

Data Analytics helps Cancer Research stand up to the big C.

Overview.

As the name suggests, Cancer Research UK (CRUK) is a charity dedicated to raising awareness and driving pioneering research to prevent, control and cure cancer. Since their establishment in 2002, its ground-breaking research activities, awareness campaigns and public policy influence has helped double survival rates. Almost entirely funded by the public, CRUK depends on advancements in technology and science to develop state-of-the-art validated know-how, tools, reagents and bioinformatic platforms that can bring new insights into cancer biology and clinical drug recovery.

The Challenge.

Challenged with the monumental task of keeping one step ahead of cancer, CRUK launched a joint initiative with a leading pharmaceutical provider to conduct functional genomics research. The goal was to support the cutting-edge application of CRISPR (clustered regularly interspaced short palindromic repeats), with new labs set up to perform screening, analyse results and develop new tooling.

ECS were invited to help design and build a bioinformatic data analytics pipeline that would bolster the existing initiative. Whilst initially built as an on-premise solution, CRUK were keen to execute the pipeline using an entirely cloud-based solution. As a charity, it also needed to be economical with a focus on long-term operating costs. This included mechanisms to easily monitor health and troubleshoot pipeline components and keeping ongoing maintenance costs to a minimum.

The Solution.

CRUK engaged with ECS to help them incorporate a centralised configuration into their existing solution, build a CI/CD pipeline that could automate the deployment of bioinformatician code and integrate a lean tool stack. As well as delivering a sustainable analytics pipeline, CRUK were specific that the solution needed to reduce maintenance costs, minimise broken deployments and allow easy rollback in the case of failure – these formed the blueprint of the infrastructure design.

ECS joined CRUK during the delivery phase of the project. The first step was to initiate a project inception to confirm and refine the requirements on a progressive basis. This helped onboard the right internal team and stakeholders so ECS could dovetail around the existing CRUK skillsets.

Over the five-month engagement, ECS designed and built a CI/CD analytics pipeline that today provides CRUK with the mechanism to run analysis of DNA screens from their research lab. To enable more advanced management and categorisation of their data, all data sources were downloaded into AWS and transformed into a structure suitable for analysis before moving to an Amazon S3 bucket.

We also integrated the following tools to help support additional business needs:

- The failure or successful completion of a data transformation is logged using AWS CloudWatch.
- Batch analysis is performed using Nextflow which is used as a harness to sequence AWS Batch job execution, as well as allowing the pipeline code to be abstracted from the infrastructure.

The above capabilities allow the pharmaceutical partner and CRUK infrastructure implementations to vary without impact to the analytics pipeline.

What value did ECS bring?

Cancer Research UK chose to partner with ECS due to the breadth and depth of the consultancy's digital transformation team, and its deep AWS experience.

By adopting a Pod-based approach, ECS had the flexibility to embed handpicked specialists into CRUK's team in a series of outcome-focused sprints. As the project unfolded, ECS created comprehensive documentation and well-structured code to reduce long term maintenance cost and on-boarding time. The team also ensured the analytics environment and pipeline could be launched by research staff or automatically executed based on new screening data becoming available. This means multiple analytic data sets per day can be processed.

All significant analytics pipeline events are audited, and key exceptions or failures are notified to nominated Microsoft Teams groups. By doing so, the technical support effort required to maintain the analytics pipeline has been minimised. It's also now possible to update the pipeline using an automated process – reducing administrative overhead and maintenance costs, while also providing a platform suited to automated testing.

In the words of Chris Moore, director of engineering at Cancer Research UK:

"Cancer Research UK's ambition is to accelerate progress so that by 2034, 3 in 4 people will survive their cancer for at least 10 years. By working with ECS, we are building the systems needed to help realise this ambition. We have been impressed by ECS's professionalism when upskilling our staff, and when designing, building and delivering this state of the art and cost effective bioinformatic data analytics pipeline."

Benefits.

- Keeping development costs to a minimum by using a JIT approach that scaled up and down automatically in AWS.
- Reducing the likelihood of huge data sets being inadvertently corrupted by human interaction by automating the workflows.
- Going cloud-native, using a wide range of tools including AWS CloudWatch and Nextflow.
- Upskilling the charity's teams by training them to build pipelines for future software releases, use various AWS tools, and agile ways of working.



Go cloud-native whilst embracing automation and transforming your data into a structure suitable for analysis.

Start your journey today.

ecs.co.uk | 0207 403 0477
hello@ecs.co.uk