**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | EISCAT 3D incoherent scatter radar system | |
| **Vertical (area)** | | Environmental Science | |
| **Author/Company/Email** | | Yin Chen /Cardiff University/ [chenY58@cardiff.ac.uk](mailto:chenY58@cardiff.ac.uk)  Ingemar Häggström, Ingrid Mann, Craig Heinselman/  EISCAT Science Association/ {[Ingemar.Haggstrom, Ingrid.mann, Craig.Heinselman}@eiscat.se](mailto:Ingemar.Haggstrom,%20Ingrid.mann,%20Craig.Heinselman%7d@eiscat.se) | |
| **Actors/Stakeholders and their roles and responsibilities** | | The EISCAT Scientific Association is an international research organisation operating incoherent scatter radar systems in Northern Europe. It is funded and operated by research councils of Norway, Sweden, Finland, Japan, China and the United Kingdom (collectively, the EISCAT Associates). In addition to the incoherent scatter radars, EISCAT also operates an Ionospheric Heater facility, as well as two Dynasondes. | |
| **Goals** | | EISCAT, the European Incoherent Scatter Scientific Association, is established to conduct research on the lower, middle and upper atmosphere and ionosphere using the incoherent scatter radar technique. This technique is the most powerful ground-based tool for these research applications. EISCAT is also being used as a coherent scatter radar for studying instabilities in the ionosphere, as well as for investigating the structure and dynamics of the middle atmosphere and as a diagnostic instrument in ionospheric modification experiments with the Heating facility. | |
| **Use Case Description** | | The design of the next generation incoherent scatter radar system, EISCAT\_3D, opens up opportunities for physicists to explore many new research fields. 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. This challenge is typically referred to as a big data problem and requires solutions from beyond the capabilities of conventional database technologies. | |
| **Current**  **Solutions** | **Compute(System)** | | EISCAT 3D data e-Infrastructure plans to use the high performance computers for central site data processing and high throughput computers for mirror sites data processing |
| **Storage** | | 32TB |
| **Networking** | | The estimated data rates in local networks at the active site run from 1 Gb/s to 10 Gb/s. Similar capacity is needed to connect the sites through dedicated high-speed network links. Downloading the full data is not time critical, but operations require real-time information about certain pre-defined events to be sent from the sites to the operation centre and a real-time link from the operation centre to the sites to set the mode of radar operation on with immediate action. |
| **Software** | | * Mainstream operating systems, e.g., Windows, Linux, Solaris, HP/UX, or FreeBSD * Simple, flat file storage with required capabilities e.g., compression, file striping and file journaling * Self-developed software   + Control & monitoring tools including, system configuration, quick-look, fault reporting, etc.   + Data dissemination utilities   + User software e.g., for cyclic buffer, data cleaning, RFI detection and excision, auto-correlation, data integration, data analysis, event identification, discovery & retrieval, calculation of value-added data products, ingestion/extraction, plot   + User-oriented computing   + APIs into standard software environments   + Data processing chains and workflow |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | EISCAT\_3D will consist of a core site with a transmitting and receiving radar arrays and four sites with receiving antenna arrays at some 100 km from the core. |
| **Volume (size)** | | * The fully operational 5-site system will generate 40 PB/year in 2022. * It is expected to operate for 30 years, and data products to be stored at less 10 years |
| **Velocity**  **(e.g. real time)** | | At each of 5-receiver-site:   * each antenna generates 30 Msamples/s (120MB/s); * each antenna group (consists of 100 antennas) to form beams at speed of 2 Gbit/s/group; * these data are temporary stored in a ringbuffer: 160 groups ->125 TB/h. |
| **Variety**  **(multiple datasets, mashup)** | | * Measurements: different versions, formats, replicas, external sources ... * System information: configuration, monitoring, logs/provenance ... * Users’ metadata/data: experiments, analysis, sharing, communications … |
| **Variability (rate of change)** | | In time, instantly, a few ms.  Along the radar beams, 100ns. |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | | * Running 24/7, EISCAT\_3D have very high demands on robustness. * Data and performance assurance is vital for the ring-buffer and archive systems. These systems must be able to guarantee to meet minimum data rate acceptance at all times or scientific data will be lost. * Similarly the systems must guarantee that data held is not volatile or corrupt. This latter requirement is particularly vital at the permanent archive where data is most likely to be accessed by scientific users and least easy to check; data corruption here has a significant possibility of being non-recoverable and of poisoning the scientific literature. |
| **Visualization** | | * Real-time visualisation of analysed data, e.g., with a figure of updating panels showing electron density, temperatures and ion velocity to those data for each beam. * non-real-time (post-experiment) visualisation of the physical parameters of interest, e.g.,   + by standard plots,   + using three-dimensional block to show to spatial variation (in the user selected cuts),   + using animations to show the temporal variation,   + allow the visualisation of 5 or higher dimensional data, e.g., using the 'cut up and stack' technique to reduce the dimensionality, that is take one or more independent coordinates as discrete; or volume rendering technique to display a 2D projection of a 3D discretely sampled data set. * **(Interactive) Visualisation**. E.g., to allow users to combine the information on several spectral features, e.g., by using colour coding, and to provide real-time visualisation facility to allow the users to link or plug in tailor-made data visualisation functions, and more importantly functions to signal for special observational conditions. |
| **Data Quality** | | * Monitoring software will be provided which allows The Operator to see incoming data via the Visualisation system in real-time and react appropriately to scientifically interesting events. * Control software will be developed to time-integrate the signals and reduce the noise variance and the total data throughput of the system that reached the data archive. |
| **Data Types** | | HDF-5 |
| **Data Analytics** | | Pattern recognition, demanding correlation routines, high level parameter extraction |
| **Big Data Specific Challenges (Gaps)** | | * High throughput of data for reduction into higher levels. * Discovery of meaningful insights from low-value-density data needs new approaches to the deep, complex analysis e.g., using machine learning, statistical modelling, graph algorithms etc. which go beyond traditional approaches to the space physics. | |
| **Big Data Specific Challenges in Mobility** | | Is not likely in mobile platforms | |
| **Security & Privacy**  **Requirements** | | Lower level of data has restrictions for 1 year within the associate countries. All data open after 3 years. | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | EISCAT 3D data e-Infrastructure shares similar architectural characteristics with other ISR radars, and many existing big data systems, such as LOFAR, LHC, and SKA | |
| **More Information (URLs)** | | <https://www.eiscat3d.se/> | |
| **Note:** <additional comments> | | | |