Laboratory Informatics Guide 2023

Understanding the needs of the laboratory

www.scientific-computing.com

From the publishers of

SCIENTIFIC COMPUTING WORLD

Laboratory software

Connect to a greater scientific experience

Thermo Scientific SampleManager LIMS

Our next-generation laboratory software uses innovative technologies to connect your organization to the future. Join us on a journey to work smarter not harder – to use digital capabilities to streamline your workflows and eliminate manual tasks. Digital transformation is critical to your success.

Version 21 of Thermo Scientific™ SampleManager™ LIMS delivers transformational laboratory management for the lab of the future. This latest release connects to industry-leading calculation engines, and combines with powerful data analytics to deliver crucial insights and more informed decisions, coupled with extended reality (XR) connectivity for hands-free control.

Discover more at **thermofisher.com/samplemanager**

thermo scientific

CONTENTS

Welcome to Laboratory Informatics Guide 2023

4. Supercharging the silo

Bioinformatics software helps scientists make sense of complex scientific data

8. Managing change in genomics

Changes to the interface between science and technology are shaping future genomics research

10. Software impact on industry and academia

Chemistry software supports scientists in academia and industry, writes **Sophie Ktori**

14. Machine learning and the role it plays in classifying patients

Eddie Cano-Gamez discusses how machine learning is helping researchers to classify patients with sepsis

16. Genomic potential

Genomics software is helping to transform large volumes of unstructured data into actionable knowledge, writes **Sophia Ktori**

20. Finding the right laboratory software

With many competing products on the market, how can scientists and researchers find the right laboratory software?

22. Open automation

Sophia Ktori discusses the importance of integration and open systems in supporting laboratory automation

26. Tackling reproducibility with digitisation

Dr Birthe Nielsen discusses the role of the Methods Database in supporting life sciences research by digitising methods data across different life science functions

28. Driving adoption of the paperless lab in India

Sachin Bhandari provides an overview of his talk at Paperless Academy India

30. Supplier's directory

The Laboratory Informatics Guide aims to highlight trends and changes in the laboratory informatics software market over the past 12 months. The content for this year's edition consists of a mix of features from throughout the year, coupled with expert interviews from prominent research organisations.

This year, the content focuses on key industry verticals, such as bioinformatics and pharmaceuticals, coupled with critical technology trends such as AI and digital transformation. The interviews also support these themes with a focus on large-scale research, machine learning and paperless lab technology.

On page 4, we have the first feature, which explores the role bioinformatics software plays in uncovering complex data patterns. Following that, on page 8, Dr James McCafferty discusses how changes to the interface between science and technology are shaping future genomics research. Sanger is one of the largest genomics research facilities in the world, so they have significant challenges in meeting the demands of large-scale research.

Starting on page 10, we have a feature that explores how cheminformatics software supports scientists in academia and industry. We then have the second interview, on page 14, which explores the use of machine learning to better characterise patients with sepsis.

The next feature, on page 16, also explores the use of genomics software and how it is helping scientists to transform large volumes of unstructured data into actionable knowledge with open source tools. On page 20, we have an article exploring the steps taken to select the right laboratory software.

The final feature of this year's LIG starts on page 22, with a look at automation and how this trend will shape future labs. Complementing this topic we have an article that highlights the role of the Methods Database in supporting life sciences research by digitising methods data across different life science functions.

Finally, we have an interview focused on driving digital transformation in India. On page 28, Sachin Bhandari provides an overview of his talk at Paperless Academy India.

SUBSCRIPTIONS: The *Laboratory Informatics Guide 2023* is published by Europa Science Ltd, which also publishes *Scientific Computing World*. Free registration is available to qualifying individuals (register online at www.scientific-computing.com). Subscriptions £100 a year for six issues to readers outside registration requirements. Single issue £20. Orders to ESL, SCW Circulation, 4 Signet Court, Cambridge CB5 8LA, UK.
Tel: +44 (0)1223 221030.
Fax: +44 (0)1223 213385.

US copies: *Scientific Computing World* (ISSN 1356-7853/USPS No 018-753) is published bi-monthly for £100 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 image: Yurchanka Siarhei/ Shutterstock.com

All other images: Shutterstock.com

EDITORIAL AND ADMINISTRATIVE TEAM
Managing editor: Annabel Ola (editor.scw@europascience.com)
Editor: Robert Roe (robert.roe@europascience.com)
Feature writer: Sophia Ktori
Circulation/readership enquiries: subs@europascience.com

ADVERTISING TEAM
Advertising Sales Manager: Lexi Taylor (lexi.taylor@europascience.com)
Tel: +44 (0) 1223 221039

DESIGN TEAM
Production manager: David Houghton (david.houghton@europascience.com)
Senior graphic designer: Justin Zwierzanski (justin.zwierzanski@europascience.com)

CORPORATE TEAM
Managing director: Warren Clark
Head of content: Mark Elliott
Web: www.scientific-computing.com

LABWARE®

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

www.labware.com

LABWARE® Results Count

STARLIMS

ONE PARTNER, ONE POWERFUL SOLUTION.

More than just working towards a paperless lab, digital transformation may allow companies to derive more intelligence from their data, ultimately improve product and process safety, and enable faster regulatory audit or approval timelines.

For more than 35 years in the market, STARLIMS is a mission critical application that supports this digital transformation from the ground up.

Discover how our solutions can support you in the digital transformation journey.

Supercharging the silo

Sophia Ktori explores the role bioinformatics software plays in helping scientists to make sense of complex scientific data

elenabsl/shutterstock

A key ongoing issue for organisations involved in drug discovery research is how to manage chemical and biological data in combination.

Abraham Wang, head of marketing at Collaborative Drug Discovery (CDD), said: 'Drug discovery and development inevitably combines chemistry and biology, but there has traditionally been no intuitive, data-rich way to hold and interrogate data on both types of entity in one repository.'

This results in what Wang calls 'a data silo' between biologists and chemists: 'Our customers tell us there are plenty of software systems that cater for the chemists who are synthesising compounds, and separate systems for the biologists who are running the assays to test the efficacy of the compounds. But you end up with an artificial division of data because the one won't work alongside the other.'

CDD's flagship CDD Vault offers a complete informatics platform and ELN for housing, managing and querying all compound and experimental data and metadata in one space. 'Our strengths have traditionally been in managing small molecules and structure-activity relationships,' Wang said. 'But we also realised an ever-growing need for a platform that could also handle biologics, and this represented an important step in our evolution.'

The goal was to fill this gap between the management of chemical and biological data. 'We wanted to develop the Vault into an interconnected platform for biological and synthetic molecule data management and querying – after all, 50 per cent of new drugs being developed today are biologics,' Wang noted. 'Its about bringing chemists and biologists together, not keeping them apart.'

With enhancements to the platform released at the end of 2021, CDD Vault users can now register and analyse their biologic entities both alongside, and in combination with, synthetic molecules. 'It's a great start,' Wang said. 'Our customers who have been using CDD Vault for compound management can now register and manage their plasmids, antibodies, peptides, proteins, nucleic acids and even mixtures and complex entities, such as antibody drug conjugates that combine an antibody with a linker and a synthetic compound.'

Key data is stored for each molecule in a particular registration entity – say, nucleotide, amino acid mixture – and then properties, such as molecular weight and other compositional information, are automatically generated by the Vault for each registration: 'Within the next few months, we are going to add some key additional features, including the capability to carry out sequence searches and visualise plasmids,' Wang added.

> "It's about bringing chemists and biologists together, not keeping them apart..."

The flexibility of the CDD platform means users can also either set up one single Vault for all of their entities, biological and chemical, or establish separate Vaults for, say, cell lines, nucleic acids, antibodies and compound libraries, so that specific entry fields can be set for each type of entity. Wang said: 'What's important is the Vaults can be

cross-interrogated, so there is no loss of intelligence by having separate Vaults.'

As far as CDD is aware, this combined depth of management, degree of flexibility and interoperability isn't available elsewhere in a single platform for both synthetic and biologics. Ultimately, it allows organisations to retain and manage all their key data, and importantly, capture and retain metadata associated with experiments – with connection to the design, synthesis and testing of synthetic and biologic molecules – on a single platform, to help ensure there is no loss of either content, or context.

Wang said: 'Data is a lab's most valuable asset, and by using CDD Vault, organisations can keep their chemical and biological data relevant, clean and accessible. We all understand that spreadsheets and paper notebooks can't be searched or cross-referenced easily, and it's also hard to keep these types of files up to date. With CDD Vault, organisations now have what we call a single source of truth immediately available. Everyone is looking at the same, current data, which facilitates decision-making and collaboration in real time.'

CDD Vault gives users an interconnected network view of compounds and biologics and experimental data across the ELN and registration system. Register every entity and reference how they are used in experiments so there is a deeper understanding and confidence in assay results. 'You never lose track of the relationship between your entities and experiments,' Wang pointed out.

Importantly, CDD works with other software providers to help establish the Vault as an integral part of a lab's informatics ecosystem, said Wang: 'Biotech companies, and especially the smaller ones, aren't likely to have a complete in-house infrastructure

> "Everyone is looking at the same, current data, which facilitates decision-making and collaboration in real time"

necessary to carry out the end-to-end discovery, optimisation and development as a seamless workflow. It's also unrealistic to expect any single vendor to do everything for them. No one software can do everything, so any drug discovery pipeline will likely involve working with a number of vendors to cover all the bases.'

CDD thus works with a range of partners to integrate the Vault with complementary software platforms. 'Our strengths are in storing, managing and mining data, and allowing organisations to collaborate easily and securely,' said Wang. 'But the drug discovery and development pipeline requires a whole raft of capabilities, and no one vendor can offer all of that, so we have established partnerships with multiple vendors, so users can integrate CDD with these other specialty applications.'

It's what Wang calls 'a best of breed' approach: 'You take the best in class for compound management, which is CDD Vault. You then make a connection, via the CDD API, with what we consider are the best platforms for, say, inventory management, lead optimisation, or analytics. If our customers choose to work with these providers, relevant data can be pulled directly out of Vault into these platforms, and then resulting data imported directly back into the Vault. It also means all of this new data is available, as it's generated, in the Vault.'

Partners in this ecosystem include Certara, SarVision, DataWarrior, Elixir, Knime, Schrodinger, PostEra, Microsoft (CDD Vault's ELN directly integrates with Microsoft Office), Titian and Optibrium.

Wang said: 'So while we don't have the resources to be everything for everyone, we integrate CDD Vault as part of an informatics environment with these key providers and their platforms. Our customers can leverage the technologies they need for their research and pipeline development, with CDD Vault at the centre of their data management.'

Another major challenge for drug discovery and biotech generally, is that generating data is hugely resource intensive, explained Matt Segall, CEO at Optibrium: 'Big pharma has dedicated departments for screening and will generate large amounts of data, but biotechs may have to be very selective about what they measure for which compounds, on a cost basis, particularly if they're outsourcing to contract research organisations.'

Even for big pharma, measuring all the properties and activities of interest for every compound of interest is cost prohibitive, and so there will inevitably be incomplete, or sparse data, he continued: 'Another challenge is that data is typically noisy. Biology is messy, and measurements may be subject to experimental variability. When experimental errors creep in, you can really waste a lot of time and resources either pursuing a hypothesis that turns out to be incorrect, based on incorrect data, or – perhaps catastrophically – discard a promising potential compound because of false negative results.'

The prospect of using AI to help make informed decisions on compounds is thus huge, Segall noted. 'There's a real appetite in biotech for AI. And again, AI isn't just a topic for big pharma, we have a wide range of organisations talking

> "We've developed products and platforms that bring AI within the reach of even niche biotechs"

to us about intelligent solutions for their specific challenges.' Optibrium is pioneering predictive modelling as an aid to decision analysis in drug discovery, and has developed AI-based platforms for small molecule design, optimisation and data analysis, and what it describes as 'augmented chemistry'. The company's Cerella platform harnesses a unique deep learning approach to help overcome limitations or gaps in drug discovery data, and ultimately reduces costs and speeds drug discovery cycles.

'We've developed products and platforms that bring AI within the reach of even niche biotechs,' Segall noted. 'Cerella effectively helps to highlight high-quality compounds with confidence, and prioritise compounds and experiments. The platform exemplifies how we can offer state-of-the-art, turnkey AI solutions that are intuitive, affordable, and that can have a significant impact on drug discovery timelines, decision-making, and ultimately, we hope success.'

Using deep learning imputation, Cerella looks at all of this very sparse

elenabsl/shutterstock

and very noisy and messy data that is generated experimentally, and essentially fills in all the blanks, to find otherwise unrealised opportunities. 'It's far more than you could even imagine doing with a conventional cheminformatics approach,' Segall stated.

'And it doesn't just "dump" all of that data onto a desktop,' Segall noted. 'Cerella is far more proactive. It can tell you if you have missed compounds that might fulfill specific sets of criteria – and highlight compounds with high probability that they're going to achieve your objectives. You can then validate those propositions experimentally. You get a much broader view of the molecules you are exploring, with biological and experimental context. This is something the cheminformatics space won't achieve … It's meeting chemistry and biology in the middle of their respective spaces.'

Traditional cheminformatics techniques are based on visualisation of data, and analyses, such as structure activity relationship analysis (SAR), to make sense of that data. This provides an understanding of the relationship between the structure of a compound and, say, its activity against a target. Cerella can relate that to the biological relationships between the different things being measured, and the overall outcome.

'And it brings all that together and understands all those relationships, both chemical and biological, to make much more accurate predictions, potentially even in the much broader context of previous projects, and other information that might be tucked away in a database that hasn't been looked at for years,' Segall said. 'It's a proactive approach that can help inform potentially new directions for projects. And that's where the promise of AI can leverage much more value from that data.'

> "We've demonstrated how Cerella can use in vitro ADME data to predict in vivo pharmacokinetics of a compound"

And for small biotech, which may only have a few projects in its pipeline, the ability to maximise intelligence from valuable screens at any stage is critical, Segall noted: 'Some of the biggest challenges these companies face is how to use the data they do have more effectively, to make decisions in the course of a project, and avoid those missed opportunities. AI can help in the decision-making process to give confidence that you are actually running the most valuable experiments, and the resulting data is going to add the most information to make those better decisions.'

Importantly, Cerella doesn't require what Segall describes as 'a bunch of expert Python programmers, huge libraries and a team of data scientists' to work with it: 'We've implemented Cerella as a cloud-based platform, so essentially, it plugs into your data source, acquires the data you give to it securely and safely, and cleans that data.' And while the task of data cleaning can otherwise be incredibly time consuming and tedious, 'Cerella does all that data cleaning, and then prepares the data for modelling', Segall continued. 'It builds and validates the models – of course, you can look at those results and carry out further validation – and then automatically fills in the blanks and makes this very, very rich data accessible in a very, very intuitive way.'

This means Cerella can be interrogated using simple questions – for example, to find compounds that may have activity against a particular target. It will also suggest compounds that may not have yet been tested in that assay.

Cerella doesn't just have utility at the level of screening, Segall pointed out: 'We can use the platform to predict in vivo response, either in preclinical in vivo models, or potentially in human clinical trials. Through a collaboration with AstraZeneca, for example, we've demonstrated how Cerella can use in vitro ADME data to predict in vivo pharmacokinetics of a compound. And that's an incredibly powerful capability. Or, the ability to help predict human safety outcomes using preclinical data, as another example.'

While Segall isn't suggesting throwing out a particular compound based purely on a prediction, 'even though it will be a higher quality prediction than conventional QSAR models', he acknowledged, 'what it does do is inform you there's a higher potential risk and informs what experiments you should do to mitigate that risk'.

A key part of Cerella is that the AI explains its reasoning. 'It's one thing to have a black box that spits out an answer and you have to sort of trust it, but actually understanding why it's made a prediction is really important,' Segall concluded. 'This means scientists can validate in their own minds whether a prediction makes sense, formulate hypotheses and derive new questions to answer or future avenues for experimentation.

2022 Analytical Data Management Survey

Effective data management and access has never been a bigger focus in scientific R&D than it is today.

We asked* those responsible for analytical data to share their experience with analytical data management.

>92% of scientists

- use multiple techniques
- collect data on numerous instruments
- process data using diverse software

Analytical data is managed and stored in a multitude of systems

80% Microsoft applications

70% Instrument software

34% In-house software

24% LIMS

13% ELNs

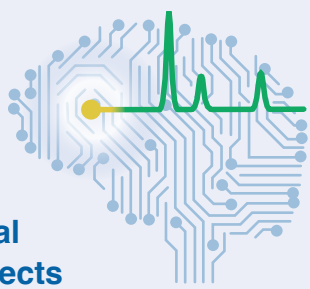
68% say data is hard to access and share

9/10 need analytical data on a daily basis to make decisions

~50%

feel that cloud-based data management is important

Only **6%** are using analytical data in AI /ML projects



70% want more investment in data management technologies

*A market research survey was conducted in 2022 in partnership with The Analytical Scientist. Survey participants: 41% academia/non-profit, 45% industry, 11% government, 3% other.

Learn more: www.acdlabs.com/ADMSurvey

ACD/Labs®

info@acdlabs.com ☎ 1-800-304-3988 (Toll free, USA & Canada) +44 (0) 1344 668030 (UK)

Managing change in genomics

Dr James McCafferty discusses how changes to the interface between science and technology are shaping future genomics research

How is genomics research changing?
Dr James McCafferty: You'll hear people talk about the move from wet lab to dry lab science. Wet lab is the folk with the test tubes and the white coats, and dry lab generally relates to IT, informatics and research software. The BBSRC, which is the main funder in this space for the UK, talks about a move from 80% wet lab, to 80% dry lab. That's the kind of transformation we're seeing in the sciences, particularly the biosciences.

This can massively accelerate research. There's all manner of very sophisticated lab instruments that are mining terabytes of data on the kind of things we study. We look at genetics itself, but we also look at the proteins, we look at the chromosomes, the operation of cells, and [we've also taken] quite a significant move towards imaging data as well. The data generated is massive.

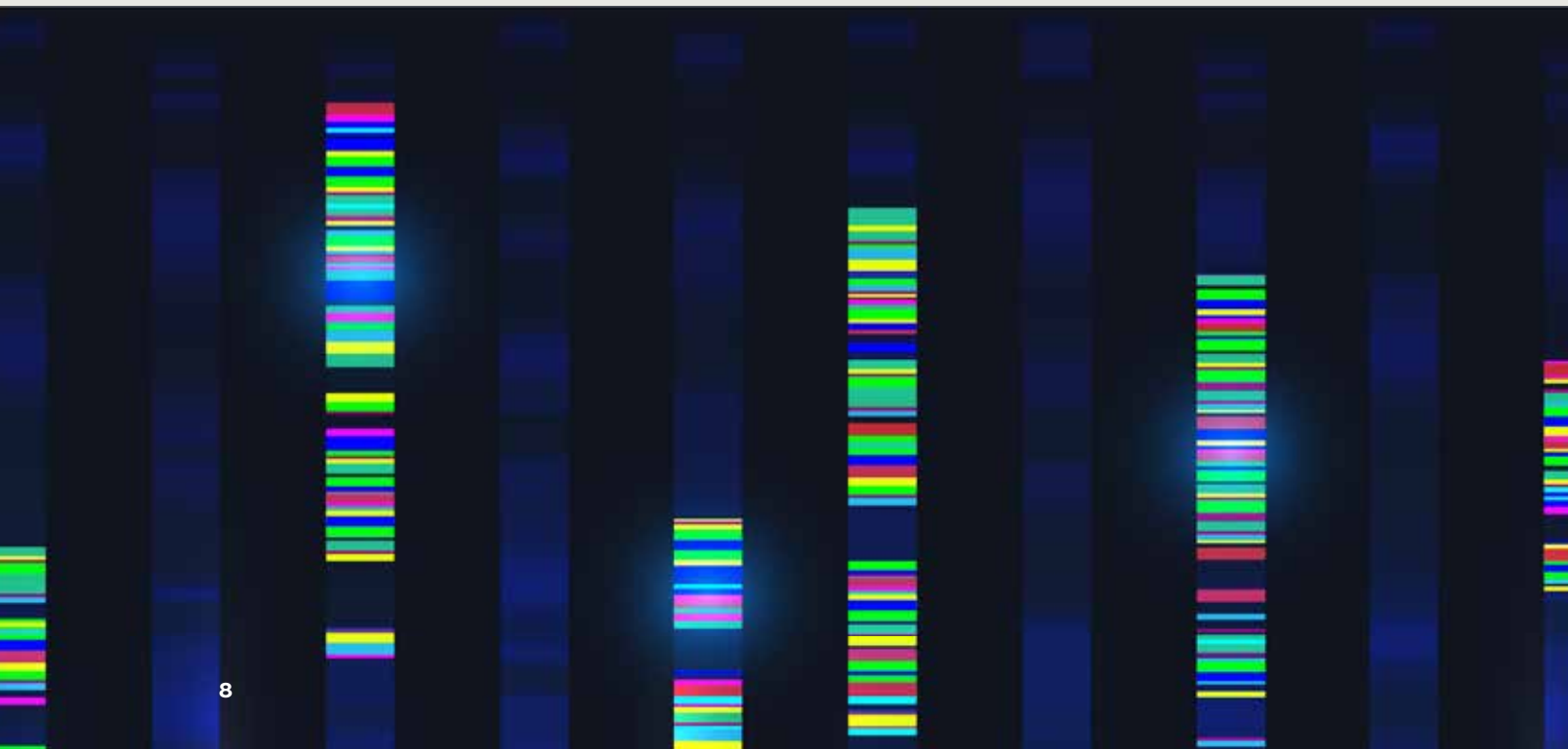
> "Part of my job is to make sure we're properly efficient with what we generate and what we store"

How is imaging data being used in genomics?
To give you an example, for the spatial data, if you consider a cell, that cell is within the context of a tissue, so it's got lots of other cells around it [which provides information]. And the way the cell behaves in that context [provides information too]. So if you're looking at, let's say, a cancerous tumour, you want to understand where the cell is, and where it is in relation to other cells.

In addition to that, by looking at what's happening inside the cell, so looking at its genomics, looking at the transcriptome – the proteins that the cell is generating – you can see what the cell is actually doing. If you capture that information, you can not only work out what type of cell it is, but what the cell is actually doing at any one time. For example, it could be growing; it could be dying; it could be splitting in two.

By combining the image, which is the cell in its context, and including into that the genomics and transcriptomics data, that yields a massive data source, allowing scientists to study things like cancerous tumours. But when you are dealing with image data, these are not small files.

How does this impact your ability to support scientists at Sanger?
This is changing the demands of your



IT systems. We have a massive data explosion here. We need to support huge amounts of storage and be able to interact with very large datasets. We need to pull that data out, manipulate it and put it back into storage.

For our workloads, it's about shifting data. In my previous role at UCL, the high performance computing (HPC) was geared up for astrophysics and materials science. And it's interesting to compare that kind of environment with the Sanger, because although we are nowhere near as big as UCL, we store more than double the amount of data they do for their research.

This is because our systems are geared up for data-intensive research. So rather than HPC, it's more like high throughput computing (HTC).

Just to give you a kind of sense of scale, we have about 90 petabytes of genomics data here in the Sanger, and our sister organisation in the same campus store about 10 times that amount. Admittedly, that is not just genomics data, they store lots of other stuff, but it gives you a sense of scale.

Part of my job is to make sure we're properly efficient with what we generate and what we store. But in addition to that, part of my job is about making sure we get the best value, the greatest insights and the highest scientific output from that data.

There's an increasing dependence on GPUs as well. That's largely driven by the adoption of machine learning and deep learning techniques in biosciences. So here in the Sanger, we have just recently invested quite heavily in Nvidia GPUs.

Do you see this demand continuing to grow? Do you think Sanger will invest further in GPU technology?

I think that's the way it's going. To be honest, there's always been a lot of machine learning-type applications in informatics, just by the very nature of what it does. So there's quite a lot of random forest-type or support vector machine applications. Neural networks, on the other hand, are much newer in the genomics world but are gaining traction steadily. I think we will see a lot more of that in the future.

Part of my role is not just about getting the kit there ready for people to use, it's helping to make sure they're getting the best out of it. So training and support, advice and things like that.

And going back to the point I made at the start [about almost] reshaping what we're doing for IT within the Sanger, now we are required to be more proactive with the science community. This means bringing new technologies and new tools to the table to help them with what they are trying to achieve. Having the best tools and the best IT means our researchers can be weeks and months ahead of other researchers, and also in some of the big scientific challenges.

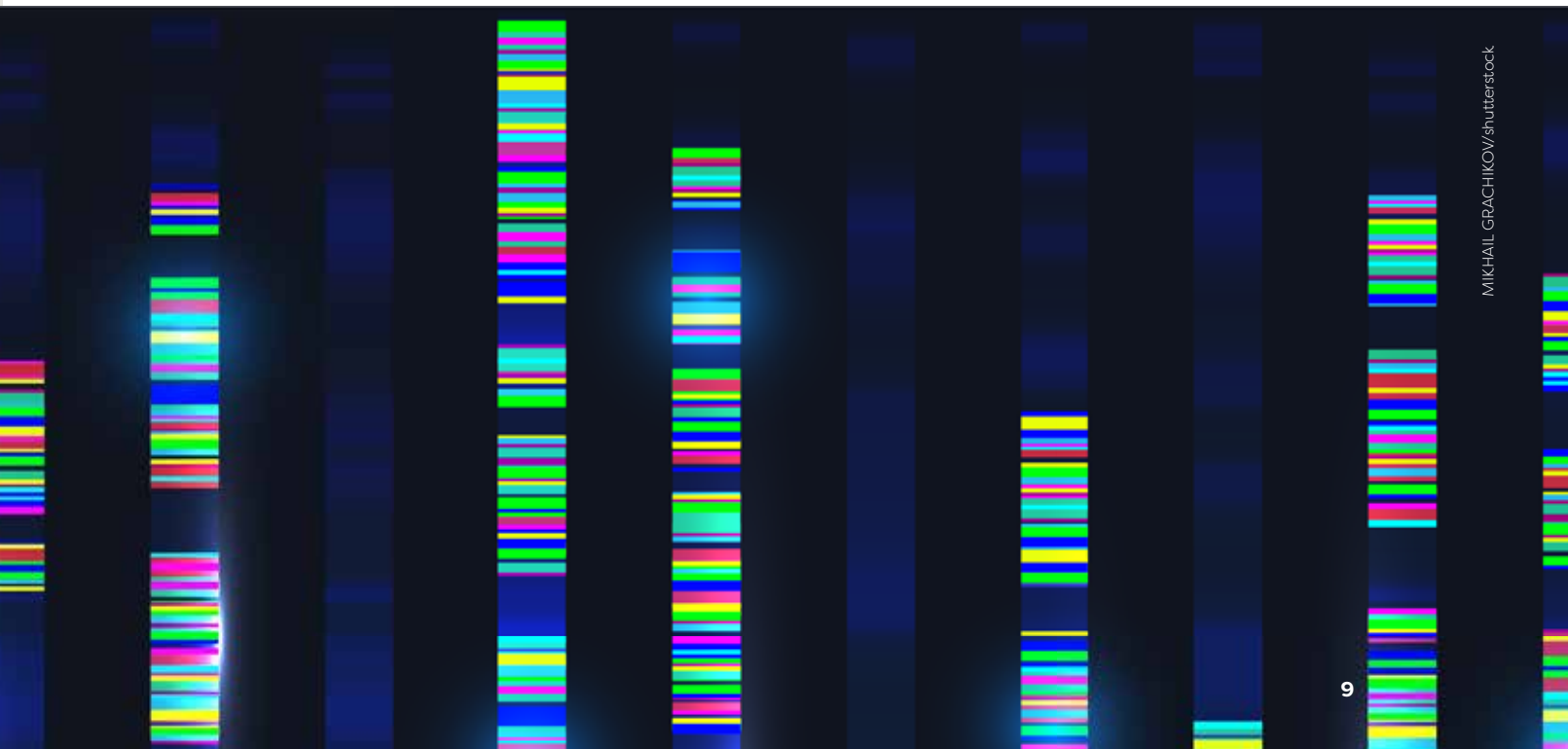
Sanger operates at scale. We've got the second-biggest sequencing fleet in the world for genomic sequencing. It's a massive operation. That means Sanger scientists can draw upon that when doing massive studies. Other research institutes still study one or two genes at a time, whereas we do entire genomes in one go.

Sanger is special and unique, but when you think about it, our data goes all the way from biological samples through to data and insights. Having a digital infrastructure that carries that all the way through is really important. One recent example, one of the things we're working on at the moment, is equipping our researchers so they have the digital tools to do what they need.

The hope is that our researchers will use Jupyter Notebooks for developing and trialling techniques. This is absolutely an IT artefact and it's exciting to see how that stitches into tools like ELN and LIMS to impact biological research.

That's absolutely the way of the future, because every single one of our scientists will be an informatician. They'll know how to write scripts, they will know how to write software and know how to do analytics. That is the direction of travel.

Dr James McCafferty is Chief Information Officer at the Wellcome Sanger Institute. His role encompasses IT strategy, delivery and operations to support the goals of the institute. This encompasses research IT, research data, research software/informatics, enterprise IT and information security. Dr McCafferty previously worked as Chief Information Officer and Director of Research IT, University College London.



MIKHAIL GRACHIKOV/shutterstock

Software advances speeding drug discovery

Gorodenkoff/shutterstock

Chemistry software providers play an important role in supporting industry and academia, writes Sophia Ktori

OpenEye Scientific develops and offers large-scale molecular modelling applications and toolkits primarily aimed at drug discovery and design. 'Although our tools are used by a wider industrial sector, from large pharma and small biotechs, through to agrochemicals and materials scientists,' commented Ashutosh Jogalekar, now Head of Product at OpenEye Scientific.

OpenEye Scientific was established some 25 years ago, to build on the founders' conviction that shape and electrostatics are two key factors responsible for molecular recognition, which, in turn, drives how drugs work, '… because most drugs are small molecules that interact with proteins', Jogalekar continued. 'So, being able to accurately model shape and electrostatics was really the foundational principle of the company.'

Cloud computing has been a significant driver of speed and scale as the need to work with huge numbers of compounds, and volumes of disparate data, has increased, Jogalekar continued. 'We now have access to almost unlimited processing power, and we can use platforms such as Amazon Web Services (AWS) to recruit hundreds of CPUs and GPUs, if required, to help us search through a billion compounds in just a matter of hours.' That's been really transformative, Jogalekar commented. 'Running our solutions on AWS means we can look at solving modelling queries and investigations that would previously have been impossible due to hardware or compute time limitations.'

A one-stop shop for drug discovery

OpenEye Scientific's flagship solution, Orion, a molecular design platform, is – as far as the company is aware – the only cloud-native, comprehensive molecular modelling and cheminformatics platform, which Jogalekar described as 'a one-stop shop where scientists can come and do all kinds of cheminformatics and modelling calculations on compounds, including large-scale virtual screening, calculation of physicochemical properties and molecular dynamics'.

Orion provides users with an integrated web-based solution for designing, calculating, viewing and analysing in the chemical space, and offers a dedicated platform for managing data and applications. This single platform negates the need to swap between apps, and also means data doesn't have to be transferred from one tool to another, but can remain in the Orion platform. 'Orion effectively wraps up all the other applications OpenEye Scientific has developed into a single solution, which includes tools for calculating shape and electrostatics,' Jogalekar continued, 'along with at least a dozen other applications that can be used to carry out all kinds of other cheminformatics protocols.'

Orion presents a set of workflows, which OpenEye Scientific has termed 'floes'. 'So, one floe will calculate key properties of compounds, such as their solubility, while another can search for compounds that are similar to an existing patented compound of interest. All of these floes are available, essentially at the simple click of a button. All the user has to do is log in to their account, upload critical inputs and go from there.'

Solving data integration and curation issues and providing the tools that will help to rapidly evolve machine learning and AI tools are, respectively, the main challenges and future-focused goal for the evolution of cheminformatics across fields, he suggested.

'People are going to want to visualise and analyse not just more data, but more diverse data from chemistry, biology and pharmacology, say, at the same time. And so here at OpenEye we are particularly interested in machine learning, because we have all these tiers of data available, and we want to see the correlations between them. That will make it easier to answer critical questions, such as how a virtual screening ties in with the end result of your downstream workflows. And from our perspective, while this is a challenge, if you have an integrated platform where all of the rigorously validated data is available – in the right format – then developing and applying those kinds of machine learning tools just becomes much easier.'

But at the same time, there is a real drive to consider technical complexity, and make the use of tools much simpler, so that additional complexity in the chemical space can be addressed, he suggested. 'The drug discovery space, for example, is moving beyond the traditional classes of small molecules into compounds that have completely novel modes of activity. These compounds present a different level of complexity, and gaining access into biological space will obviously be very significant.'

Open source chemistry resources, The European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), maintains a range of freely available cheminformatics resources that allows users to share data, and undertake and analyse the results of complex queries in the chemicals space.

'Probably the best known of these resources is ChEMBL, an open data resource of binding, functional and bioactivity data,' explained Andrew Leach, head of chemical biology and head of industry partnerships at EMBL-EBI. 'But our suite of resources also includes SureChEMBL, a searchable database that contains information extracted from patent documents, together with ChEBI, a dictionary of small chemical molecular entities, and UniChem, which gives users the ability to cross-reference chemical structures across different databases.

'Containing data on more than 2 million compounds and 1.4 million assays data', Leach commented, 'ChEMBL is widely considered to be the world's leading expertly curated resource of its type that is completely open, and can be used without restriction by anyone in the scientific community. Its utility ranges from basic searches for compounds that have specific properties and activity against particular targets, to the development of new AI algorithms and machine learning tools.' Launched more than 10 years ago, the ChEMBL database is now on its 29th release, and is derived from data in more than 80,000 published documents relevant to the life sciences sector. ChEBI includes information about the function, or role, of the compound in its biological context. Importantly, many external resources link to ChEBI due to its careful curation, the availability of a stable identifier for each entry, and its ontology, which can be combined with other ontologies to enable reasoning, Leach suggested.

> "Running our solutions on AWS means we can look at solving modelling queries and investigations that would previously have been impossible due to hardware or compute time limitations"

Additionally, EMBL-EBI provides a resource known as UniChem, which allows scientists to link chemical structures across different databases. This includes external resources that may not be focused on small molecules, but which contain some small-molecule information. This capability relies upon the ability to unambiguously define chemical structures using the International Chemical Identifier (InChI), which resolves the question of whether two compounds represented in different resources are the same molecule. 'By deriving the InChI, you can be confident that any compound with that InChI, wherever you find it, is the same. It's a very simple idea but also really quite powerful, which UniChem uses to easily allow users to navigate to data from different resources on the same compound.'

Challenges do still exist with respect to managing the diversity and volume of data being produced in the chemical space, Leach further pointed out. 'There are new data types being produced using the latest experimental

methods, so there is a continual challenge with respect to managing and integrating these different kinds of data.' So there will always be a need for expert curation, 'especially when it comes to interpreting information from multiple sources, which can be challenging to interpret or might be ambiguous'.

The small molecule sector can, nevertheless, perhaps learn from other biological data resources, Leach suggested. 'The biological community already deposits large volumes of data directly into the relevant databases. The type of data in ChEMBL is much more heterogeneous than (say) DNA sequence data; in addition, it is often generated in individual academic labs, which may not have the necessary informatics systems nor expertise to undertake the necessary processing and validating steps. Nevertheless, it is important for us to continually explore how we might address this challenge and indeed, we continue to work with a number of laboratories that directly deposit data into ChEMBL.'

EMBL-EBI is funded by its member state governments and other external funders, which include UK Research and Innovation (UKRI) and the UK's Research Councils, as well as the European Commission, the US National Institutes of Health and the Wellcome Trust. The Wellcome Trust has been the most significant funder of the ChEMBL family of databases, Leach noted. EMBL-EBI is also a key partner in large-scale collaborative projects, he continued. 'We are involved in a large public-private collaboration called EUbOPEN, which is funded by the Innovative Medicines Initiative (IMI). This project aims to create a publicly available chemogenomics dataset that will comprise a set of compounds representative of a wide diversity of bioactivities and also deliver at least 100 open-access chemical probes. The ChEMBL team is also involved in a number of other IMI collaborations, the NIH-funded Illuminating the Druggable Genome project, Open Targets and BioChemGraph, which aims to integrate data from ChEMBL, the Protein Data Bank in Europe, and the Cambridge Structural Database (CSD).

Increasing model reliability

Alvascience is a young cheminformatics software company established three years ago, which offers a suite of desktop tools designed to help streamline the entire Quantitative structure-activity relationship (QSAR)/ Quantitative structure-property relationships (QSPR) process, from data curation to the deployment of prediction models.

Cheminformatics and in silico techniques have been growing in focus for the past two decades, suggests Matteo Bertola, Alvascience co-founder and head of software development. 'Cheminformatics has traditionally been viewed as something of a niche field within the pharma/ biotech sector but, in fact, it's critical for many sectors of discovery and R&D,' he explained. 'Just about every agrochemical, food, oil and gas, pharma or materials science organisation has a cheminformatics department because these organisations all need to work with molecules, and try to understand the chemistry of foodstuffs and crop products, drugs, petro- and speciality chemicals, and new materials.'

Organisations today expect to work with software that allows them to screen huge numbers of molecules and create reliable models to test specific properties or evaluate a prediction, Bertola noted. 'They may commonly have some endpoint that was acquired experimentally, on which they want to build a mathematical model that allows them to test new molecules against that endpoint, and rule out or rule in molecules they may want to take to the next level.'

One of the main challenges, Bertola suggests, is 'how to create models that are reliable, and explain how the molecule will behave in the real world, so you can test structures with some degree of confidence.' Another challenge is how to explore the sheer size of chemical space that is available and find the best molecules that fit the required properties and, ultimately, functionality.

With these goals and challenges as starting points, Alvascience has built a suite of desktop QSAR and cheminformatics tools to aid the complete workflow. Data curation is often the first step of a QSAR pipeline, and the firm's alvaMolecule platform has been developed to allow users to analyse, visualise, curate and standardise a molecular dataset.

'In the next step, alvaDesc calculates molecular fingerprints and thousands of molecular descriptors in an efficient way,' Bertola continued. 'Molecular descriptors are also key components for the development of models to predict given endpoints, so we've developed alvaModel to enable users to generate QSAR/QSPR regression or classification models to predict the endpoint you need.

> "Orion effectively wraps up all the other applications that OpenEye Scientific has developed into a single solution"

'The software, making use of genetic algorithms, can search for high-performing models by selecting the descriptors from those previously calculated in alvaDesc. Once your models have been created, you can share them with your colleagues, who can use alvaRunner to apply the models to new molecular datasets. In this way, you do not need to use other software for applying models as alvaRunner provides a single solution.' alvaModel and alvaRunner can thus effectively be applied together to build and deploy QSAR/QSPR regression and classification models, with alvaRunner offered as a software tool that allows users to apply models, created using alvaModel, on a new set of molecules.

Alvascience also offers a tool for de novo molecular design, called alvaBuilder, which has been developed as a user-friendly software that lets users generate new molecules with a set of desired properties, starting from a defined training set. 'The suite of tools effectively addresses this concept of a cheminformatics pipeline, starting with data curation, and then getting to the deployment of a model,' added Bertola.

All Alvascience's tools are offered solely as desktop solutions, available for Windows, Linux and macOS. 'We've stayed away from the cloud as many customers don't want to share any of their data with third parties,' Bertola pointed out, 'but we are not ruling out possible expansion into cloud offerings in the future.'

CDD VAULT® Complexity Simplified

Discover more with CDD Vault

CDD Vault is a complete platform for drug discovery informatics, hosted through an intuitive web interface.

Helps your project team manage, analyze, and present chemical structures and biological assay data.

Smart Software Saves Time®

Electronic Lab Notebook
Capture experiments digitally

Archive and search all of your experiments with ease and collaborate securely.

Inventory
Keep track of compounds

Track and locate materials in your inventory whether in-house or shared externally.

Activity & Registration
Store & organize your data

Store, organize and analyze your data with ease. Share and collaborate securely across project teams.

Visualization
Plot data sets & mine them

Analyze large data assets to find interesting patterns, activity hotspots and outliers.

BioHarmony™ Drug Data Store
Structured data on *any* drug

Real-time semantic data streams on your drugs of interest. Standardized, up to date, and ready to use.

cddvault.com

Predicting infection

Eddie Cano-Gamez discusses how machine learning is helping researchers to classify patients with sepsis, leading to more effective, personalised treatments

Sepsis is a life-threatening condition caused by an uncontrolled response to infection. While advances in life support have reduced the mortality of this condition, researchers' understanding of sepsis is incomplete, and there is an urgent need for new treatment alternatives. Dr Cano Gamez's research focuses on using high-throughput data to build reliable methods for precision medicine in sepsis, with a focus on gene expression and how genes are turned on or off under different conditions in the immune system.

Why is it so important to develop a better understanding of sepsis?
Eddie Cano-Gamez: The problem has several intertwined issues. One is that sepsis is what you can call a syndromic disease. It's not defined based on anything you can specifically measure. It is a diagnosis by the doctor using a combination of several measurements in hospitals. [One of] the criteria is you have to have evidence of an infection. That could be where doctors have detected bacteria in your blood or a viral infection, or something tells them you possibly are infected with something. Then your organs start to fail. You can have respiratory failure, kidney failure, or a very sharp drop in your blood pressure. The symptoms are very non-specific.

That means the disease is not very homogeneous. There are a lot of different types of patients and types of infections that come under the same umbrella term of 'sepsis'. That is one of the reasons why people believe so many clinical trials have failed to identify new drugs, because it's possible that actually what we call sepsis is a combination of different – sometimes very different – groups of patients: some might respond to treatment, some might not. In fact, some might actually be harmed by the treatment. Having them all together in a single group can make it very difficult to determine whether a treatment is working.

How does your research help to solve this problem?
Researchers have known since 2016 that patients can be split into subgroups. We look at the variety of genes – in this case, you can look at between seven and 19 genes. We know we can use that information to classify patients. But until now, there wasn't a reproducible method to do that.

This was a big issue, because lots of other groups and people studying sepsis wanted to classify patients and know which group they fall into, or how much they are at risk. There wasn't an easy way to do that, because the study hadn't really been done with a predictive angle in mind. The original work was more like a discovery study, so the only way to find out was to re-analyse the patients and the data.

My study aimed to make it as straightforward as possible for people studying sepsis, or infections in general, to get a set of patients and immediately classify them into subgroups and get a sense of their risk.

I predicted two machine learning methods. One predicts the groups of patients, so it's either a healthy-looking person, so that's what we call group three, or it's a sepsis patient that can be either low risk, so that's group two, or high risk, which is group one. That's the first model, which predicts three outcomes, depending on which group of patients we are talking about. The second method is where we realised these two ends of the spectrum – the healthy versus severe sepsis – actually are not completely separate. They tend to form this progression. We took that to our advantage to derive a score for

> "My study aimed to make it as straightforward as possible for people studying sepsis or infections in general"

thinkhubstudio/shutterstock

risk. This score goes from zero to one. And if it's zero, it means you're basically healthy, or your immune system looks like a healthy immune system. And the closer the score gets to one, the more severely ill you look in terms of your immune function. I trained models to predict this.

In terms of the specific machine learning models we use, all of what we present are 'random forest' models. We took a set of patients and then built what looks like a flowchart or a decision tree. And we ask, for this patient, is the first gene on or off, and how active is it? Then we subdivide the patients. Then we go to the next gene, and so forth, for a series of decisions. We use that type of decision tree to predict the groups or to predict the score.

Why did you choose a random forest model for this research?

In the past, the research group used simpler models. These were linear regressions. And that worked well for a while, but the problem is that sometimes each of the genes we are measuring do not act in isolation. Sometimes, two or three of them might work together or correlate with each other. That is not picked up as easily by something as simple as a linear model, where each variable is assessed independently.

The advantage of these random forest models is they can look at interactions. If we have two genes that tend to be together, or that tend to be active or inactive most of the time, then based on these different decision branches, you can get that to capture these nonlinear interactions. That was my motivation for trying this type of model.

They definitely have improved our results compared to the linear models the lab used before. And I think the accuracy was good enough that we decided to go for that, but there's no reason why we couldn't have used other types of models that may be more complicated – for example, neural networks or support vector machines and other types of approaches. The random forests seem to work well on our end and I think it was suitable for the type of data we had. But it's certainly not the only possibility.

What were the main challenges in developing this new model?

The real challenge here was that the models would work really well in our dataset, and then we would apply them to another dataset, and they wouldn't necessarily work as well.

That was the motivation for integrating so many different datasets first into this training set. When the users want to classify the patients, the algorithm will first do that integration. It will take new samples and then align them or integrate them with our reference set, and then the prediction is done. So I think this preparation step, where the samples are aligned, is really the crucial one because it removes all of these differences in scale.

For that, I took a lot of inspiration from the field of single-cell biology. Single-cell biology is an exploding field that has grown over the last decade, where you can get detailed information separately for each cell. That means you generate really huge datasets. Nowadays, you can have in

> "That was the motivation for integrating so many different datasets first into this training set"

the order of a million data points, and for each data point, you have thousands of measurements. And so that real increase in the size of data in biology has led to a bunch of new data science methods that have been created for sorting all of those problems.

One big problem in 'single cell' is, if I have one experiment, and then another experiment, can I integrate them? So basically, I took inspiration from that as the same problem. But rather than integrating single cells here, I'm integrating patients from one hospital versus another. ●

Dr Eddie Cano-Gamez is a postdoctoral researcher studying the host immune response during sepsis at Wellcome Centre for Human Genetics. Dr Cano-Gamez has a background in immunogenomics, with a particular emphasis on the use of single-cell technologies to study cellular functions. He completed his PhD in Cambridge (2020), funded by a Gates Cambridge Scholarship and trained under Dr Gosia Trynka at the Wellcome Sanger Institute.

Genomic potential

Genomics software is helping to transform large volumes of unstructured data into actionable knowledge, with open-source tools and data architecture that is applicable for clinical genomics at scale, writes Sophia Ktori

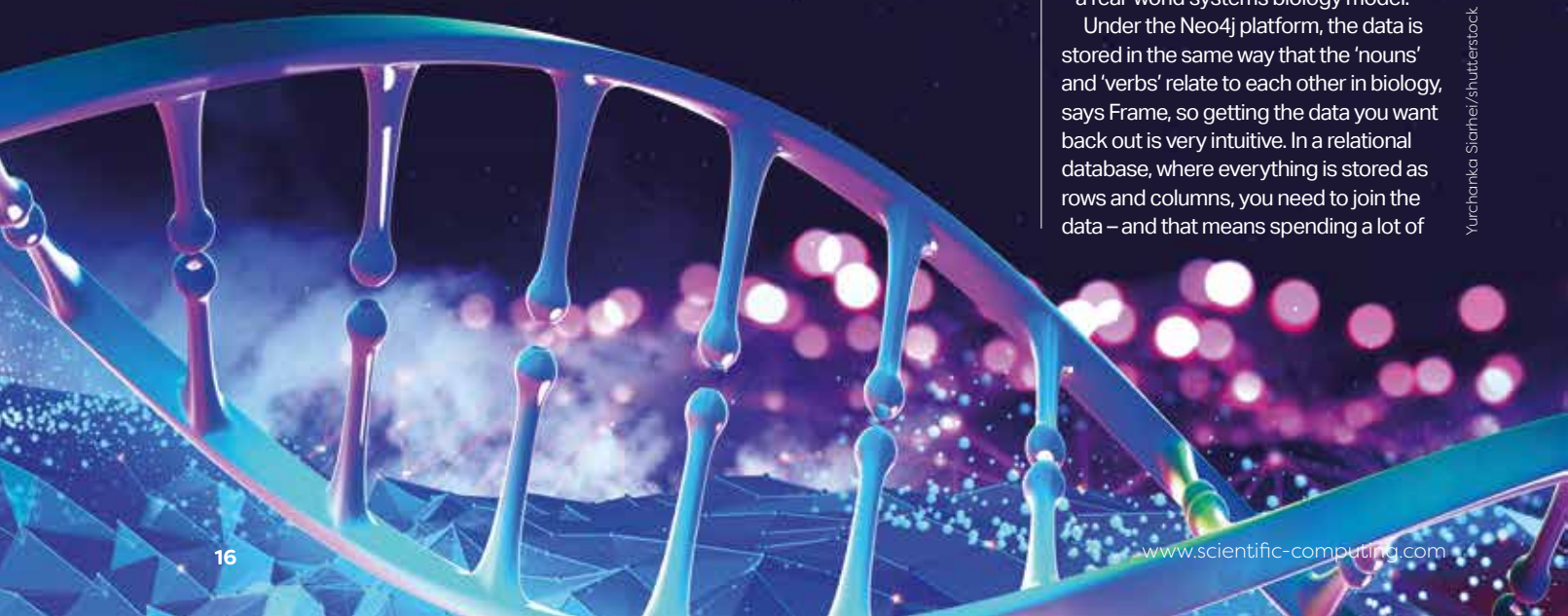
Today's DNA sequencing technologies now make it possible to sequence whole human genomes cost-effectively and with speed. Sequencing initiatives are generating vast volumes of data that – theoretically – give scientists a starting point to drill down into individual patient genomes in the hunt for disease-related variants, and also to mine collectives of huge public datasets to aid our understanding of the genetic basis of disease, unpick disease mechanisms, identify drug and diagnostic targets, and stratify patients for clinical trials and personalised medicine.

In practice, analysing this wealth of genomic data, in the context of associated biological and clinical data, is challenging. Gene variants identified through genotyping studies are stored in variant call format (VCF) files, but deriving patterns and insight from these files and connecting disparate data types isn't necessarily intuitive. And with relational datasets generated through large public and private initiatives (containing potentially millions of variants from many thousands of individuals) there are immediate issues associated with scale, as well as with how one can formulate the right queries.

In contrast with relational databases, graph databases can help to transform large-volume unstructured data into actionable knowledge, explains Alicia Frame, formerly Senior Director, Graph Data Science at Neo4j. 'In the case of genomic research, the key problem is how to integrate the large volumes of highly heterogeneous data and gain maximum insight,' she says. This is whether for diagnosis, personalised therapies or drug development, she is keen to stress. 'Graph databases are an ideal way to represent biomedical knowledge and offer the necessary flexibility to keep up with scientific progress. Using graph databases, a well-designed data model and query can deliver in seconds what previously took days of manual analysis.'

Graph platforms are effectively a way of representing and storing data as connected concepts, Frame explains. 'You can think of the graph as built on nodes that are concepts and then the relationships that connect them,' she says. 'In "everyday speak", we might well consider the nodes as nouns. So, in the genomics or bioinformatics space, these "nouns" are the genes, chemicals, diseases, variants and phenotypes. And then, of course, the relationships between them are effectively the verbs, which connect the concepts. It's – kind of – a real-world systems biology model.'

Under the Neo4j platform, the data is stored in the same way that the 'nouns' and 'verbs' relate to each other in biology, says Frame, so getting the data you want back out is very intuitive. In a relational database, where everything is stored as rows and columns, you need to join the data – and that means spending a lot of



Yurchanka Siarhei/shutterstock

> You can think of the graph as built on nodes that are concepts and then the relationships that connect them

time thinking about how the computer stores that data and trying to map how to connect it. Cypher lets a domain expert query far more naturally for patterns in the data. 'So the user can literally ask the database to find chemicals that bind to receptors for particular genes associated with a particular disease,' says Frame. This makes it very easy to effectively express a "mental model" and phrase the questions naturally, and retrieve the relevant information from the underlying database.

'If you've ever worked with a relational database, you have to typically join data across lots of tables,' she says. 'The more complex the query, the more complex it is to join the proverbial dots in the table. The more joins you have, the slower it is and the more difficult it is to write the query,' Frame acknowledges. 'Use a labelled property graph model based on nodes and relationships and there is no need to consider joins between tables, because the data is already joined.' It also becomes intuitive to add new data as it is derived.

Open-source and user-friendly

Graph databases also make it much easier to build applications for every end-user – think again, clinicians and researchers – and, at the back end, it becomes relatively easy for the person building the graph to maintain the resource, update it and deliver it to those end-users.

Neo4j has focused on making the open-source platform easily accessible and user-friendly for novices and smaller initiatives. 'For the community edition, we offer the database, plug-ins for data science and visualisation tools,' explains Frame. 'If you are a researcher or an individual, you can download our database and our software from our website for free. In fact, many groups start there.' The pivot point between the free, open-source version and the commercial enterprise platform will depend on the volume of data and the number of people who will be using the system, she adds. 'One of the primary

bluesroad/shutterstock

differences between the free community version and the enterprise system is parallelisation. The community platform will use up to four cores, whereas users of the enterprise platform can tap unlimited numbers of cores for faster computation when datasets are really huge and speed is important.'

In fact, many public genomic datasets are already encoded as graph databases. 'The NCBI, for example, has downloadable graph representations of many of its public databases,' Frame says. 'We also have a 'graphs for good' programme, through which we offer the commercial, enterprise software for free to non-profits, charities, researchers and academics for them to do their research; we also licence the database and the plug-ins to drug discovery companies, such as Novo Nordisk.'

The most obvious – although not the only – challenge associated with managing and analysing genomics data is its scale, comments Ignacio (Nacho) Medina, CTO of Zetta Genomics and founder of the open-source computational biology (OpenCB) platform. OpenCB is a bioinformatics suite designed to allow genotypic data management and analysis on a scale relevant to the massive sets of genome sequencing results that the research and clinical communities are generating. Medina describes the platform as a full stack open-source software solution, enabling large-scale genomic data storage, indexing, analysis and visualisation.

Scalable genomics research

The need for a dedicated, genomics-focused platform became increasingly evident to Medina more than a decade ago with the emergence of next-generation sequencing technologies and with the application of genotyping – not just for basic disease research, but also in clinical settings for potential applications in disease diagnosis and the development of personalised medicine.

As the first scalable solution enabling genotypes – recorded in variant call format – to be stored in a variant database, OpenCB is a high-performance solution for indexing and analysing many hundreds of thousands of samples, he believes.

Medina, who has been Head of the Computational Biology lab on the HPC team at the University of Cambridge since 2015, conceptualised and founded the OpenCB project while working in Spain during 2012. Within a few years, the platform was gaining the attention of some major genomics research initiatives. 'At first, it was just a prototype – very small – but this was enough to raise the attention of EBI, the University of Cambridge and Genomics England in

> The community platform will use up to four cores, whereas users of the enterprise platform can tap unlimited numbers of cores for faster computation when datasets are really huge and speed is important

2015, which adopted and contributed significantly to its development,' he says. During this period, Medina remained the platform's architect and has led the design and development of OpenCB. 'Today, OpenCB also includes a metadata and clinical database, fine-grained security management and a knowledge database, representing a complete genome data interpretation platform,' Medina notes.

As an open-source platform, OpenCB is accessible and free-of-charge for any organisation looking to manage and analyse genomics data in a non-regulated setting. In 2019, Medina spun Zetta Genomics out of Genomics England and the University of Cambridge to commercialise the OpenCB technology as XetaBase – a regulated, clinically validated and technically supported data architecture and software solution that is applicable for clinical genomics data management and evaluation at large scale.

'Zetta Genomics is, effectively, the commercial venture established to extend the scope of OpenCB, and XetaBase – OpenCB's commercial name – was created and launched in 2020,' says Medina. 'XetaBase is now becoming a certified platform that meets the regulatory requirements for data in clinical settings, while also addressing the need for customer support and implementation skills "built-in". It's offered as a software and through a service model, so we provide updates, fixes and training, along with ongoing support.'

Medina remains the CTO of Zetta Genomics, which is now also the main contributor to OpenCB. In June 2021, Zetta won £2.5m in VC seed funding. This investment is being focused on growth, improving performance, stability and implementing new analysis. Some is also enhancing the company's partnership network while it expands from the UK to open both Spanish and US offices. Resource is also being focused on talent; securing additional team members with software, development and commercialisation expertise. Importantly, the OpenCB and XetaBase data architecture supports regulatory governance for clinical and genomic data management, including NHS digital security and privacy policies.

'Regulatory and security issues aside, clinical labs face particular challenges with respect to how you deal with patients' genotyping test data,' Medina explains. These challenges relate to the sheer numbers of tests that are performed and the volumes of data generated but, also, the almost inevitable shortfall in human resources to analyse all the data for each patient in the hunt for a gene variant that might be the pathogenic cause of a disease.

Another challenge the OpenCB platform and XetaBase address is one of data sharing between scientists. Typically, if a clinician identifies a new disease-related variant that explains pathogenesis and disease symptoms, that finding may stay buried in the clinician's notes.

'In some cases they can submit or publish their findings but, even if that happens, it can take as long as 12 to 18 months for peer review and publication,' says Medina. 'Clinicians really need to be able to share their findings – with all of the patient data-related regulations in place – across hospitals. With the new federation feature, XetaBase will finally address that need to make findings available within minutes, not months.'

XetaBase is cloud-hosted and this simplifies data management and scalability, with a huge emphasis on making data secure and, effectively, available in real time.

'You may have several gigabytes of genotypic and other contextual data and metadata per patient,' explains Medina. 'The server for our platforms runs in the cloud and so this fact allows customers to easily scale to their needs, [supporting] tens or hundreds of thousands of patients in some cases, while we take care of and provide all the services that they need for the platform.'

Importantly, the OpenCB platform is built on a fileless infrastructure. 'Other solutions rely on a file-based system, but then how can you easily search across, say, 20,000 files to look for a disease-related variant that matches that of your patient?' he asks. 'In OpenCB, in contrast, all of the genomic variant data is aggregated in one indexed database. The largest example we have is Genomics England – for which there are about 140,000 whole genomes in one single installation, accounting for about 300 terabytes of data,' says Medina. 'And this fileless system means that, despite this massive volume and breadth of data, we can scan the whole database within minutes or scan any patient or the entire family in a few seconds.'

In fact, the OpenCB architecture makes it possible to include hundreds of different pieces of information relevant to each genetic variant and still query the whole platform. 'One analogy we can use to help explain this is Google,' says Medina. 'When you search for something on Google, the system doesn't search through the one trillion pages of content individually. Rather, Google has every page indexed so, when you query Google, you query that index and it takes just milliseconds.

'We have done something similar with OpenCB. We take all of the billions of mutations from large datasets and put them into one index on the system to enable incredibly fast analyses.'

And, of course, this is critically relevant, whether the query is for insight into one patient, such as searching for patients with the same mutation, or for the disease researcher who might be querying different variants across all of the different samples, Medina adds.

The ultimate vision is for a platform such as OpenCB and XetaBase to help reduce drug development times, increase the speed of disease diagnosis and aid decision-making for patients.

'My goal for the next five years is to demonstrate we can have a significant impact on research and healthcare, and help to reduce drug development times by potentially years,' says Medina. 'We also want to enable researchers to communicate their findings in a secure way, so they can re-analyse data and ensure no patient is forgotten.'

ILES - LAB INFORMATICS CLOUD PLATFORM

IVENTION LAB AUTOMATION

- Integrated LIMS, LES & ELN Platform
- Cloud and web-based software
- Continuous Delivery of Updates

CONTACT US FOR A
FREE PRODUCT DEMONSTRATION

iVENTION
Lab Automation

sales@ivention.nl • +31 38 452 83 75

ivention.nl

Finding the right laboratory software

With many competing products on the market, how can scientists and researchers find the right laboratory software, asks **Robert Roe**

Laboratory Information Management (LIMs) and Electronic Laboratory Notebook (ELN) software provide the backbone of electronic data capture and, increasingly, data management in modern laboratories. But choosing the right software for your laboratory is not straightforward. Software requirements can change based on a wide array of factors, including the organisation's workflows, scale, budget, regulation and compliance, and interoperability with other software and hardware currently used in the organisation.

While these software packages will have specific specialisations, most offer integration with instruments, cloud or web-based deployment methods, and customisability or configurable settings to tailor the software to your organisational needs. Beyond basic distinctions, such as discovery-based research or a focus on compliance and regulated industries, separating software packages can be difficult.

In a recent White Paper, *A complete guide to LIMS selection*, Thermo Fisher Scientific created a guide to help scientific organisations better understand the options available and select the right software package. The White Paper outlines three key points that any LIMS should deliver:

- Little-to-no customisation to support the existing lab workflow
- Be an easy-to-use system that can be implemented and qualified within a reasonable timeframe
- Address the key challenges of each industry, including internal and external regulatory requirements.

Ultimately, any LIMS or ELN system needs to be able to capture, store and manage laboratory data, but increasingly, the requirements are now growing to include the integration of instruments and other systems and support for secondary use cases of data, such as AI/ML, increased reproducibility and the inclusion of metadata.

These requirements are becoming increasingly important as scientific organisations try to realise the full value of laboratory data. Data is no longer used for a single experiment and then stored for potential regulatory proof. Now it needs to be accessible, transformable and usable for a much longer timeframe.

The demand for accessible data is a pre-requisite for AI/ML or big data applications where scientists might need to combine data sources from other areas in the business, collaborators or even public health sources. This can be neccessary to build more robust models, or to make use of large datasets for AI training. These potential use cases are not possible without easy access to laboratory data and the ability to clean or normalise datasets so they can be shared by partners.

An integrated laboratory

Integration of laboratory instruments to connect data streams straight into LIMS or ELN software is one of the laboratory's first steps in digital transformation.

Connecting instruments directly with LIMS or ELN and supporting that with a data lake – centralised data storage that allows scientists to store structured and unstructured data – helps to support streamlined processes and increased efficiency.

Analytics, data sharing and collaboration, and AI/ML initiatives can all benefit from the connected, integrated approach to integrating data sources with data management tools.

However, there are still questions about which approach to standardising data is right for a particular use case. Some organisations support standard initiatives such as Analytical Information Markup Language (AnIML), The Allotrope Framework, or FAIR data principles, which are all options for data standardisation.

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*. The principles emphasise machine-actionability (the capacity of computational systems to find, access, interoperate and reuse data with minimal human intervention).

AnIML is an open data format standard for chemistry and biological data. This is similar in some ways to Allotrope, which is also a data format standard that targets raw data, results and evidence across laboratory data.

When approaching the topic of integration, it may be obvious that an organisation may want to consolidate and standardise data, but choosing the specific format and standards can be much more complex.

Thermo's recent White Paper outlines their strategy to support the integration of instruments and other data sources in a unified data management strategy. It states: "Thermo Fisher has partnered with leading instrument and system vendors, as well as leading data standards organisations – such as the Allotrope Foundation and the Pistoia Alliance – to support a common language for scientific collaboration, FAIR data principles, and thus, facilitate data sharing and automation."

The document also shares insight into the benefits of choosing to manage laboratory data in this manner: "A LIMS, combined with data visualisation capabilities, such as dashboards to display information from your laboratory (for example, key performance indicators (KPI'S)), provides significant insight and benefits. Connecting LIMS to elements such as instruments and equipment and other enterprise systems offers a far more holistic view of operations and supports data integrity and compliance."

There are other strategies for lab connectivity. Scitara, for example, has its own Scientific Integration Platform (SIP), which it recently announced would be integrated with Agilent's laboratory informatics software, including chromatography software and lab workflow management solutions.

This new solution, created from the partnership between Agilent and Scitara, aims to increase data mobility

and provide a platform to support digital transformation across the laboratory.

The partnership will enable bi-directional communication with SLIMS, initiating lab integration workflows that will facilitate data exchange from, and back into, SLIMS using Scitara's Digital Lab Exchange DLX, providing universal connectivity within a GxP-compliant environment.

However, there will always be an argument about open standards vs standards or formats developed by a single vendor. While many vendor-specific platforms may aim to be technology agnostic, if they should stop supporting their framework at some point in the future, that could cause a setback in an organisation's path towards digital transformation.

Ultimately, the choice may differ for each organisation based on many factors, including existing experience, cost, functionality, ease of set-up and integration with legacy systems.

Deployment models

Cloud-based software is becoming increasingly popular across the LIMS and ELN markets – particularly in the new offerings from software providers in these markets. There are, however, many laboratories that have not adopted cloud computing for their lab management software. But to explore new solutions today would largely involve products that are either cloud-based, or where cloud is one possible solution. Several new and established players offer the cloud as an option, or deliver their entire platform as a cloud-native solution.

Earlier in 2022, TetraScience announced a five-year, $500m investment in developing and delivering a cloud-native, open and purpose-built scientific data cloud that lays the foundation for AI and ML research. The Tetra Data Platform (TDP) was expanded to include manufacturing and quality control (QC) data.

These scientific applications aim to reduce time to value by addressing the challenges of specific scientific lab operations and workflows. The platform ingests raw/primary life science data from thousands of sources, engineers them, extracts metadata, harmonises content, and publishes them to the cloud in a vendor-agnostic format.

These systems all share the concept of providing seamless data access and connectivity and easily configurable routine and complex workflows in a scientific laboratory. But the underlying mechanisms, technology and deployment methods have significant differences. This enables scientists to be more efficient and leverage data analytics, AI/ML techniques and decision-making tools.

Thermo Fisher Scientific, for example, offers its LIMS and ELN solutions in the cloud, either managed by Thermo or by the users. But the systems can also be deployed 'on-premises' if the user wants to make use of that option. This more flexible model allows the organisation to break up that initial investment in hardware and IT personnel, allowing them to scale up and test the cloud as needed in their organisation. Equally, those labs that are not ready to adopt cloud technology can still make use of the software through an on-premise installation.

Reference:
http://www.nature.com/articles/sdata201618

Open automation

Sophia Ktori discusses the importance of integration and open systems in supporting laboratory automation

Automation, robotics and digital transformation will be critical to the journey towards Lab 4.0. This vision of seamless, hands-off routine lab tasks, experiments and data handling is motivating investment in hardware and software that will increase lab efficiency, flexibility and throughput, while reducing costs and failures.

However, achieving integration of all of this lab equipment, and getting systems to communicate with each other remains a struggle. Even highly automated labs are today faced with managing islands of language-barriered hardware, and carrying out workflows that are interrupted by the requirement for error-prone, time-consuming manual tasks.

At a foundational level, this punctuation results, in part, because vendors of lab equipment have traditionally paired their systems with proprietary software that was not designed to talk to that of other suppliers, suggested Pantea Razzaghi, head of design at Automata. And this has been "oftentimes intentional", she suggested. 'Some of the larger players had a monopoly within the industry,' and it may not have been in their interest to make instrumentation that communicated easily with systems outside of their brand. Ease of integration would smooth the way for customers to switch to competing systems at upgrade, or when expanding or diversifying their labs.

This lack of interconnectivity means labs often have to maintain software that doesn't fit with the evolving lab environment, and retain equipment that generates data requiring manual housekeeping for downstream utility, Razzaghi pointed out. Labs may even decide to sideline equipment that works perfectly well and does a great job, because it remains disconnected from the lab set-up.

Whatever the outcome, this disconnect is likely to be costly, time-consuming and result in interrupted workflows and the need for repetitive manual tasks. It's also likely the format of the data "doesn't fit well" with that required by the next instrument using that data. You may then have silos of data that are not standardised or optimised, and so there's no way to maximise its utility, she noted. 'We may then need a middle layer of translation before that data can be used to its full benefit.'

Think about at which point lab functionality is most reliant on human intervention, and somewhere near the top of the list will possibly be the requirement to manually pull data out of devices and transfer that data to the next stage. 'It's almost comical how manual this process commonly still is,' Razzaghi commented. 'We see people literally walking up to an instrument, inserting a USB stick, downloading the data, walking over to a computer and uploading it into that system.'

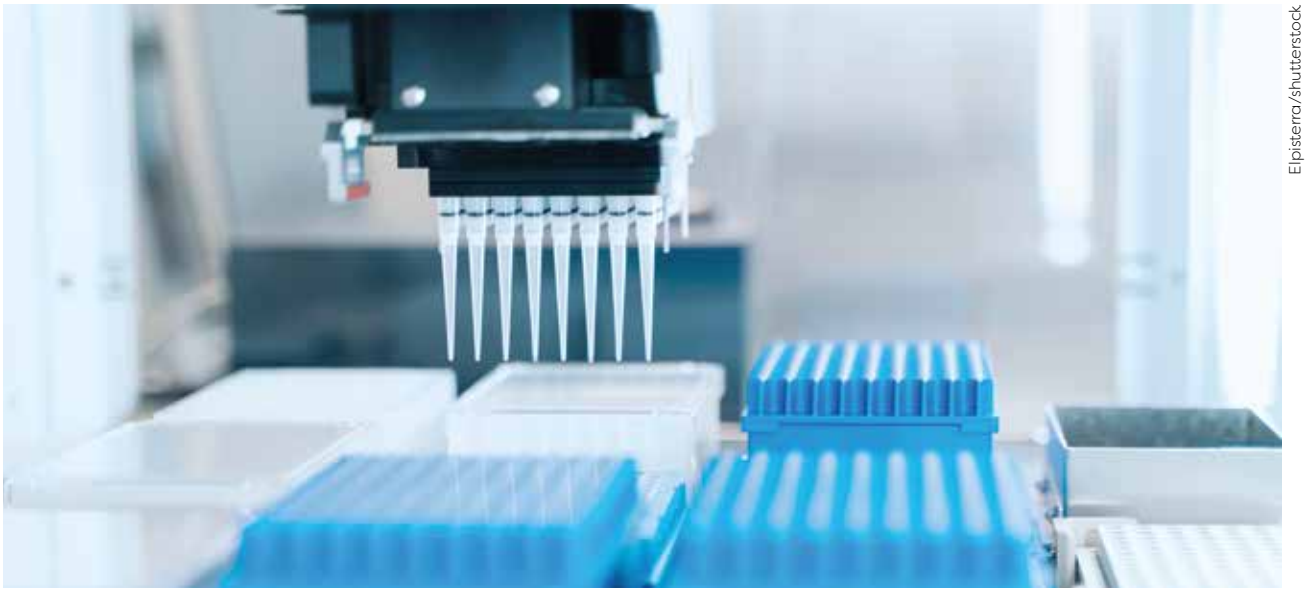
The lab today thus fosters equipment that has a level of "intelligence", let's say, that is analogous to that of the early era of digital cameras, she further suggested. 'To use these early digital cameras we had to insert a memory card into the camera, take the photo, then take out the memory card, put it into a reader, connect it to a computer, pull the images out of it, and store them on that computer. But today we can just snap a photo on a smartphone and send it directly to someone else, wirelessly and in an instant… ' It's this sort of ability that we need to bring into the lab space. 'It's not just about making scientists more efficient, but removing punctuation in processes and the requirement for manual data input, retrieval and transfer will save scientists from having to engage in multiple, repeated manual steps as part of everyday experiments.'

> "Our aim is to really help labs leverage the key instrumentation they already have, as well as implement new robotics hardware, through software"

Progressive changes

Fortunately, the philosophy in the vendor space is changing, Razzaghi suggested. An increasing number of what she described as "more progressive" companies are developing systems designed with an open architecture that can more easily be configured to interconnect and communicate. Vendors are also recognising that culture and expectations are changing within labs themselves.

Scientists and lab technicians are increasingly becoming more interested in engaging with the different layers of a system's software, to help it "play together". As she explained: 'Automation scientists may come from a scientific background or an engineering background, but today are interested in extending the utility of how they relate with a device. And that means they're actively looking for new ways to modify a system – whether that's through drivers or APIs – to orchestrate different instruments to



Elpisterra/shutterstock

connect and communicate together.'

This cultural change is also driving a shift in expectations. 'Similar to consumer markets in other industries, the move within the lab sector is to diversify options on the market, and particularly to give users far greater flexibility,' she said. This means that suppliers and developers are evolving their own mindset. 'They're realising that to survive in this space they have to make sure they offer this flexibility – think open API – and develop systems that offer a set of drivers, or advanced tools that give more advanced users within a lab space the option to interact with that software,' she added.

This interaction may be the responsibility of the organisation's automation engineer, or automation scientist – 'who may be few in number and in great demand,' Razzaghi noted. 'These are the people who the lab will call on when they want to scale up an assay or experiment, or transition from manual tasks to a partially or even fully automated task. But even with open APIs and inbuilt tools, there's still a great deal of work that needs to happen to enable that progress to automation.'

Not every organisation will have its own library of drivers, or automation tools, so even with a more open system, enabling that connectivity and communication device to device can be a major task. 'It''s still very early in the process for most labs, and it's going to take time for them to have a robust, reusable library of software tools – a fact which itself opens up another interesting question,' she pointed out. 'Should every lab have to do that?

> "The lab space is offering great opportunities for designers to help make the world a better place"

Should each lab have to develop its own library of tools and drivers to enable that lab integration?'

As consumers we now expect our technology to be plug and play, and to work with whatever else we've got on our home or office networks. We no longer have to download drivers or other integration tools when we set new systems up. So why has the pharma industry lagged behind? It's partly down to the already mentioned complexity of the lab environment, and also the diversity of automation and robotics systems now available in the lab sector, Razzaghi commented. And while 'democratisation is now happening within this space', these closed systems are still commonplace in labs.

Thinking to the future, system developers in both R&D and the manufacturing space realise the imperative to reduce the risk of market entry, and that making systems more "amenable" to integration will help to attract potential customers. Interestingly, Razzaghi said the lab is becoming a much more stimulating space from the perspective of user experience and interface designers. 'Whereas there has historically been a huge focus on developing consumer-oriented tools in fields such as gaming design or application design, the lab space is offering great opportunities for designers to help make the world a better place.'

Coming back to that cultural shift within the lab, we can also see that, with scientists today having a far greater understanding of how software works, the imperative is again there for vendors to open up their system's configurability. 'Compared with scientists who were graduating 10-15 years ago, scientists these days are far more knowledgeable about software tools, and how to configure them. Their personal toolset is very different. It's far more common for scientists who graduate today to be Python-savvy, for example.'

Scientists want to demonstrate more value out of their workflow, and to use their time more productively, formulating new projects or writing up papers, for example, rather than having to spend time doing repetitive, manual tasks. 'So, if they can access a tool that has an open API, they are more likely to try to work it to get the system to do what they want.' This is also generally a more cost-effective option than having to hire someone in, and will also likely be much faster with respect to upstream and downstream connectivity, because the scientists are the ones who know how the lab functions, and what is required to optimise that functionality.

Further challenges

However, Razzaghi acknowledged, the caveat to all this is that it's never a case

Gorodenkoff/shutterstock

of walking into a lab and immediately being able to undertake a single automation project that will connect everything. And that gap between expectation and reality can, in itself, present as a significant problem when labs are looking to undertake some sort of digital transformation or automation exercise. 'A lab may, for example, have a manual protocol or experimental workflow they want to translate to an automated format. What they may not realise is that transformation may not be possible as a single step. Here at Automata, we understand this gap between expectation and reality, and so we work hard to educate – as well as provide the software solutions – to get that integration in place.'

Part of Razzaghi's role is also to teach scientists how to 'think in an automated manner', she explained. 'When you are doing something manually it's a linear process. You may have restrictions, for example, waiting for a thermal cycler to finish before you can move to the next step. Or you may only have one liquid handler and it can only be used for a certain task. But labs going through an automation transformation may be able to achieve greater parallelisation of tasks.' It can thus be possible to save hours of work by optimising processes, and this is reliant on the communication channels and integration between instrumentation.

Factor in tools such as scheduling software and the lab then becomes even more efficient. 'You can then schedule your lab resources and leverage applications to calculate your workforce and instrument capacity.' Supporting automation and interconnectivity thus makes it possible to adapt workflows and processes to maximise efficiency. 'It gives you a way to adapt that journey map to understand how, by moving different steps, or blocks, around, you're going to achieve a certain task or workflow more efficiently, whether that efficiency is how much of a reagent is used, the time it takes for the experiment to complete, or when and how many of the lab instruments are required to complete that task.' Then, of course, it may be possible to calculate the cost benefits associated with each alternative iteration of that task or workflow. 'And that's one of the greatest bits of value we can bring to the table,' Razzaghi stated. 'Our aim is to really help labs leverage the key instrumentation they already have, as well as implement new robotics hardware, through software.'

> "If they can access a tool that has an open API, they are more likely to try to work it to get the system to do what they want"

As well as offering both hardware and software to aid lab integration and automation, Automata has the industry insight and expertise that is helping labs make a smoother transition. 'So, we provide the robotic lab bench system that has different actuation devices. This can be thought of as the device framework. Then to that robotic platform customers can integrate their own devices, or we can help them through the process of putting together and purchasing a bundle of instrumentation, and then leverage software so they can write the workflow protocols and do the day-to-day runs and data collection.'

Automata partners with the scientists running the experiments, so that everyone in the lab and other stakeholders understand how systems and software work together, and what they are capable of, Razzaghi noted. 'Importantly, the lab bench system is very much vendor agnostic, so it can fit with a variety of different instruments, and this really helps to connect them together. We have an existing library of drivers and can develop custom drivers to enable that integration.

'Our software also offers a workflow design tool, which makes it possible to interact with the protocols developed for each individual instrument, and facilitate communication so you can program flexibility into your procedures,' Razzaghi continued. 'It's then possible to run the workflow through a simulator to help identify where there may be errors, and help to optimise experiments and workflows to make the most of time and generate the best quality outcome.'

So, how might these developments impact on lab function within the next few years? Razzaghi suggested: 'In five to 10 years, we can imagine labs no longer having to rely heavily on service integrators for a bespoke solution, instead [having] the ability to independently build out their own automation platform using Automata and cutting-edge instruments from the vendor of their choice.'

Informatics for Your Lab – LIMS • ELN • LES • SDMS • AI

LabVantage®

Your guide for the digital transformation journey

Successful digital transformation – of your lab or your entire organization – demands an expert guide. LabVantage Solutions is that guide, keeping you focused on the True North of your business transformation journey.

We've combined the most modern laboratory informatics platform with expert services to reimagine digital strategies in biopharmaceuticals, consumer packaged goods, chemicals, food and beverage, healthcare and more.

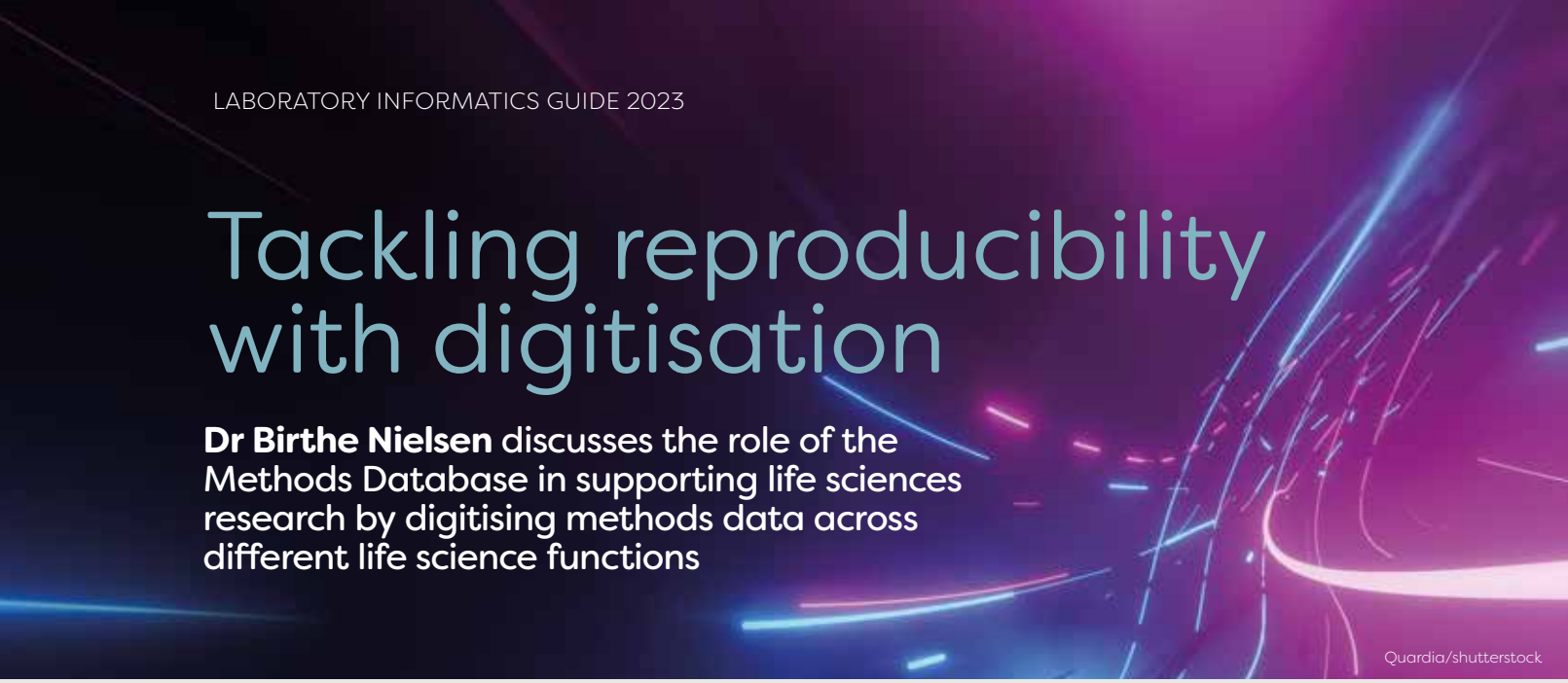
Discover why LabVantage is the platform of choice for digital transformations.

LabVantage. Leading laboratory digital transformation.

LabVantage.com

Tackling reproducibility with digitisation

Dr Birthe Nielsen discusses the role of the Methods Database in supporting life sciences research by digitising methods data across different life science functions



Quardia/shutterstock

Reproducibility of experiment findings and data interoperability are two of the major barriers facing life sciences R&D today. Independently verifying findings by re-creating experiments and generating the same results is fundamental to progressing research to the next stage in its lifecycle – be it advancing a drug to clinical development or a product to market. Yet, in the field of biology alone, one study found that 70 per cent of researchers are unable to reproduce the findings of other scientists and 60 per cent are unable to reproduce their own findings.

This causes delays to the R&D process throughout the life sciences ecosystem. For example, biopharmaceutical companies often use an external Contract Research Organisation (CROs) to conduct clinical studies. Without a centralised repository to provide consistent access, analytical methods are often shared with CROs via email or even by physical documents, and not in a standard format but using an inconsistent terminology. This leads to unnecessary variability and several versions of the same analytical protocol. This makes it very challenging for a CRO to re-establish and revalidate methods without a labour-intensive process that is open to human interpretation and thus error.

To tackle issues like this, the Pistoia Alliance launched the Methods Hub project. The project aims to overcome the issue of reproducibility by digitising methods data across different life science functions, and ensuring data is FAIR (Findable, Accessible, Interoperable, Reusable) from the point of creation. This will enable seamless and secure sharing within the R&D ecosystem, reduce experiment duplication, standardise formatting to make data machine-readable and increase reproducibility and efficiency. Robust data management is also the building block for machine learning and the stepping-stone to realising the benefits of AI.

A FAIRer future for methods data

Digitisation of paper-based processes increases the efficiency and quality of methods data management. But it goes beyond manually keying in method parameters on a computer or using an Electronic Lab Notebook; A digital and automated workflow increases efficiency, instrument usages and productivity. Applying a shared data standards ensures consistency and interoperability in addition to fast and secure transfer, of information between stakeholders.

One area that organisations need to address to comply with FAIR principles, and a key area in which the Methods Hub project helps, is how analytical methods are shared. This includes replacing free-text data capture with a common data model and standardised ontologies. For example, in a High-Performance Liquid Chromatography (HPLC) experiment, rather than manually typing out the analytical parameters (pump flow, injection volume, column temperature and so on), the scientist will simply download a method that will automatically populate the execution parameters in any given Chromatographic Data System (CSD). This not only saves time during data entry, but the common format eliminates room for human interpretation or error.

Additionally, creating a centralised repository like the Methods Hub in a vendor-neutral format is a step towards greater cyber-resiliency in the industry. When information is stored locally on a PC or an ELN and is not backed up, a single cyberattack can wipe it out instantly. Creating shared spaces for these notes via the cloud protects data and ensures it can be easily restored.

A proof of concept (PoC) via the Methods Hub project was recently successfully completed to demonstrate the value of methods digitisation. The PoC involved the digital transfer via cloud of analytical HPLC methods, proving it is possible to move analytical methods securely between two different companies and CDS vendors with ease. It has been successfully tested in labs at Merck and GSK, where there has been an effective transfer of HPLC-UV information between different systems. The PoC delivered a series of critical improvements to methods transfer that eliminated the manual keying of data, reduced risk, steps and error, while increasing overall flexibility and interoperability.

The Alliance project team is now working to extend the platform's functionality to connect analytical methods with results data, which would be an industry first. The team will also be adding support for columns and additional hardware and other analytical techniques, such as mass spectrometry and nuclear magnetic resonance spectroscopy (NMR). Plus, it plans to identify new use cases, and further develop the cloud platform that enables secure methods transfer.

Dr Birthe Nielsen is the Pistoia Alliance Methods Database project manager.

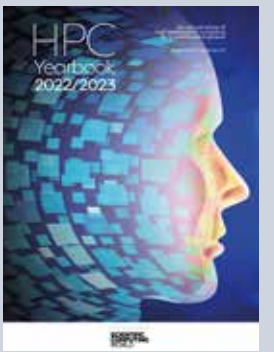
Subscribe for free*

Do you compute?

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

Also published by Europa Science

HPC 2022/2023



*Registration required

Do you subscribe? Register for free now!

scientific-computing.com/subscribe

Driving adoption of the paperless lab in India

Sachin Bhandari provides an overview of his talk from the Paperless Academy India 2022 event

In 2022, The Paperless Lab Academy hosted the third edition of its conference in India, but it was the first time that a live, in-person event was possible. The event showcased discussions to help laboratory organisations to implement paperless technology and strategies for digital transformation.

Event organisers' aim was to provide a platform to discuss key milestones to generate business insights, while also connecting like-minded experts to share their experience and best practices for digitisation and digital transformation.

Sachin Bhandari was one of the expert speakers lined up for the PLA India keynote presentations. His talk, "Milestones of a digital transformation journey", highlighted key points on the journey towards lab digitisation and share experience gained from carrying out this process at Sun Pharma.

With more than 20 years of experience in pharmaceutical and healthcare compliance, Bhandari is now Sun Pharma's Senior General Manager (Global Head QA-IT and Quality Digitisation Projects). He is responsible for IT compliance and helps to lead quality initiatives pertaining to IT and the readiness of the organisation for wider IT compliance.

'I've been with Sun Pharma for almost eight years now. And, primarily, my job is to make sure that our IT systems stay compliant at all times. I take care of QA-IT compliance, and also drive the digitisation of CSV processes for Sun Pharma and the digitisation projects for our QA/QC, and IT compliance,' stated Bhandari.

He has considerable experience in GXP, 21 CFR part 11, Annex -11, GAMP 5 compliance (Validation and Quality). He has also primarily worked on validation activities related to quality systems, lab systems validation, IT infrastructure validation, supply chain validation, quality systems and service-desk compliance.

'Digital transformation is an investment," he said. 'Digitisation gives you an opportunity to re-invent your processes. It helps you to optimise and make those processes more aligned with changing business needs – that's how I look at it. This is an investment that has a very high return,' Bhandari continued.

> "Digital transformation is an investment. Digitisation gives you an opportunity to re-invent your processes"

'Digitisation improves compliance, and all of us in the pharma industry know what the cost of non-compliance is. Digitisation helps us to improve our level of compliance, and it makes overall compliance faster and more efficient.'

Bhandari is currently driving paperless initiatives at Sun Pharma, supporting the implementation of technological innovations to improve quality and compliance. He was Chairman of ISPE GAMP steering committee India Chapter, and is a prominent speaker at various forums and seminars.

'The Paperless Lab Academy was a very exciting opportunity because the lab forms a very critical part of our entire quality gamut. In addition, a lot of organisations have different approaches towards achieving a paperless lab. If we look at my talk, for example, it's about

the milestones in achieving this kind of transition towards digital transformation and the paperless laboratory.

'Different organisations will be at different stages. When you talk about the different milestones it's not a turnkey approach, it's a process,' commented Bhandari. 'In my talk, I considered the reasoning behind Sun Pharma's decision to adopt digital technology. An organisation must consider the different steps or milestones they need to achieve. That will help people to understand where they are, and the roadmap ahead.'

> "Have you digitised all your instruments? Is your quality analysis digitised? Are you using artificial intelligence and robotic processes to assist in compliance?"

While there are many paths towards digital transformation, Bhandari highlighted several milestones that can be crucial steps for lab-based operations around the world: 'Have you digitised all your instruments? Is your quality analysis digitised? Are you using artificial intelligence and robotic processes to assist in compliance?

'I also spent some time discussing the key pitfalls and challenges that can come in at various milestones, when you go from one level to another.'

The advice and expertise that can support digital transformation is now becoming more readily available for lab users and managers that want to begin the process within their own organisation. Bhandari noted there have been three main changes that have caused this shift towards digitisation.

He described how, in the past, professionals thought that digitisation just meant converting processes from paper-based systems to digital. However, Bhandari stressed that, if thought out properly, digitisation provides an opportunity to re-invent processes and increase operational efficiency. 'Now they have started to understand it is a lot more than that. That's the number one change,' he told the audience at the event.

'Second, it used to be true that we did not have the vendors, or the business analysts available in the market. We had IT guys who did not understand pharma and pharma guys who did not understand IT,' said Bhandari. 'Today, we have lots of business analysts, and a lot of use cases already available in the market. We have vendors with products that are configurable and can be easily deployed. This was not there, but now we have a lot of choices.

> "I'm looking forward to very interesting knowledge exchanges, because there is no single set formula for digitisation"

'Third, I think the manufacturers have started to understand they need to have more open interfaces, so that their data can be part of data lake and data analytics. Previously the market was dominated by closed systems,' stated Bhandari. 'Now they have started opening up a bit, not everything is proprietary. These are the three primary changes I came across.'

Directory of suppliers

A list of leading suppliers, consultants and integrators

Autoscribe Limited

1 Venus House, Calleva Park
Aldermaston, Reading, RG7 8DA, UK
Tel: +44 (0) 118 984 0610
info@autoscribeinformatics.com
www.autoscribeinformatics.com

Autoscribe Informatics offers industry leading configurable future-proof LIMS software. With 30 years' experience our solutions are configured to match user requirements, and can easily adapt as needs evolve, using the built-in configuration tools. Matrix Gemini LIMS allows you to track and manage testing activities within a controlled environment with ease.

Thermo Fisher Scientific Informatics

St. George's Court Unit 1, Hanover
Business Park, Altrincham WA14 5TP,
United Kingdom
+44 161 942 3000
www.thermofisher.com/informatics

Digital transformation is no longer an innovative vision but is becoming a strategic imperative for companies to be competitive. It is crucial for businesses to find a way to create and implement strategies to take advantage of digital technologies.

Thermo Fisher Scientific has industry-leading scientific and laboratory operational excellence, paired with an extensive portfolio of digital capabilities to aid in your journey of digital transformation. Our complete suite of digital solutions enables you to connect everything - lab automation, data management and digital partners, giving you complete control and oversight of your lab.

LabWare

Denzell Lodge, Dunham Road
Altrincham, Cheshire,
WA14 4QE
Tel: +44 161 927 5600
infoeu@labware.com | www.labware.com

LabWare is recognised as the global leader of Laboratory Information Management Systems (LIMS) and instrument integration software products. The company's Enterprise Laboratory Platform combines the award-winning LabWare LIMS, LabWare ELN/LES and LabWare Mobile, which enable companies to optimise compliance, improve quality, increase productivity, and reduce costs. A wide range of industry template solutions are also available to provide best practice and help fast-track installations. LabWare is a full-service provider offering software, professional implementation services and validation assistance, training, and world class technical support to ensure our customers get the maximum value from their LabWare products.

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

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

iVention
+31 38 452 83 75
https://ivention.nl

LabVantage
+44 (0) 1494 477977
www.labvantage.com

STARLIMS UK Ltd
+44 161 711 0340
www.starlims.com

STARLIMS

ONE PARTNER, ONE POWERFUL SOLUTION.

More than just working towards a paperless lab, digital transformation may allow companies to derive more intelligence from their data, ultimately improve product and process safety, and enable faster regulatory audit or approval timelines.

For more than 35 years in the market, STARLIMS is a mission critical application that supports this digital transformation from the ground up.

Discover how our solutions can support you in the digital transformation journey.

LABWARE®

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

www.labware.com

LABWARE® Results Count