Kiruna November 21, 2018

EISCAT_3D Data Management Plan

Version: 2018-11-20

Introduction

The EISCAT_3D system represents a significant upgrade to EISCAT's incoherent scatter radar (ISR) capabilities. The system is designed with a great deal more flexibility than the present EISCAT ISRs can provide, largely through the use of signal digitization very early in the processing chain and the use of a very large number of smaller antennas. Internally, this represents an extremely large amount of data bandwidth, on the order of a few terabytes per second. The data streams from the individual antennas are, however, of little use unless they are combined and this combination can be accomplished at various stages in the system, thus greatly affecting the load on the data processing and communication subsystems. Data management and data processing management are, as a result, important considerations in the system architecture and operation.

ISRs are primarily used for probing the near-Earth plasma environment, though they have capabilities in other areas such as meteor research and radio astronomy. The desired performance capabilities have been compiled by a large number of researchers and documented in a science case (EISCAT_3D Science Case, EISCAT Preparatory Phase Project WP3), scientific publications (e.g. The science case for the EISCAT_3D radar, McCrea et al., Progress in Earth and Planetary Science, 2015), and proposals to funding agencies. As an environmental research infrastructure, the envisioned operating lifetime for the radar is on the order of at least 25 years. Experience with other ISRs has shown that systems must be upgraded significantly over the decades of their use to adapt to changes in scientific understanding of the environment and to take advantage of technological advances. As a result, one goal for EISCAT_3D implementation is that architecture decisions not limit upgrade potential except where absolutely necessary. On the other hand, the initial configuration need not be maximally flexible as long as it supports the current science cases. This allows the implementation to limit costs in areas for which affordable technologies are not presently available while maintaining options for future upgrades as technology costs come down.

One overall architectural decision was made during the European Commission-funded fp7 Preparatory Phase project: that the antennas be combined into physical subarrays of 91 antennas and that these subarrays need have no more than 10 parallel data stream outputs at 30 MHz of bandwidth without compromising the final measurements. Similarly, at any given site the data streams from subarrays need be combined into a total of no more than 100 parallel data streams without compromising the radar capabilities, at least for ISR research goals. At the next level of the architecture, the measurement results from all three sites in the Stage 1 implementation must be combined to provide the unique vector measurements that EISCAT_3D will be capable of. Precisely which data streams are needed from each site is a potential degree of freedom in the implementation and one that can easily evolve over the life of the infrastructure.

Headquarters
P. O. Box 812
SE-981 28 Kiruna, Sweden
Phone: +46 980 79150

Kiruna Site
P. O. Box 812
SE-981 28 Kiruna, Sweden
Phone: +46 980 79062

Sodankylä Site
Tähteläntie 54B
FIN-99600 Sodankylä, Finland
Phone: +358 40 8669096

Tromsø Site
Ramfjordmoen
N-9027 Ramfjordbotn, Norway
Phone: +47 776 20730

EISCAT Svalbard Radar
P. O. Box 432
N-9171 Longyearbyen, Norway
Phone: +47 776 25270

An additional consideration for EISCAT_3D data relates to the potentially sensitive nature of data accidentally captured during the course of normal operations. In particular, the data streams might contain returns from operational satellites, etc., which might be considered classified in certain countries. In order to minimize this risk, EISCAT has implemented a set of Data Policy Procedures. The present version of these procedures is contained in Appendix A.

Baseline Requirements

The core requirements for EISCAT_3D include support for vector ionospheric measurements at time resolutions ten times shorter and spatial resolutions ten times smaller than those available with the present EISCAT systems. Additionally, EISCAT_3D must support vector measurements for the entire ionosphere in the transmitter antenna pointing direction. The transmitter must be able to steer on a pulse-to-pulse basis toward any direction less than 60 degrees from zenith and the receivers must be able to follow that steering, though quasi-real time decisions on sets of steering directions can be made with second time scales (as opposed to millisecond time scales for pulse-to-pulse steering).

An additional aspect of the realization of these capabilities is the need (or lack of need) for real-time processing of the returns. Most of the anticipated use cases do not require optimized nor immediate results, which then add considerably to the processing requirements. Some use cases do require rapid estimates of plasma parameters (e.g. in support of launch decisions for sounding rockets), but those can normally be supported by estimates at either lower time cadences or with reduced spatial resolutions. One class of measurements that does require rapid decisions is meteor head echo vector velocity determination because it can affect the system's ability to capture the bulk of the meteoroid trajectory through beam steering. This case, however, is not especially compute intensive, at least not for first order estimates. In each of these cases, the most intensive processing can be carried out at each site individually with the greatly-reduced results combined to form the vectors with minimal data communication bandwidths or computation loads.

A large class of experiments can also be performed with a greatly reduced receive bandwidth. Most experiments utilize either the ISR ion line (with a few KHz spreading beyond the transmitter modulation bandwidth) or a Doppler-shifted hard target return with bandwidth much less than the modulation bandwidth. For these kinds of experiments, a total bandwidth per site of approximately 6 MHz should be sufficient, even for the raw measurements. A class of experiments that this does not support, however, is those making use of the plasma line of the ISR spectrum. Plasma line measurements may require up to the full 30 MHz of bandwidth, though the plasma lines at any given altitude are expected to be quite narrow (and thus amenable to more clever detection/processing schemes).

These considerations led to plans for the EISCAT_3D implementation that started with reduced-bandwidth processing at each site. In particular, independent 6-MHz channels were considered for the outputs of the subarrays, noting that this would still allow plasma line measurement using several of the 10 parallel data streams at offset frequencies. The subarray hardware must support the full 30 MHz, but the second stage beamformer, of which there is one at each site, could use less capable systems initially and later be upgraded to faster compute hardware (an upgrade that was included in the budgets).

In addition to processing speed and bandwidth, data archival must be considered. Even for the baseline requirements, the overall data volumes will be considerably greater than those for the existing EISCAT systems (at least a factor of 100 greater, probably somewhat more than that). Even during commissioning EISCAT will need quite large storage and archival capacity.

Expansion Potential

Ideally, the EISCAT_3D system will evolve into one that combines raw data streams from all subarrays into one central processing location. Such a configuration would allow maximum flexibility in extracting physical parameters from the data for presently-planned and as-yet-undetermined scientific applications. This evolution can be readily supported as long as it doesn't require extensive and expensive hardware upgrades throughout the system. In particular, it is important to maintain the full capabilities in the subarray hardware and not throttle those data sources unnecessarily, because this would lead to high replacement costs. Additionally, it is important to implement sufficient data bandwidth between the individual sites and the internet backbone/data archive site. Even this connection might be readily upgraded under regular budgets at some future time, but when such infrastructure is being put in place it makes sense for the fibres to support the full future system needs, to the extent possible.

In a similar vein, the longer-term evolution of EISCAT_3D may require significant increases in processing power at a location where data from all sites is combined. Work carried out in the NeIC EISCAT_3D project suggests that processing loads as high as 500 Tflops might occasionally be useful. This extreme amount of compute capacity is likely to be needed only rarely and probably not for real-time applications initially, so there are ways to greatly reduce the peak loads by spreading the computations out over much longer times. Nonetheless, as processing capacity becomes less expensive it may make sense to have the full performance available for things like reprocessing of large sets of historical EISCAT_3D data. This is obviously a long-term goal and not something that must be implemented immediately.

Implementation

The planned implementation approach has EISCAT_3D purchasing the general-purpose computer hardware as late as possible. This should allow the project to obtain the best price/performance ratio and, given the delays in starting the project, may allow it to fully realize a 30 MHz bandwidth in the second-stage beamformer from the start. It also allows EISCAT to evaluate the budgetary situation and utilize any savings on purchases of hardware for enhanced processing capacity. Similarly, the archival and storage capacity will be purchased or leased fairly late in the implementation to ensure maximal benefit from overall trends in the industry. Several data archival volumes are possible, from fairly narrow bandwidth products of fully beam-formed returns to full bandwidth products from individual subarrays. The specific volume of data archived can thus be tailored to the funding available and, later, adjusted upward for maximum flexibility. Even the narrow bandwidth, beam formed data fulfil the Baseline Requirements but they may not support future, more exploratory, demands.

Data Curation

As an organization, EISCAT is required to retain and curate the data collected with its systems. The general data policy is spelled out in the EISCAT Data Policy appendix to the Blue Book. Though this document specifically calls out the present radars, as indicated in Appendix 2 of the EISCAT AGREEMENT portion of the Blue Book, these policies will also apply to EISCAT_3D. While the EISCAT_3D volumes will be larger than those for the presently-operated ISRs, the general principles do not change. EISCAT retains ownership of all data produced, stores as much Level 1 data (raw samples) as finances will support. EISCAT also provides access to Level 2 data for exclusive use of the requesting user(s) for a period of 1 year, provides access to Level 2 data for all EISCAT members for the following 2 years and makes the data publically available after that time. EISCAT archives Level 3 data (extracted plasma parameters) and retains them for the exclusive use of the user(s) that request the data for a period of 1 year (unless released by the user(s) for public access).

For the longer term, EISCAT's plan is to ensure that the Level 3 data are available in perpetuity. The volume of Level 3 data will be manageable and affordable, even with current storage technologies. Were EISCAT to dissolve as an organization, it would work to find a new curator/owner for these data sets so they could remain available to the global scientific community.

As indicated above, EISCAT will also attempt to retain as much Level 1 and Level 2 data as is practical and affordable. At present this retention would come at a not insignificant cost. As a result, EISCAT cannot yet *guarantee* that these lower level data products can be retained after the organization ceases to exist. If a suitable organization can be found to take ownership of the data, however, EISCAT plans to transfer that ownership with the permission of Council.

Appendix A - Data Policy Procedures

Background

The EISCAT Statutes include an appendix spelling out the EISCAT Data Policy. That policy indicates areas that require special attention to ensure that EISCAT systems adhere to the goals of the Association, but it does not prescribe detailed procedures for avoiding or handling potentially politically sensitive observations. This document describes the procedures implemented by EISCAT to adhere to the Data Policy.

The Data Policy originated largely from recommendations made by the EISCAT Council's Expert Group on Satellite and Space Debris Observations. The expert group was formed in response to concerns about the future capabilities of the EISCAT_3D radar and the potential for abuse of this new system, either accidentally or intentionally.

The EISCAT Data Policy indicates, in section *4. Transparency and Logging of EISCAT Operations*:

> *1. EISCAT shall strive to have full transparency in its operations and with respect to the data generated. All observation campaigns shall be clearly documented and the campaign log shall be available for inspection in accordance with the EISCAT Agreement.*
> *2. Generally, data segments containing radar echoes from resident space objects shall be filtered out at a low processing level. When such filtering is not adequate to reach the objectives of the approved observation campaign, special care must be taken to avoid a breach of the Objects and Means of the association as laid down in the EISCAT Agreement.*
> *3. EISCAT raw data containing radar echoes from satellites shall not be distributed to other agencies.*

The particular Objects and Means in question are as follows:

> *a. The aim of the Association is to provide access to radar, and other, high-latitude facilities of the highest technical standard for non-military scientific purposes.*
> *e. The Association may contribute to the international task of tracking objects in space (natural or man-made). For this activity, an agreed list of objects shall be maintained and the Association shall only conduct tracking of objects from this list.*
> *k. All use of observation time must be in line with the aims of the Association. Users shall not use the facilities for collecting data on military sensitive objects.*
> *l. All data obtained shall become the property of the Association and shall be managed according to the EISCAT data policy.*

Procedures

While no operational procedure can be 100% effective in avoiding the collection and dissemination of data containing classified or sensitive objects, especially not in an environment where state-of-the-art radar and signal processing techniques are being advanced, the following overlapping measures greatly reduce such risks. The expectation is that these procedures will minimize the risk without significantly impacting the scientific results from the measurements. The EISCAT Council's Expert Group on Satellite and Space Debris Observations specifically noted, in their report from October 2015: "*Although some high-precision measurements of satellite range and velocity can be made using the EISCAT radar, the capabilities of the present system in terms of orbit accuracy are such that tracking capabilities can be considered rudimentary compared to the open Space-Track catalogue. The statistical surveys that can be done with EISCAT are not restricted*." As a result, some of the Data Policy Procedures are not implemented on the legacy EISCAT systems (particularly those procedures requiring specialized technical solutions). Other procedures are, however, more generally applicable for all observations.

1) The first level of protection comes from the experiment request procedures. This is a web-based process where experimenters are required to explain the purpose(s) of their measurement campaign(s) and the specific measurement techniques (e.g. pulse sequences, etc.) to be employed. This vetting of the experiment requests allows EISCAT to block users who unknowingly wish to attempt prohibited measurement campaigns. The vetting process will occur in two stages. In the first stage, experiments with no apparent conflicts with the aims of the Association can be simply approved. This includes operations for basic ionospheric measurements using standard pulse coding and processing schemes and measurement modes not amenable to characterizing classified or sensitive space objects. If there is any question about the acceptability of the goals of an experiment, the request will be passed on to the second stage.

In the second stage, the experiment request will be evaluated by a group of radar experts, called the Experiment Review Panel, to determine whether the experiment goals/capabilities are at odds with EISCAT's mandates. The panel will be chaired by the EISCAT Head of Operations and may include experts from outside the organization. This second stage evaluation may require additional information/clarification from the person(s) requesting the measurement(s). In the case of an experiment aimed at measuring the orbital parameters of one or multiple known satellites or other kinds of objects in space, the details of the target satellite(s) will be provided by the experimenter(s) and the satellite(s) will be compared to a list of approved objects (White List). For objects not on the White List at the time of the request, it is up to the requestor to justify an addition and demonstrate that the requested object is allowable. In any event, the Experiment Review Panel's analysis must indicate that the experiment does not conflict with EISCAT's mandate before said experiment can be performed.

The White List will be openly available and initially based on the Open Space-Track Catalog (www.space-track.org). Any Associate or Affiliate can anonymously request to have an object removed from the White List. Non-manmade objects (e.g. near earth objects) can generally be added to the White List with minimal review. Ablating meteoroids are automatically on the White List and are identified primarily by their altitudes (below operational satellite altitudes).

2) The second level of protection comes from the automated removal of hard-target returns from the low level data stream. This protection level is specific to EISCAT_3D and it entails the detection and removal of returns from the output data stream of each subarray of 91 antennas. The removal is only done for altitudes above 200 kilometers (thus leaving most meteor head echoes unaffected) and for Doppler velocities less than the escape velocity of 11.2 km/sec (thus ignoring objects that cannot be in Earth orbit). The algorithms employed are based on standard matched-filter processing and the threshold is set to ensure a satisfactorily low false alarm rate while detecting the vast majority of actual targets. The specific objects of greatest interest for removal are, in any event, operational satellites and, as such, tend to have relatively large radar cross sections. Thus, if the algorithms fail to detect some objects, those objects will be primarily smaller pieces of space debris, etc. The algorithms will be periodically tested with objects of various cross sections from the White List and test reports will be available on line.

Some experiments will require that this automatic removal be disabled. In particular, it cannot be used for experiments that target objects on the White List or, for instance, searches for meteoroids prior to their interaction with the atmosphere or near earth objects (asteroids). When the automatic removal algorithm is disabled, the metadata for the radar returns will indicate this via dedicated tags. Reports will be sent to EISCAT Council periodically to specify which experiments and how much total time was used with the automatic removal disabled.

3) The third level of protection takes the form of full transparency of all EISCAT operations. No data will be collected without the details being added to an open and publicly accessible campaign log. This log will be automatically generated and updated during operations. Metadata, which describe operations as they occur (e.g. radar mode, antenna steering, etc.), are also openly available. Furthermore, all sites are available for inspection by the relevant agencies in the hosting countries. When practical, they will also be available for inspection by non-hosting countries, though this may be subject to local law (e.g. the need for travel visas, etc.). Inspections will have access to both the physical resources (after safety concerns have been taken into account) and the software source code and data.

If any operations are discovered that violate the aims of the association, experiments can be immediately terminated by EISCAT and data can be made inaccessible.

4) The fourth level of protection comes from the fact that all data collected on EISCAT systems remain the property of the Association. Raw data shall not be disseminated to other agencies, either automatically or individually. Raw data can be accessed by the scientists who requested the experiments or, after the embargo period, by all Associates and Affiliates, but only processed results will be made openly available. Scientists may, at times, need to cache data from their experiments at their home facilities. These temporary caches are allowed if reasonable precautions are taken to ensure that the raw data are not disseminated further. Users must also ensure that their caches of EISCAT data are purged after use and that they, in any event, are not retained longer than two years unless explicitly permitted by EISCAT, in writing, and under extraordinary circumstances (e.g., to support PhD students during their university studies).