



# GÉANT and EMBL-EBI: Empowering the global scientific community

International collaboration pushes the frontiers of life sciences, enabling discovery in medicine, agriculture and environmental science

## The data deluge

The life sciences have undergone a revolution over the past 20 years. Sequencing the DNA of thousands of organisms, including humans and agriculturally relevant species, is generating unprecedented volumes of data in hundreds of research organisations around the world. This disruptive technology has changed the face of research: what used to take years of work can now be done in minutes – and cheaply. Because of this, bioinformatics – the science of storing, managing and integrating data from biological experiments – has become a fundamental component of modern research.

EMBL-EBI is leading this revolution in information sharing. Part of the European Molecular Biology Laboratory (EMBL), the EBI is the European node for many global projects. Its mission is to create and provide bioinformatics services that make a world of data freely available to all. Established in 1994, it shares a campus with the Wellcome Trust Sanger Institute, a leading genetic research institute that has been responsible for high-profile projects such as the original decoding of the human genome.

In collaboration with hundreds of partners around the world, EMBL-EBI manages data from countless life sciences projects that produce terabytes of information. It relies on high-speed, robust networks that can carry large amounts of information securely so that researchers can access – and upload – data and information freely. This is no small task, as EMBL-EBI's website gets around 1 billion hits a year.

This demanding level of performance requires sophisticated engineering and a robust network. EMBL-EBI uses JANET, the UK's National Research and Education Network (NREN), and the GÉANT pan-European research and education network to communicate across Europe and beyond. Together, JANET and GÉANT helped this key research institute to explore and understand its network options and provided tailored consultancy to ensure that service levels can meet its exact requirements well into the future.

As well as providing robust links at speeds of up to 1 gigabyte per second between EMBL-EBI and other European networks, GÉANT securely interconnects academic networks around the world – including Internet2 in the US and CERNET in China – allowing truly global, real-time research collaboration. Over 80 terabytes of data is transmitted by the EBI over the high speed, high bandwidth JANET and GÉANT networks every month.



## The Challenge

To store, manage and integrate the skyrocketing volume and variety of biological information produced in life sciences research, and to make it available to the global scientific community.

## The Solution

Working in partnership, the JANET and GÉANT networks provide seamless, high-performance links between the EMBL-European Bioinformatics Institute (EMBL-EBI) in Cambridge, UK and scientists located throughout the world, enabling real-time, global access to the world's largest collection of molecular databases.

## Key Benefits

The results of large-scale, ground-breaking initiatives such as the 1000 Genomes Project can now be distributed and made available for analysis quickly. This kind of research has already been used in several important medical studies, and shows how genetic research is enabling rapid advances in the medical sciences.

Data generated by biological experiments is doubling every five months, driven by vital collaborations such as the 1000 Genomes Project. Our mission at the EMBL-EBI is to make the results of these international collaborations freely available to the scientific community wherever they are located. To do this we need an infrastructure that is robust, flexible and high performance, linking us to our partners across the globe. Our close working relationship with the JANET and GÉANT networks delivers the speed and capacity that we need, giving us confidence and allowing us to focus on sharing data to push forward scientific progress.

*Dr Paul Flicek, Head of Vertebrate Genomics, EMBL-EBI.*

### Thousands of genomes

EMBL-EBI is currently involved in a number of large-scale research projects that are pushing the frontiers of scientific knowledge. One of these is the 1000 Genomes Project, a major international research effort to sequence the genomes of 2500 people worldwide in order to understand human variation. Chaired by Richard Durbin at the Wellcome Trust Sanger Institute in the UK and David Altshuler at the Broad Institute in the US, the Project includes participants from Europe, the Americas and Asia.

The genes of any two humans are more than 99% the same. Understanding the very small genetic variations that make us different is key to advances in the diagnosis, prognosis and treatment of disease. For example, it is helping to explain why people react differently to certain medications, and how genes contribute to susceptibility cancer or addictive behaviour such as smoking.

The pilot phase of the project, completed in 2010, generated 4.9 terabases of DNA sequence, uncovering 8 million variations that had never been seen before. The full project, scheduled for completion in 2012, is expected to produce between 60 and 80 terabases of data (around 250 terabytes of data). At the outset of the project, researchers at EMBL-EBI, led by senior researcher Paul Flicek, created a strategy to characterise these variations and the bioinformatics infrastructure to support the massive movement of data.

Data from the project currently flows to EMBL-EBI from seven sequencing centres in China, Germany the UK and the US. All of it is then mirrored to the National Human Genome Research Institute (NHGRI) in the US and supplied to 40 groups across the world for initial analysis. Final datasets are then accessed by another 100 research groups of various sizes. To enable even faster access to information, the results of the pilot phase of the 1000 Genomes project is available online through Amazon Web Services.

In addition to GÉANT, the 1000 Genomes Project uses the EU-funded ORIENT link to facilitate collaboration with its research partners in China, and the Internet2 network to share data with its NHGRI partners in the US.

### connect • communicate • collaborate

GÉANT and JANET provide EBI with the high capacity, global connections that enable seamless sharing of biological data. The combination of the ground-breaking research undertaken by EMBL-EBI and its international partners and the power and reach of the GÉANT network provides the perfect platform for global bioinformatics.

### Sharing genetic data globally – Ensembl and EGA

Since the decoding of the human genome in 2000 and the increasingly rapid deciphering of the genomes of other plants and animals, there has emerged a pressing need to share a growing volume and diversity of research data with scientists across the world. Ensembl provides a centralised resource for geneticists, molecular biologists and other researchers studying all manner of species. A joint venture between the EBI and the Wellcome Trust Sanger Institute, Ensembl allows scientists to access and query genetic information such as completely sequenced genomes through a freely available web interface.

Another crucial service is the European Genome-phenome Archive (EGA), which looks at the differences between different human genotypes. Designed as a repository for genotype experiments, the EGA provides a single point of access to the most up-to-date genotype data. The data held in the EGA is used to identify small genetic differences between individuals that affect their predisposition for certain diseases, and as such is an important resource for biomedical research.

Large-scale data-sharing projects all require reliable, fast networks. JANET and GÉANT make it possible to deliver data wherever it is needed, driving closer scientific collaboration and facilitating breakthroughs in research. They provide the fundamental structure that makes real-time bioinformatics-based research possible.

### The future – ELIXIR

By 2020, biological data will be generated at thousands of times the current rate. This growth far exceeds predicted increases in storage capacity, and it is unrealistic to expect that a single organisation or nation could cope with such demand. ELIXIR ([www.elixir-europe.org](http://www.elixir-europe.org)) unites Europe's leading life sciences researchers and data providers to create a robust, distributed bioinformatics infrastructure that can support the science of the future.

The results of publicly funded research – the data – belong to everyone, and must be safeguarded. ELIXIR, an ESFRI project coordinated by EMBL-EBI, will play an important role in many other life science infrastructures. Its success will rely on high-speed networks, such as GÉANT, JANET and other national research networks across Europe to deliver data in real time to a global scientific community.

The volume of biological information available to researchers is rapidly expanding. Being able to transmit, store and share this data is vital to the cross-disciplinary, international collaboration that is at the heart of today's pioneering biological research projects. The combination of EMBL-EBI's essential bioinformatics services and the power and reach of the GÉANT network is enabling tremendous advances in life sciences research the world over, and contributing to European innovation.

### For more information:

GÉANT <http://www.geant.net>

EBI <http://www.ebi.ac.uk/>

EMBL <http://www.ebi.ac.uk/embl/>

ESFRI [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)

ORIENT <http://www.dante.net/server/show/nav.1678&t>

JANET [www.ja.net](http://www.ja.net)

