

Conference: EGI Community Forum 2014: Call for Lightning talks

Submitted by: Dr. CHEN, Yin

Title: Towards the Big Data Strategies for EISCAT-3D

Primary Authors:

Dr. CHEN, Yin (Cardiff University) <cheny58@cardiff.ac.uk>

Dr. HAGGSTROM, Ingemar (EISCAT Science Associate) <ingemar.haggstrom@eiscat.se>

Mr. PINTO, Salvatore (EGI.eu) <salvatore.pinto@egi.eu>

Co-authors:

Ms. KRAKOWIAN, Malgorzata (EGI.eu) <malgorzata.krakowian@egi.eu>

Dr. SIPOS, Gergely (EGI.eu) <gergely.sipos@egi.eu>

Dr. SAVOLAINEN, Ville (CSC) <ville.savolainen@csc.fi>

Abstract presenters:

Dr. HAGGSTROM, Ingemar

Brief Description of your work

The design of the next generation incoherent scatter radar system, EISCAT-3D, opens up opportunities for physicists to explore many new research in the field of Earth upper atmosphere studies. On the other hand, it also introduces significant challenges in handling large-scale experimental data which will be massively generated at great speeds and volumes. During its first operation stage in 2018, EISCAT-3D will produce 5PB data per year, and the total data volume will rise up to 100PB per year in its full operations stage in 2023. This refers to so-called big data problem, whose size is beyond the capabilities of the current database and software technology. Since February 2013, a pilot project has been set up within the ENVRI project to help establish a partnership between EISCAT and EGI. The pilot investigation has identified and allocated useful solutions that directly benefit EISCAT-3D and beyond.

Why was EGI important to conduct your research?

The continuous operation with large data volumes of EISCAT-3D demands an e-Infrastructure capable of providing both data storage and computing resources. The volumetric data from the radars are so rich, and can be studied in many different ways, depending on users interest. The data is generally divided into three linked levels, from the raw voltage data, via spectral data to the ionospheric parameters. In addition a fourth level, the value added data of the users, is desired. The computation at the e-Infrastructure is needed to do searches on the different levels of data, going between them and finally to visualize the results. With the worldwide distribution of EISCAT users, EGI is the only one capable of providing these resources in a distributed network.

Which computing resources made available by EGI/NGIs did you use? And how?

The large volume of EISCAT experimental data need to be stored, catalogued and pre-processed to produce previews and extract metadata. The goal is to ease the browsing of the data and the access from the EISCAT community. To do so, EGI provides a Data Catalogue and Dissemination Platform-as-a-Service, which integrates Cloud Computing (Infrastructure-as-a-Service) and Cloud Storage (Storage-as-a-Service). The usage of Cloud resources ensures the flexibility for both computing and storage. The choice to integrate everything into a Platform-as-a-Service, provided by EGI directly to the community, permits EISCAT to use a simple Web Interface to manage the system, upload files, removing the need to directly access the API of the Cloud Storage and Computing service. Regarding the resource providers, Computing and Storage resources are provided by FZJ (DE) and INFN (IT) Federated Cloud sites.

What is the overall impact of your research?

The new data processing and searching strategy will offer more flexible way for EISCAT users to analyse and discover interesting data patterns which are not yet available. Space physicists will be able to make better use of the observation data and exploit the growing wealth of them. This will eventually lead to a new data-centric way of conceptualising, organising and carrying out research

activities which could lead to an introduction of new approaches to solve problems that were previously considered extremely hard or, in some cases, impossible to solve and also lead to serendipitous discoveries and significant breakthrough.

Please provide a list of up to five relevant publications.

1. Chen, Y., Häggström, I., Hardisty, A., Sipos, G., Krakowian, M., Ferreira, N. L., Savolainen, V. (2013). "Towards the Big Data Strategies for EISCAT-3D", *EISCAT International Symposium 2013*, Lancaster, the UK, 2013.
2. Häggström, I., Chen, Y., Hardisty A., Sipos, G., Krakowian, M., Ferreira, N., & Savolainen, V., "Towards the Big Data Strategies for EISCAT-3D", *Radiometenskap och Kommunikation 2013: Generation, Real-Time Processing, Transport, Distribution and Management of Large Raw Data Volumes in the Physical Sciences*. 11 - 12 November 2013, KVA, Royal Academy of Sciences, Frescati, Stockholm, 2013.