By analysing large microbiome data sets, researchers based at the Science and Technology Facilities Council (STFC) Hartree® Centre are working with established agricultural research centre - Rothamsted Research – to understand how soil microbiomes can shape soil health.

Challenge
The ability to understand the microbiological profile of soil, its biochemical properties and fertility is crucial to address global challenges including food security and climate change. Soils are the most complex biological systems on Earth but understanding their microbiomes in detail requires analysis of huge amounts of data. This is a computationally demanding exercise that needs skills in data analytics to provide real insight. Rothamsted Research needed to speed up their metagenomics data processing capability and identify methods that could provide new insight to help understand soil health.

Approach
The IBM Research team developed highly-parallelised workflows suitable for processing large scale metagenomics datasets on the Hartree Centre’s high performance computing (HPC) systems. Using their expertise in data science, the team devised methods to distribute datasets, ultimately reducing the run time from days to hours. By further developing APIs – application programming interfaces – to mine a range of diverse public databases, the team created a rich, integrated knowledgebase used to develop machine learning models that were capable of offering a detailed picture of biochemical activities across soil samples and their connection with overall soil health.

Benefits
This work provided a HPC-enabled solution to the data processing bottleneck experienced in soil metagenomics. The computational and data analytics capabilities of the team complemented the expertise of Rothamsted Research to create a highly-integrated and effective collaboration, providing novel insights towards the understanding of soil health. These tools - completed part of the Innovation Return on Research (IROR) programme, a collaboration between STFC and IBM Research - can now be applied to processing any microbiome datasets, transferable across a broad range of industry challenges.

Working with the Hartree Centre has been extremely positive, allowing us to broaden our horizons - imagining more complex, larger studies and helping us establish world-leading research in areas critical to food security and global climate.

Andrew Neal
Rothamsted Research

www.hartree.stfc.ac.uk
@hartreecentre /company/stfc-hartree-centre
At a glance

- Were able to process large datasets on high performance computing platforms, reducing run time from days to hours
- Machine learning methods provided a detailed view of soil properties by analysing the microbiome
- Led to a better understanding of factors affecting soil health
- Resulted in a set of transferable tools that can be applied to any microbiome datasets

Who we are

- 60+ computational scientists and technologists
- World-leading supercomputing and AI infrastructure
- Bespoke small teams built around your project
- Tailored business development support
- Access to our network of industry, academic and technology partners

What we do

- Boost productivity and enhance innovation for industry
- Big data analytics and artificial intelligence (AI)
- High performance computing and quantum simulation
- Training and skills development
- Insights into future technologies

Our impact on UK industry and society

The Hartree Centre was created by UK Government to transform industry by accelerating the adoption of high performance computing (HPC), big data analytics and artificial intelligence (AI) technologies. We play a key role in realising UK Government’s Industrial Strategy by stimulating applied digital research and innovation, creating value for the organisations we work with and generating economic and societal impact for the UK.

The Science and Technology Facilities Council (STFC) Hartree Centre is part of UK Research and Innovation.

www.hartree.stfc.ac.uk

@hartreecentre /company/stfc-hartree-centre hartree@stfc.ac.uk