Sensitive Data (SD) services for research:
with a few clicks a researcher can launch a personal secure computing environment

Francesca Morello works as Customer Liaison Officer for CSC’s sensitive data services. Morello and her colleagues are developing tools and services to analyse, share and publish data. CSC will also host the Finnish part of the Federated EGA (European Genome-phenome Archive), a distributed network of repositories for sharing human sensitive biomedical data.

SD Connect is a service for collecting and storing sensitive research data during the active phase of a research project while SD Desktop users can directly access and manage that data in a virtual computing environment. The services are accessible via a web user interface, from the user’s own computer.

“Once users upload their sensitive data to CSC, these data are always kept encrypted when stored, transferred or processed within our services. Decryption is done only when data is made available for authorised users within the SD Desktop service”, says Morello.

This cloud-computing environment is easy to use. Morello is very enthusiastic about the fact that accessing the workspace does not require specific technical expertise.

“Researchers can access the workspace with just a few clicks. While SD services are suitable for managing sensitive data from any research field, we are working on further facilitating the use of the services, from fully automating data encryption to streamlining the computing environment customisation.”

The services are available to researchers and students affiliated with Finnish academic organisations, research institutes, and their international collaborators. Using CSC services requires to register a CSC account. While SD Connect and SD Desktop have been designed to facilitate collaboration between organisations, the data is always stored in CSC’s cloud services in Finland.

Human sensitive data and organoids
Seppo Vainio is Professor in Developmental Biology at the University of Oulu, and his research area is organoids. An organoid is a simple and small version of an organ, grown from stem cells. The study of organoids involves the processing of plenty of sensitive data.

“Organoids can be used to model cellular and molecular changes that are either normal or abnormal in terms of organ functioning. The crucial thing is that researchers have developed methods based on the use of human cells to create cells capable of multiple tasks. These can then be used to channelled in various pathways using developmental biology signals. This means we have methods – recipes, if you like – to make cells do different things, such as create organoids that model the normal development of a kidney. We are able to create development models of these and other organs. We also have methods to..."
create similar genetic changes in human stem cells found in the human genome."

According to Seppo Vainio, organoid study is a current scientific megatrend. Now we are able to model human disease processes in a new way. When organoids are combined, we can also experimentally study the interaction of tissues and organs during their formative phase in humans. Organoids are useful in studying human diseases and developing new drugs and treatments.

"We have human stem cell libraries in Europe, and in principle, we could create a stem cell storage of every human being in a biobank. Where necessary, we could then draw on this to create a personal disease model for the study of the health and disease of each person."

Vainio says that as the goal is to have personalised health technology and data, both the stem cell and cell biobank, we have realistic means of finding out how diseases develop. However, all this requires investments in research. Vainio hopes that the Finnish biobank system could be developed in this direction.

One interesting area involves a type of stem cell called an induced pluripotent stem cell (iPSC). Human embryonic stem cells were first grown in the late 1990s, and iPSCs, which are very similar to these, were first generated in 2007. These iPSC lines can be created from, for example, a patient’s skin or blood cells, and they can be programmed to differentiate as desired.

"Because iPSCs originate in individuals, the projects also involve a large amount of processing of sensitive patient material. Our goal is to link observations in organoids better and better into patient records. In this context, the Finnish Social and Health Data Permit Authority Findata offers ways for the utilisation of Finland’s numerous register data system."

For example, the goal with the FinnGen research project study is to gain a better understanding of disease mechanisms and the development of new treatments by combining genomic and health data. It contains the genetic data of more than 500,000 Finns. The data is returned, as agreed, into Finnish biobanks, from which it is freely available to researchers. FinnGen has identified a number of genetic variants related to diseases.

"Researchers have modelled variants identified in organoids experimentally in order to study in more detail genetic changes associated with diseases, or pathogenesis. Once this research is later connected to automated drug screening performed with various chemical libraries and biomarkers, the process will create a foundation to accelerate the development of new treatments with organoids."

**Data is controlled by the research organisations**

Sensitive data include human personal data, ecological data or confidential data. Processing of personal data is regulated by the European General Data Protection Regulation.

"The data controller is an organisation or a legal representative who takes all the decisions on how the data is used. With SD services, we aim at providing all the tools for researchers and their organisations to manage the data access during collection, analysis and reuse," says Francesca Morello.

Processing health or register data for secondary use is strictly regulated by national
laws. SD Desktop is a certified secondary use environment that has been audited against Findata (the Finnish Social and Health Data Permit Authority) regulations. In this case, data access and data exports are managed by Findata and by CSC’s helpdesk. According to Morello, these services are designed so that they provide researchers and data controllers all the instruments to keep their data safe but the services remain flexible and user friendly.

Suitable especially for life science researchers

Sequencing, storing and processing genetic sequences is a time consuming process. As a first step, DNA sequences can be uploaded by the sequencing facility directly into researchers’ workspace in SD Connect. The encrypted data can be easily shared with other researchers, via a URL. When the data collection phase is over, researchers can spin up a virtual computer with SD Desktop and analyse the data stored in SD Connect via data streaming. They can also decide to give read only access to their collaborators from other organisations, for example, to analyse their data together.

When researchers have created their results from their genetic analyses, they can publish their data under controlled access using the Finnish Federated EGA service. In this case, the dataset will be assigned a permanent identifier and the data will be advertised internationally via EGA for reuse. The data remains in Finland while approved researchers can access the data via data streaming using the SD Desktop service.

Only one copy of the same dataset is uploaded to CSC and used during all the different stages of research. The Federated EGA (European Genome-phenome Archive) together with its fully compatible American counterpart dbGAP are the primary global resources for access of sensitive human biomedical data consented for research use.

Ari Turunen

MORE INFORMATION:

University of Oulu
www.oulu.fi/en

CSC – IT Center for Science
is a non-profit, state-owned company adminis-
tered by the Ministry of Education and Culture. CSC
maintains and develops the state-owned, centra-
lised 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 Eu-
ropean 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 2821s ● e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org