Secure, resilient storage for 100,000 genomes in the cloud

Run by Genomics England, the 100,000 Genomes Project is the largest national sequencing project of its kind in the world. It aims to create a genomic medicine service within the NHS, and to kick-start a UK genomics industry. De-identified data is securely stored on the UKCloud platform, where authorised researchers have access to it. The project has already delivered diagnoses to some families, transforming the way NHS patients are cared for.

When Dave Brown joined Genomics England as Head of Infrastructure, he knew he had to get things moving fast. The first data from the 100,000 Genomes Project was due to arrive within three months, but Genomics England had no IT infrastructure in place to receive it. Brown knew that the traditional route of buying tin and installing it in an academic data centre would take too long. Instead, he opted for cloud.

"With such a short time to deployment, we needed infrastructure that was quick to buy and set up," says Brown. "Only a cloud platform available on a government framework contract could deliver."

Using the Digital Marketplace, Brown researched cloud providers on the G-Cloud framework, and chose UKCloud. "UKCloud is clearly a leader in its field: a mature company with the resources and breadth of services we needed," he says. "And unlike some of the other providers we looked at their cloud platform was real and ready to use."

Security, protection and resilience built in

Given the sensitivity of the data generated by the project, Brown needed to be sure that the cloud platform he selected was secure and resilient. UKCloud met his requirements on all counts.

"The UKCloud platform is based in England, which does away with data residency and protection issues that are often associated with clouds delivered from outside the UK," says Brown. "It has all the required industry accreditations and certifications, so you know it's secure. On top of that, it was Pan Government Accredited, which adds yet another layer of assurance."

Because UKCloud operates out of two data centres, it could provide a resilient environment with automated failover. "UKCloud met all the specs we had at the start, and we've worked together to refine the setup as the 100,000 Genomes Project has developed," says Brown.

Connectivity options to fit every use case

The project involves individuals and their families consenting to provide DNA samples, which are analysed at the sequencing centre. The centre sends the resulting DNA data to the Genomics England platform at UKCloud over secure, dedicated 10Gbit/s connections using UKCloud's HybridConnect service.

Health data about participants is provided by the hospitals and clinics that recruit them, and is sent to the platform over the N3 network. "UKCloud having N3 aggregator status was key to making this straightforward," says Brown.

All the data is de-identified. Authorised researchers access the data over the internet, using a secure virtual desktop built on top of the UKCloud platform, and managed by Genomics England. Their access is closely controlled, and they're not permitted to download any of the data.

ukcloud **health** | **case study**

Pioneering a repeatable platform

Working together to build the infrastructure has been the touchstone of the relationship between Genomics England and UKCloud. UKCloud provides the Compute as a Service platform for data processing and analysis; and hosts and manages specialised storage equipment which is owned by Genomics England and integrated into the cloud environment. "We provided our own storage because of the sheer volume of data that will be generated by the project," explains Brown. The data from a human genome needs around 240Gb of storage: 24Pb will ultimately be needed to store 100,000 genomes.

"Working with UKCloud, we've pioneered the definition and development of a computing environment that can deal with the scale and complexity of genomics sequencing data — something most hospital data centres would struggle to support," says Brown. What's more, it's a repeatable environment: other NHS bodies and hospitals doing genetics sequencing can confidently buy the infrastructure they need from UKCloud, simply and easily via G-Cloud.

"Right from the start, it's been a truly collaborative effort that meant we got the infrastructure up and running quickly and efficiently," says Brown. "We simply couldn't have done it without the G-Cloud framework, or without UKCloud being an accredited G-Cloud supplier."

About Genomics England

Genomics England is a wholly owned company of the Department of Health. Genomics England is working together with NHS England, Health Education England and Public Health England to deliver the 100,000 Genomes Project. This flagship project will sequence 100,000 whole genomes from NHS patients and their families. The project is focusing on patients with rare diseases, and their families, as well as patients with common cancers.

genomicsengland.co.uk