes Berkeley Lab, the Laboratory for Nuclear Physics and High Energy at Sorbonne University, the Centre for Astronomical Research of Lyon, the Institute of Physics of the 2 Infinities at the University Claude Bernard, Yale University, Germany's Humboldt University, the Max Planck Institute for Astrophysics, China's Tsinghua University, the Centre for Particle Physics of Marseille, and Clermont Auvergne University.

MODELLING AND SIMULATION

Microsoft supercomputer to boost Met Office's weather forecast

The Met Office has signed a multimillion-pound agreement with Microsoft for a supercomputer to accelerate weather and climate forecasting in the UK.

The supercomputer – expected to be the world's most advanced dedicated to weather and climate – will be in the top 25 in the world and twice as powerful as any other in the UK. Its data will be used to provide more accurate warnings of severe weather, helping to build resilience and protect the UK population, businesses and infrastructure from the impact of increasingly extreme weather.

Penny Endersby, chief executive of the Met Office, said: 'We are delighted to be working in collaboration with Microsoft to deliver our next supercomputing capability. Working together, we will provide the highest quality weather and climate datasets, and ever more accurate forecasts that enable decisions to allow people to stay safe and thrive. This will be a unique capability that will keep not just the Met Office but the UK at the forefront of environmental modelling and high-performance computing.

'This investment by the UK government is a great vote of confidence in the Met Office's world-leading status as a provider of weather and climate science and services, as well as in our national commitment to build a more resilient world in a changing climate, helping build back greener across the UK and beyond,' Endersby said.

The supercomputer will also be used to take forward ground-breaking climate change modelling, enhance the Met Office's global expertise in climate science. The precision and accuracy of its modelling will help to inform Government policy as part of the UK's fight against climate change, and its efforts to reach net zero by 2050.

The news follows an announcement by the UK Government in February last year that committed £1.2bn to develop this state-of-the-art supercomputer.

Business Secretary Kwasi Kwarteng said: 'This partnership between the Met Office and Microsoft to build the world's most powerful weather and climate forecasting supercomputer is a ringing endorsement for the UK's credentials in protecting our environment, as we prepare to host COP26 later this year.

'The new supercomputer, backed by a billion-pound UK government investment, will act as a catalyst for unlocking new skills, technologies and jobs right across our economy – from data scientists to AI experts – all as part of our efforts to build back better and create a cleaner future.'

Clare Barclay, chief executive officer at Microsoft UK, said: 'The Met Office has long been synonymous with excellence and innovation in our understanding of the impact of weather and climate. To make progress with the ecological challenges we face requires innovation, technology and partnerships. The potential of the deep expertise, data gathering capacity and historical archive of the Met Office, combined with the sheer scale and power of supercomputing on Microsoft Azure will mean we can improve forecasting, help tackle climate change and ensure the UK remains at the forefront of climate science for decades to come.'

As the Met Office and Microsoft work together over the next 10 years, it is expected the collaboration will deliver scientific and technological innovation that will ensure the Met Office and the UK is ready to harness the next generation of supercomputing and data technologies – allowing experts to answer many of the big questions presented by changing climate.

The supercomputer will be based in the south of the UK and will boost employment, apprenticeships, internships, mentoring opportunities, training in digital skills and support for startups.

Met Office and Microsoft announce new supercomputer to support UK weather simulation

LABORATORY INFORMATICS

Labforward ramps up Series B financing

Labforward has added €3m to the Series B funding from January, bringing the total funds raised to €8.5m.

The majority of the new investment comes from the Fielmann Family Office. The financing is combined with funding from existing investor IBB Ventures, as well as existing and new business angels invest in the laboratory software company with headquarters in Berlin, Germany.

Laboratories have always involved a lot of manual processes that require the presence of an entire laboratory team on site. In the Covid era, the degree to which labs have already digitised or automated these processes makes the difference between being able to safely continue important scientific work or having to choose between a complete shutdown or putting employees at risk for several months.

Dr Simon Bungers, co-founder and CEO of the company explains that there has been increased demand for Labforward software solutions: 'Last year's experience showed lab teams the importance of investing in automation and digital data management to keep a lab running with fewer on-site staff. We're seeing increased demand for our solutions which let teams control devices remotely, capture data digitally and let scientists access the data in the cloud or an accessible server for analysis and process optimisation.'

Labforward software allows scientists to capture digital data remotely

LABORATORY INFORMATICS

Cytel has appointed Dr Kyle Wathen

Bayesian clinical trial design specialist Dr Kyle Wathen has joined Cytel as vice president of scientific strategy and innovation.

Dr Wathen has more than 20 years' experience in innovative adaptive clinical trial design and was a major contributor to the first adaptive platform trial – I-SPY2 – which helped to establish an 'adaptive' clinical trial model.

He will support the ongoing maturation of Cytel's Solara and East Bayes software. These solutions play a key role in Cytel's reshaping of clinical development, removing critical roadblocks, and making complex innovative trial designs more accessible.

'The growing adoption of complex designs – facilitated by enabling tools like East and Solara – is a momentous step forward for the industry,' said Dr Wathen.

'I'm dedicated to further elevating Cytel's innovative design arsenal and empowering customers to carve a smoother, more efficient path to market. My priority will be the maturation of Solara – a radical new tool that eliminates siloed decision-making for quicker identification of the best design. This is the future of clinical strategy – a leap from traditional, cumbersome clinical trial design selection processes.'

The rapid pace of drug development and the urgency of getting new medicines to market means drug developers are exploring innovative clinical trial designs to reap the benefits of reduced study duration, better decision-making and minimised costs.

But such designs have traditionally been off-limits to smaller biotechs, as they are difficult to implement, require powerful computing power, and rely on deep statistical expertise for execution.

The appointment of Dr Wathen represents the latest step in Cytel's journey of providing more equitable access to Bayesian expertise, and follows the recent acquisition of Bayesian trial design and implementation specialists Laiya Consulting, co-founded by renowned biostatistician Professor Yuan Ji. Dr Wathen has rich experience in academia, consulting and industry.

HIGH-PERFORMANCE COMPUTING

HPC improves engine cooling innovation for SMEs

Advanced Innovative Engineering (AIE), an SME based in Lichfield, UK, will use HPC resources from PRACE to improve the development of rotary engines for unmanned vehicles.

The 12th PRACE Shape call generated 16 proposals. One was taken forward from AIE, which specialises in the development of rotary engines for unmanned vehicles.

Over the next few months, domain experts at AIE will collaborate with HPC experts at the STFC Hartree Centre to develop and refine AIE's simulations to make cooling improvements to power units using HPC.

AIE manages entire project life cycles through concept, prototype and production. Working with international partners and customers, it creates technologies that combine low total-cost-of-ownership (TCO) with exceptional reliability and versatility for commercial and defence markets.

Key to achieving a highly power-dense and simple air-cooled engine is the ability of the required heat exchange area of the engine (fin or cooling pack) to reject the heat from the engine to the atmosphere as efficiently as possible.

Due to the large number of design and parameter iterations required to achieve an optimised heat exchanger design, sheer computing power becomes a significant factor in reaching a prototype design in reasonable timescales. This is where HPC comes in.

A dedicated AIE technical expert will collaborate with STFC specialists to develop and refine suitable mesh representations of candidate heat exchangers to facilitate Computational Fluid Dynamics (CFD) computations at scale on an HPC platform. The computations will involve the open source software OpenFOAM, which already runs efficiently on HPC platforms.

From these simulations, the performance of candidate heat exchanger designs will be determined with precision. All results will be validated against experimental data.

The Thirteenth Call for Applications to Shape (SME HPC Adoption Programme in Europe) closes on 1 June.

PRACE invites applications from European SMEs with an interesting idea that would benefit from HPC to increase their competitiveness.

PRACE is funded by the PRACE members. The implementation phase of PRACE receives funding from the EU's Horizon 2020 Research and Innovation Programme (2014-2020).

Suppliers' Directory

Find the suppliers you need quickly and easily
www.scientific-computing.com/suppliers

HIGH-PERFORMANCE COMPUTING

Altair
(248) 614-2400
info@altair.com
www.altair.com/hpc-cloud-applications

Ansys
ukinfo@ansys.com
www.ansys.com

Boston LTD
Tel: +44 (0) 1727 876 100
sales@boston.co.uk
www.boston.co.uk

NAG
www.nag.com
E: info@nag.com

Nvidia
Tel: +1 (408) 486-2000
Email: info@nvidia.com
www.nvidia.com/dgx

Oracle
OracleForResearch_ww@oracle.com
https://www.oracle.com/oracle-for-research/

Panasas, Inc.
Tel: +1 (408) 215 6800
info@panasas.com
www.panasas.com

PRACE
Tel.: +32 2 613 09 27
communication@prace-ri.eu
www.prace-ri.eu

Samsung
Tel: +49 89 45578-1000
semi.eu@samsung.com
www.samsung.com/semiconductor

SoftIron
UK: +44 20 3890 5466
USA: +1 855 450 CEPH (2374)
www.softiron.com
info@softiron.com

LABORATORY INFORMATICS

Abbott Informatics
Tel: +44 (0) 161 711 0340
AISalesUK@abbott.com
www.informatics.abbott

Achiever Medical LIMS Interactive Software Limited
Tel: +44 (0) 121 380 1010
enquiries@interactivesoftware.co.uk
www.interactivesoftware.co.uk

Advanced Chemistry Development, Inc. (ACD/Labs)
Tel: +1 (416) 368-3435
info@acdlabs.com
www.acdlabs.com

Autoscribe Informatics
Tel: +44 (0) 118 984 0610
info@autoscribeinformatics.com
www.autoscribeinformatics.com

BSSN Software
+49 6151 78537 39
peter.m.maier@merckgroup.com
bssn-software.com

Collaborative Drug Discovery, Inc.
Tel: +44 (0) 1223 803830
info@collaborativedrug.com
www.collaborativedrug.com

Digital Lab Consulting Ltd
Tel: +44 (0)1543 547607
info@dlc.io
www.dlc.io

Dotmatics
+44(0)1279 654 123
www.dotmatics.com/
Sales@dotmatics.com

LabWare
Tel: +44 161 927 5600
infoeu@labware.com
www.labware.com

iVention Lab Automation
Tel: +31 (0) 38 452 83 75
info@ivention.nl
www.ivention.nl

MilliporeSigma
+44 7866 116946
haydn.boehm@merckgroup.com
sigmaaldrich.com

Thermo Fisher Scientific Informatics
Tel: +1 (215) 964-6124
www.thermofisher.com/informatics

MODELLING AND SIMULATION

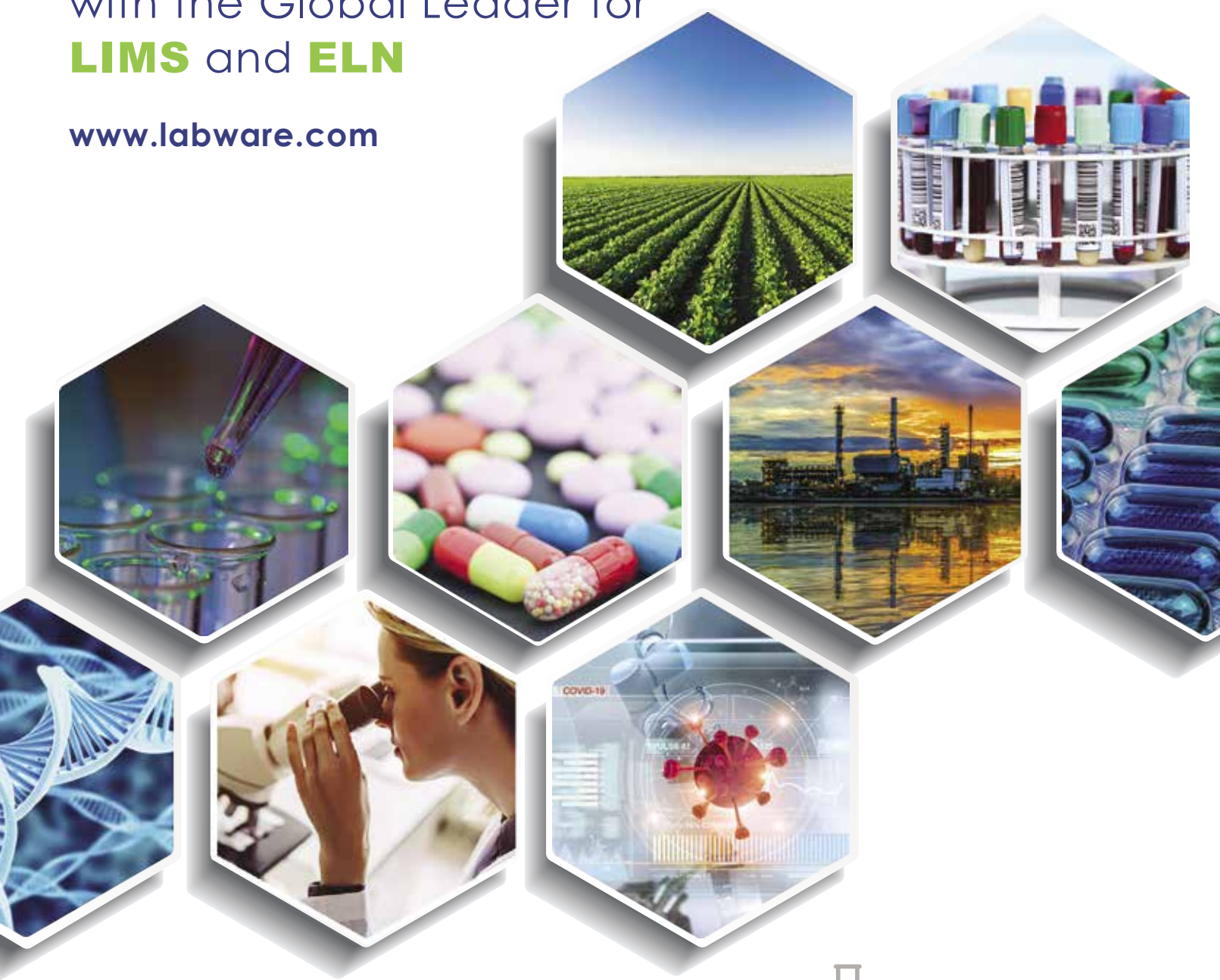
Altair Engineering, Inc
(248) 614-2400
info@altair.com
www.altair.com


MERCK
Dive into your data
BSSN™ SciDiver lets you access and analyze your laboratory data with anyone, anywhere, on any device.
Scientific breakthroughs aren't confined to the bench. Collaborate with remote colleagues and connect your instruments with SciDiver, a secure cloud solution for organizing, harmonizing, and visualizing all your analytical and biological data.
Learn more at bssn-software.com
BSSN Software is now part of MilliporeSigma
The life science business of Merck operates as MilliporeSigma in the U.S. and Canada.
© 2020 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved. Merck, the vibrant M and BSSN Software are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources. 34066 11/2020

LABWARE® 8

Automate Your Laboratory
with the Global Leader for
LIMS and **ELN**

www.labware.com



LABWARE®
Results Count