



Support International Incoherent Scatter Radars Data Sharing with Global Open Science Cloud

Yin Chen⁽¹⁾, Ingemar Häggström⁽²⁾, Xinan Yue⁽³⁾, Junyi Wang⁽³⁾, Jianhui Li⁽⁴⁾ and Lili Zhang⁽⁴⁾

(1) EGI Foundation, Netherland; yin.chen@egi.eu

(2) EISCAT, Sweden; ingemar.haggstrom@eiscat.se

(3) IGGCAS, China; yuexinan@mail.iggcas.ac.cn, wangjunyi@mail.iggcas.ac.cn

(4) CNIC, China; lijh@cnic.cn, zhll@cnic.cn

Extended Abstract

Sharing and comparing data from different international Incoherent Scatter Radars (ISRs) are important to develop better understanding of global effects of energy inputs into the ionosphere and upper atmosphere. The EISCAT_3D is under design and construction by the EISCAT Scientific Association in Europe, a next-generation ISR consisting of three Phased-Array antenna fields located in the northernmost areas of Norway, Finland and Sweden. EISCAT_3D focuses on the atmosphere at high latitude in the Fenno-Scandinavian Arctic. On the other hand, the Sanya Incoherent Scatter radar (SYISR) is also a Phased-Array radar under design and construction by the Institute of Geology and Geophysics, Chinese Academy of Sciences (IGGCAS). SYISR focuses on the atmosphere at the low latitude in the East Asia region. EISCAT_3D and SYISR share many similarities and are complementary to each other. Collaboration between the two ISRs systems have been established for data interoperability, sharing and analysing.

However, actual sharing of data is a big challenge. Large-scale experimental data will be massively generated at great speeds and volumes by the two Phased-Array ISRs. For example, EISCAT_3D will produce 40PB data per year at its full operations stage in 2023. While for SYISR, the data volume will be 3PB per year in its full operations stage in 2023. To solve the challenge of processing and storing the big data, and how to transfer and share the big data between China and Europe, an federated E-Infrastructure approach is considered.

E-Infrastructures are the combination of digital technology, computational resources and communication to support collaborative research. In recent years, many national- and regional-level e-infrastructure have been built to support large scale scientific research and innovations which are subject to data- and computation-intensive research requiring high capacity that beyond individual institutions or countries can offer. For example, the EGI Federation is one of the largest distributed computing infrastructures federating hundreds of major research data centres in Europe and worldwide, originally built to facilitate LHC experiments in CERN and now is open up for all disciplinary research communities, currently providing access to more than 1.2 Exabyte of research data and 1.2 Million CPU Cores for data processing and analysis needs to thousands of researchers. As another example, CSTCloud is a national infrastructure for Chinese Academy of Sciences scientific communities and China's top research. It federates the national computing resources including, the China Science & Technology Network, China National Grid, and National Scientific Data Centers, currently consisting of 100PB storage, 100G network connection, 300PF HPC computing capacity. However, the two digital infrastructures are not connected which prevent cross-continent data movement and analysis.

In order to create a global digital environment for borderless research and innovation, **a concept of Global Open Science Cloud (GOSC) emerged that aims to connect the different international, national and regional research infrastructures.** Both EGI and CSTCloud are among the pioneers that support the GOSC visions.

A testbed has been established to integrate the Cloud resources from EGI and CSTCloud that allow cross-continent access to computing capacity. EISCAT and SYISR are involved as one of the initial community case studies. Platform and semantic interoperability are two key issues to be addressed in EISCAT-3D & SYISR radar data cooperation, and the joint activities include:

- **Metadata & data federation:** Metadata searchable via an integrated portal, with data stored in the EISCAT repository and the SYISR repository. Such a search portal can be based on the EISCAT Portal (by either configuring the EISCAT Portal to be able to access the SYISR metadata, or by hosting a customised instance of the EISCAT Portal in the CSTCloud) and configuring it to be able to access the SYISR and EISCAT metadata.
- **Federated processing:** After discovering relevant EISCAT and SYISR data, the user could launch analysis jobs in the cloud to process those data. These jobs would run in the European or the Chinese clouds depending on the user's affiliation (e.g., Chinese users would access the CSTCloud portal and run jobs in the Chinese cloud resources, others would use the EISCAT portal and run jobs in the EGI cloud). Other clouds could also be included when necessary.
- **On-demand data movement:** If the data must be moved, the user would be able to move data from EISCAT and SYISR into a single cloud site and perform online analysis with both sets.
- Collaborations on necessary training and activities for SYISR and EISCAT community outreach.