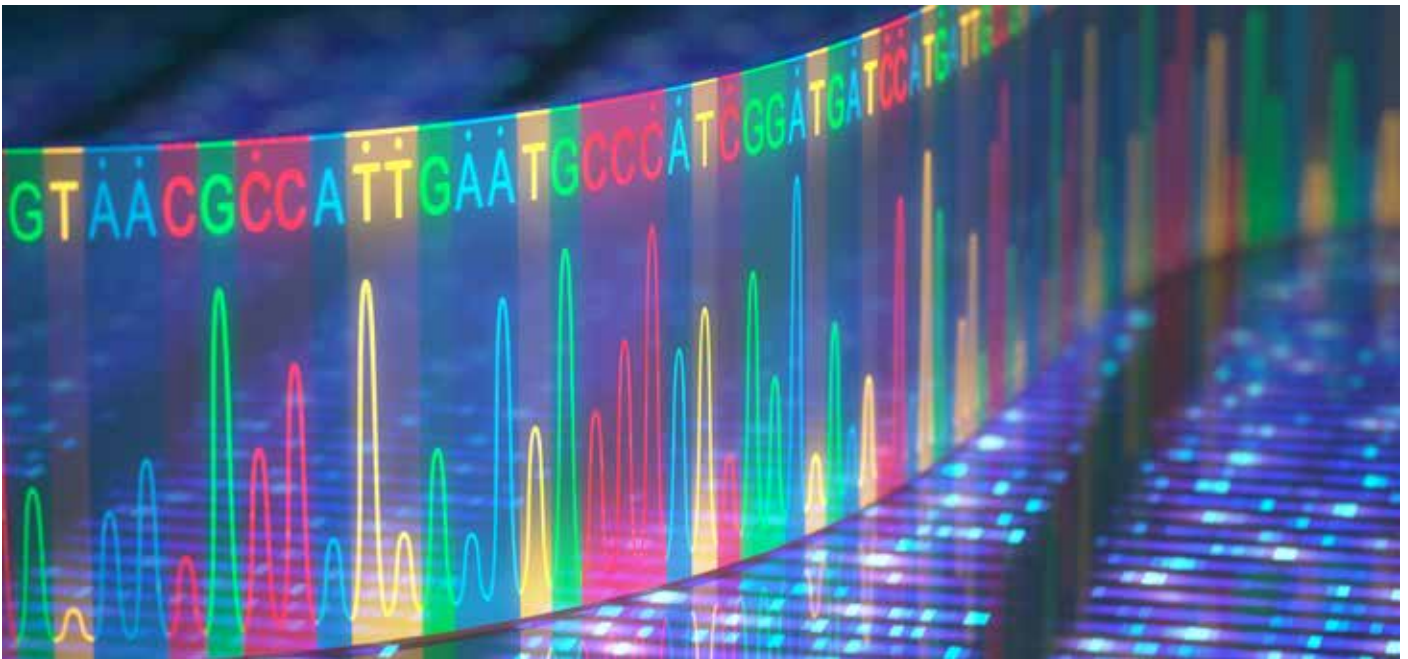


# Genetic data under control and in the desired format

Biotechnology industry research produces a huge volume of data, and the amount doubles every few months. That is why data management requires sophisticated tools. This can be implemented in cooperation between public biological data infrastructures, such as ELIXIR, and companies, such as BC Platforms.



BC Platforms offers information systems for the management of genomic data. The two heavy-duty databases developed by the company are also used in the ELIXIR infrastructure through the Finnish ELIXIR node CSC. BC Platforms is now in the process of creating an ecosystem where the data sets of biobanks from different countries can be searched by using a common user interface.

BC Platforms has more than 20 years of experience in handling large data sets. The company's information management systems can be set up in a local computing environment or the cloud. A virtual file system operates in the background. Users log in to the database and retrieve the material from the server. The changes made by users are then saved back into the database, so that huge amount of files are exported and imported by the secured network. This so-called object-based storage is particularly suitable in cases where data needs to be stored for a long time while also taking into account information security.

## **"BC Platforms wants to create an open ecosystem between researchers, pharmaceutical companies and biobanks."**

The items analysed by the customers of BC Platforms range from the data of a single person or animal to cohorts consisting of millions of individuals. The clientele also includes research organisations that produce up to 10,000 genomes a day.

BC Platforms wants to create an open ecosystem between researchers, pharmaceutical companies and biobanks. The BC I REQUEST service provides information about the data in different biobanks. Via the service's user interface, researchers and drug developers have centralised access to the material of the biobanks belonging to the cooperation network.

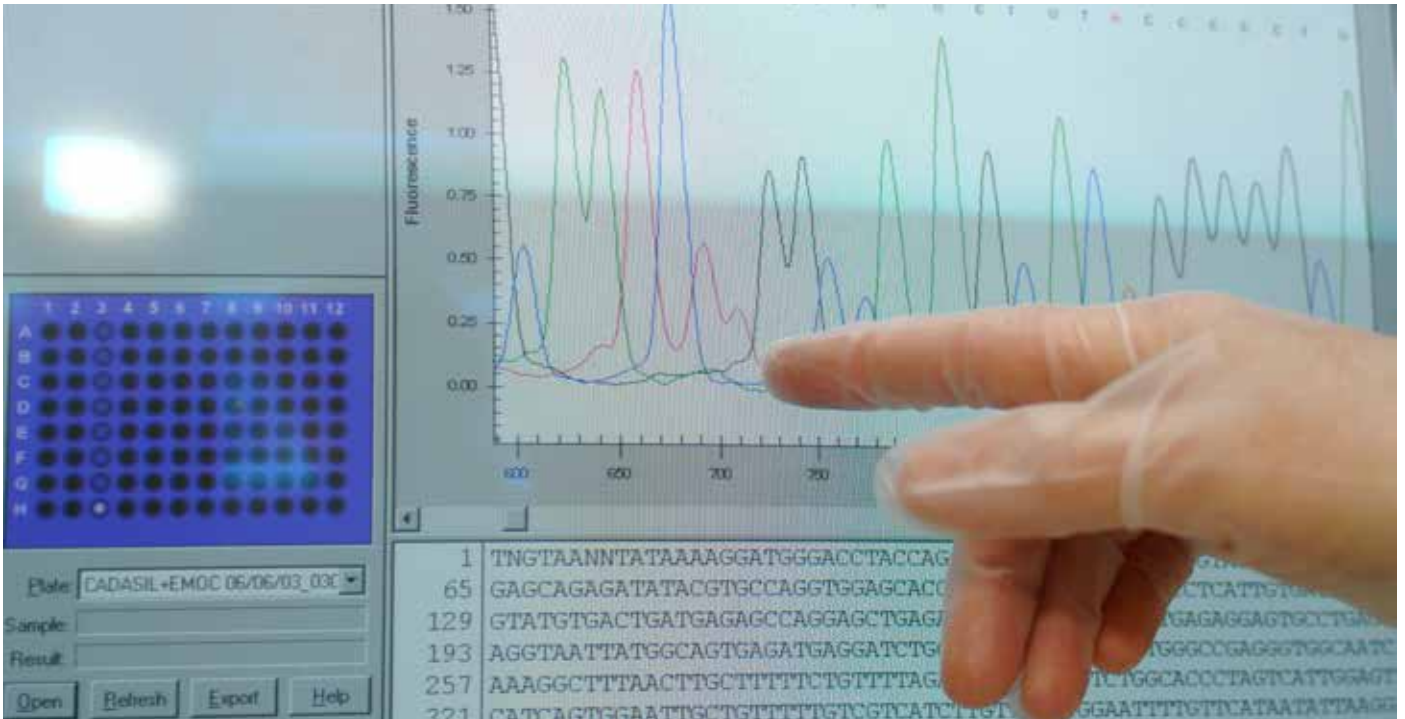
Each biobank that has joined the ecosystem has a module developed by BC Platforms that transmits biobank data to the

service. According to **Timo Kanninen**, Chief Architect at BC Platforms, a common biobank user interface benefits everyone.

"We help pharmaceutical companies find the right biobanks with data significant for them stored. For example, using the search term "asthma" allows you to see how many asthma patients have their data stored in the biobanks of different countries. In the past, it was necessary to e-mail the operator of an individual biobank, ask how many asthma patients they had and then wait for a reply."

The software automatically generates aggregate data, i.e. data collected from multiple sources. As it does not contain personal information, the data can be transferred outside national borders. Identified data of biobanks can be combined in a system once authorisation has been obtained.

"It is possible to conduct smart searches on existing data. The service and ecosystem bring together the data owners, providers and users. Because the users are companies



developing drugs, they often want to define the data they need. Our analysis tools are well-suited for this purpose.”

According to Timo Kanninen, the goal is to have the clinical and genomic data of five million patients under the search functions of the common interface by 2020.

“Now we have a broad view of what kind of data is available. We are constantly recruiting biobanks with genomic data in addition to clinical data into the ecosystem. This benefits drug designers as it allows findings to be verified in another population.”

BC Platforms’ application automatically generates metadata, improving the search results from biobank materials. BC Platforms classifies the metadata based on the existing standards. However, the harmonisation of metadata is still a challenge for efficient data processing. Recording practices vary depending on the country and hospital.

“Age, gender and diagnosis are generally known, but surgeries, operations and laboratory values are often recorded in a non-uniform manner. Different information systems add even more challenges”, says Kanninen.

Companies in the bio-industry will not wait for the results of standardisation if it takes years. They have to come up with their own solutions. However, the harmonisation and standardisation of metadata as well as the provision of public databases

in standard format would be a big relief and resource. ELIXIR aims to this.

### From genetic tests to drug ingredients

Genetic data is increasingly being used in patient care and the industry. The clientele of BC Platforms includes one of the largest companies providing genetic tests in the world, for which BC Platforms produces the genetic data. Finnish research groups utilise the systems of BC Platforms in analysing plant, animal and human genomes. The University of Helsinki conducts, for example, research related to animal breeding and the researchers need tools to manage genomic data. The data analysed with the BC Platforms system is also used to look for new target areas for drugs and to study the efficacy and safety of drug ingredients.

“We digitise the genetic data into a format that researchers can use in their analyses. It can then be combined with other data, such as clinical or patient data”, says **Anita Eliasson**, Director of Administration and Development at BC Platforms.

Genomic data can be utilised in cancer research when determining the patient’s cancer type. Based on genomic data, it is possible to know what the drug response is like and what kind of treatment should be recommended.

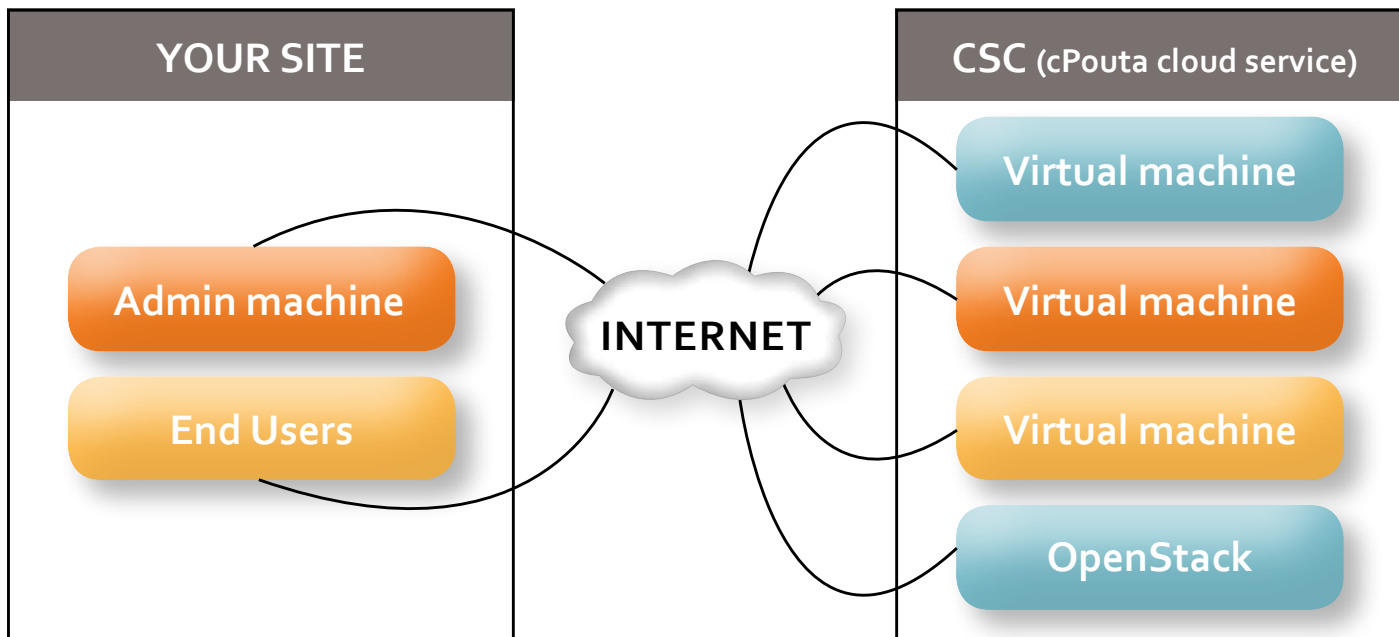
“We use public databases with information on what kind of genome finding typically has specific treatment responses or what type of cancer it is when a person has a certain genome. This is combined with other information. The patient can be treated in the correct way from the start, saving time and money. Being able to select the right medication saves lives.”

Even though the main database system was developed by BC Platforms, Eliasson stresses that BC Platforms is an ecosystem company that places great importance on a partner network.

“We have developed our information systems together with researchers for a long time. Genetic research is now entering a new phase as information is also needed for uses other than research. We are not aiming to provide analysis services for every purpose. That is why our information system has open interfaces. It can then easily be connected with other analysis methods, such as artificial intelligence.”

### Services available as part of the ELIXIR infrastructure

BC Platforms’ two information systems, BC I Genome and BC I Insight, are available in the ELIXIR infrastructure through the Finnish ELIXIR node CSC. Research groups have their own virtual server with BC Platforms’



The BC Platforms databases used by researchers at the University of Helsinki are actually on virtual servers in CSC's environment.

databases and tools. Virtual servers operate on CSC's computing platform and, if necessary, the ePouta cloud service with increased level of information security.

"Researchers can use these to store genomic and other research data while also being able to perform a very wide range of different genome analyses in the same environment by combining data in different ways."

The research environment is currently being used by groups from the University of Helsinki studying animal genes.

"It is possible to connect more applications to this environment because BC I Genome and BC I Insight have open interfaces. When analysing human data, if necessary, the material could be stored in an environment with stricter information security, such as CSC."

Because the processing and combination of data are automated, the research group does not have to perform data conversions or worry about data formats.

"Maintenance is efficient because the environment is consistent. Few research organisations can afford to acquire such a heavy-duty solution and its maintenance for a single research group. This is now possible for bioscientists through the ELIXIR infrastructure."

According to Anita Eliasson, companies like BC Platforms are in great need to utilize replicated public databases, automatically extracting local copies of the database. Bits do not travel quickly enough from the EMBL databases. Physical distance is a factor when it comes to transferring really large data masses.

"Transferring all the data is not sensible. That is why databases should be replicated at the nodes of Finnish ELIXIR. Companies that want to analyse large data masses with artificial intelligence seek out locations that are physically close to the databases due to data transfer costs."

**Ari Turunen**

#### MORE INFORMATION:

[www.bcplatforms.com](http://www.bcplatforms.com)

#### CSC – IT Center for Science

is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.

<http://www.csc.fi>

<https://research.csc.fi/cloud-computing>

#### ELIXIR

builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.

<http://www.elixir-finland.org>

<http://www.elixir-europe.org>

ELIXIR FINLAND

Tel. +358 9 457 2821 • e-mail: [servicedesk@csc.fi](mailto:servicedesk@csc.fi)  
[www.elixir-europe.org/about-us/who-we-are/nodes/finland](http://www.elixir-europe.org/about-us/who-we-are/nodes/finland)

[www.elixir-finland.org](http://www.elixir-finland.org)

ELIXIR HEAD OFFICE

EMBL-European Bioinformatics Institute  
[www.elixir-europe.org](http://www.elixir-europe.org)