KAO DATA

# How Kao Data is supporting EMBL-EBI to become the most comprehensive and easily accessible public biological data library for the global scientific community

The European Bioinformatics Institute's Head of Technical Services - Dr Steven Newhouse, cited seven key considerations behind their decision to sign with Kao Data: network connectivity, infrastructure resilience, space & power capabilities,green credentials, proximity to Cambridge and Kao Data's competitive pricing.

EMBL-EBI

The European Bioinformatics Institute (EMBL-EBI) has been using external data centres for more than a decade to complement its own on-campus data centre capability at the Wellcome Genome Campus in Hinxton. In 2019, with EMBL-EBI's IT activities increasing significantly year-on-year, the institute signed a contract to colocate its IT infrastructure at Kao Data's state-of-the-art, 40MW campus in Harlow, primarily to support its ongoing and expanding bioinformatics service and data storage capability. The move highlights the life sciences' need for ever-increasing levels of compute and storage to properly record data and continue their life-changing work within a worldwide community. Kao Data's commitment to achieving technical and operational excellence without compromising on sustainability also resonated deeply with the organisation.

## The Situation

Dr Steven Newhouse, EMBL-EBI's Head of Technical Services, says: "EMBL-EBI's main role is to archive, for the international research community, the open data that it wishes to share, and to provide added value services that turn data into knowledge. We do this across a whole range of data types, from short genetic sequences through to large multi-channel images – and this scale of data storage and analysis drives many of our IT strategies. For example, we have over 300 Petabytes of raw storage across our data centre environments to help us sustain our activities."

KAO DATA

Over the last 10 years, there has been a marked increase in the amount of data that EMBL-EBI curates, with the commoditisation of genetic sequencing and its widespread adoption. EMBL-EBI has been able to respond to this growing need with the support of funding from UK Research and Innovation.

Since the start of the current Covid-19 pandemic, there has been a further increase in data uploads and particularly data downloads as the organisation secured funding from the European Commission and worked with the global scientific community to set up the **COVID-19 Data Portal** encompassing related data sets and services in one place.

Steven says: "We support a number of analysis platforms, including in-house HPC clusters and private cloud, and we additionally operate a range of public web services and databases hosted on virtual infrastructure. Therefore, we need to use a range of different storage mechanisms to support such varied workloads. Data centre storage, not just now but in the future, is therefore hugely important for us. As is having a flexible and accessible network to host the data and enable vast amounts of data to be moved to and from users, internally and externally, across the whole of Europe and, increasingly, across the whole world.

"To support this, we needed an external data centre which would enable us to host significant storage and compute capacity, with scope to expand, and which was, additionally, located close to the Wellcome Genome Campus in Cambridge. As we operate 24/7, we also needed a highly resilient data centre that could offer 100% uptime with multiple, redundant paths for electrical and network connections.

"Clearly the 24/7 operation has to be sustained by ensuring that there is more than adequate access control to the premises. We have many millions of pounds worth of equipment housed within the Kao Data campus, we obviously want that to be secure. On top of that physical infrastructure, we layer our own security IT environments to protect the data that service our own internal and external activities."

KAO DATA

## The Solution

Very few scientific and research organisations operate at the scale that EMBL-EBI does in terms of data volumes and data analysis. EMBL-EBI's compute profile includes HPC servers, graphics processing units (GPUs) and a range of storage options that operate at high densities, well in excess of normal enterprise standards.

EMBL-EBI has already finessed its IT platform to achieve a x10 scale of power density and over 25kw per rack. But, given the increasing use of AI and graphics-based processors in their data analysis, and the elevated levels of compute required to do this, the organisation is now looking to increase processing speeds and capabilities to achieve a density of 30-35kw per rack.

Kao Data underpins EMBL-EBI's ability to do this – providing a technically robust environment with industrial scale cooling and an abundant power supply to look after and cope with its high-density computing requirements – now and into the future.

Dr Newhouse says: "One of our greatest challenges is that high performance computing infrastructure is extremely power-hungry and, in the past, as the power density has increased, we have ended up with empty space in some of our data centres as we've exhausted all the electricity capacity that we can draw to keep the equipment occupied.

"However, Kao Data has the capacity to deliver that rack density and volume of electricity into the rack space that we have, and we have an ambition to carry on expanding. With the modular infrastructure Kao Data has in place, we will be able to scale up the power delivery and carry on populating the physical space that we have without over-challenging the electrical delivery and reliability.

"We generate the connectivity between our different data resources using a range of algorithmic data analysis techniques. Increasingly, we are looking to GPUs to supplement traditional computing infrastructure as well as other computing patterns and approaches, such as artificial intelligence and machine learning, to help derive the connectivity between the different data sets that we have. Kao Data enables us to operate at the scale that we need to, and means that we can then provide a blueprint to the rest of the scientific community as to how to deal with some of the data scales, volumes and densities that they might see taking place on a national and local level; and which we are already able to see from our international perspective."

EMBL-EBI's substantial power and storage needs, however, are being challenged by the organisation's green principles and sustainability goals. From its own analysis, 50% of EMBL-EBI's institutional carbon footprint has come from its data centre activities. Kao Data's impressive green credentials, however, are now helping EMBL-EBI become more environmentally stable, which is key to the organisation's vision.

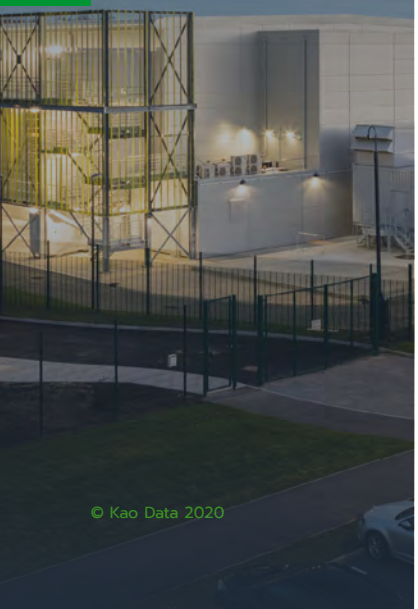
Kao Data's campus is powered by 100% certified green energy and it is the UK's first 100-percent free-cooling wholesale colocation campus. Innovative cooling technology, which removes the requirement for refrigerant, further minimises the data centre's environmental impact - and due to excellent technical design, inspired by hyperscale, the data centre boasts an exceptionally low PUE (power usage effectiveness) of 1.2 (even at partial loads).

BREEAM, the world's leading sustainability assessment method for master-planning projects, infrastructure and buildings, also certified Kao Data's first data centre, KDL1 (in which EMBL-EBI is hosted), as 'BREEAM Excellent'.


Kao Data's CTO, Gerard Thibault explains: "We have a 100% certified renewable energy guaranteed supply with EDF Energy and, because of this, we can report a zero CUE (Carbon Usage Effectiveness) and that our power supply is carbon neutral.

"We use the most efficient type of air-cooling system currently available, an indirect evaporative cooler that doesn't rely on any refrigeration cooling whatsoever and which is excellent from a sustainability viewpoint as it involves no gases and therefore no gas emissions. EMBL-EBI, like ourselves, is very focused on sustainability and energy efficiency and this really appealed to them - not just from a corporate and social responsibility point of view but from a bottom-line running cost.

"It means we can offer a capped PUE of 1.2 and the cost of actually running their facility becomes much less. A PUE of 1.2 means that, for every kilowatt hours' worth of IT energy they need and use in IT processing, they only pay for 1.2 hours to run the facility. The average PUE across the industry is 1.67 which means, at 1.2, we are able to offer an 80% reduction in overheads which is a huge benefit in terms of the total cost of operations (TCO)."



KAO DATA

## Success

From the perspective of a global organisation, the public provision of easily accessible open data and open science for accelerating knowledge development is of vital importance. This is what EMBL-EBI does, and what Kao Data helps to support them to do.

Steven says: "In January 2020, China sequenced and shared the genome of the novel coronavirus with the international community. This triggered research activities before most of the world had even heard of the Covid-19 disease or been impacted by it. The data that have been generated by the scientific community, and collected through EMBL-EBI and other organisations, have enabled that scientific knowledge to help us in understanding how the disease behaves in different people and understanding the structure of the virus itself. And all of this helps to provide knowledge that we hope will lead to vaccines and better treatments further down the road."

"One of the big changes that IT is going through – and research computing and scientific IT is no different – is: How do organisations complement their on-premise IT infrastructure with public cloud infrastructure? EMBL-EBI is going through that journey as well, and Kao Data now plays a part in ensuring that our on-premise IT is very well connected and complimentary to any public cloud resources that we use, and key to that is the very fast responsive networking and 24/7 availability which they enable.

"EMBL-EBI has recently established a bioimaging archive to complement our other long-running archives. Clearly, dealing with images involves a different type and scale of data. Enabling users to visualise and interact with the image data poses different challenges and access patterns that we now need to support.

"Fundamentally, Kao Data is a better quality, more resilient and industrial scale data centre that complements our original on-premise facility. And, importantly, it offers us the ability to expand and scale our operations. Next year, we plan to concentrate the bulk of our activities at Kao Data and, in so doing, improve the resilience of more of our services by moving them there."