COOPEUS Deliverable 2.3

Cooperation on incoherent data practices and formats: Report on EC-US workshop Incoherent Scatter Data Practices

Prepared by EISCAT Scientific Association

November 2013

1. Progress of Discussions

The discussions between EISCAT and their US colleagues started with two face-to-face meetings in Kiruna and Skype conversations during the spring of 2013. The outcome of these ongoing discussions was later presented for discussions with a broader user community during two meetings: the EISCAT_3D user meeting (held in May 2013 in Uppsala, Sweden) and the EISCAT International Symposium (held in August 2013 in Lancaster, United Kingdom).

The discussions mentioned above were made with colleagues at MIT Haystack Observatory and their contribution was central for the activities within Work Package 2. The largest part of the discussions were made using Skype video-conferencing. The face-to-face meetings with the MIT colleagues took place when they were visiting Kiruna for other planned activities. Additionally another face-to-face meeting took place at Haystack in September 2013, just prior to the annual COOPEUS meeting in Boulder. Colleagues from SRI, that had eventually received limited funding for the work, did not participate in the earlier discussions but attended this meeting. On the meeting it was agreed to follow the same line of progress as was earlier outlined with the MIT group. The future plans were concretized.

Further meetings are planned to take place in 2014, and the results will be presented to broader communities at a variety of workshops. The first of these workshops will be the EISCAT_3D user meeting in May 2014, discussions are also planned for the US Cedar Workshop in June 2014 and the Incoherent Scatter Radar School will be held at the Arecibo Radio Observatory in Puerto Rico. The outcome of these discussions will be a set of suggestions for international data format standards and specifications to be used by incoherent scatter radar facilities world-wide.

On further joint activities, we have discussed a couple of use cases. One on a collaborative international radar school to educate students using both EISCAT and US radars and another one on a response to a geophysical trigger, operate AMISR and EISCAT (at least) together to create an integrated ionospheric image. These will be substantiated in the coming meetings, primarily between SRI and EISCAT and then discussed in a broader user community.

Three types of data have been discussed between EISCAT and the US colleagues: Raw samples, Correlated data, (intermediate) and Analysed data. Based on the information summarized in the Gap Report (COOPEUS Deliverable 2.2.) and on earlier discussions it was decided to concentrate the discussion of the intermediate level data (see Fig. 1, below). The form of these data is the most convenient for exchange and storage, with a total data volume that is manageable while still retaining sufficient information for further advanced analysis. As a basis of the discussions we agreed to use a document on RF Signal Format that was prepared by Frank D. Lind from MIT Haystack Observatory within the Open Radar Initiative[1]. The content of the data is described in the next section, a subsequent section discusses the format.

2. Data Content

The content of the data of the incoherent scatter measurements is summarized in the RF signal objects described in the Open Radar Initiative document. They are: source information, signal information, conversion information, coherence information, source format, signal information and signal processing chain and are described below.

2.1 Source information

- The name of the sensor which is the source of the RF signal. This is a descriptive name used to identify the source from where the signal originates.
- A description of the type of sensor which generated the RF signal.
- A description of the originating sensor.
- A context specific identifying number associated with the source. This is a unique number associated with a particular sensor or zero if such a number is not available.
- Decimal latitude of the RF signal. This is most typically the physical location of the antenna. The reference frame is specified separately.
- Decimal longitude of the RF signal. This is most typically the physical location of the antenna. The reference frame is specified separately.
- Altitude of the RF signal. This is most typically the physical location of the antenna. The reference frame is specified separately.
- Name of the physical reference frame used for the coordinates. Typical examples are 'rectangular' or 'WGS84'.
- Description of the coordinate reference frame.

2.2. Signal information

- The central frequency of the sampled RF signal bandwidth prior to any analog down conversion, aliasing, or digital frequency translation.
- The central frequency of the RF signal bandwidth after analog down conversion but prior to aliasing and any digital frequency translation. This would typically be the final IF output of an analog receiver stage. When no down converter is used this is the same as the RF center frequency.
- The central sampling frequency of the signal after analog down conversion and aliasing but prior to digital frequency translation. This is the aliased IF frequency as seen at the A/D converter when IF subsampling is used. When no aliasing is used this is the same as the IF center frequency.
- The central frequency of the RF signal after analog down conversion, aliasing, and digital frequency translation. For many modern systems this will bebaseband (0 Hz). Some systems may use a non-zero frequency such as a fixed carrier offset.
- Indicates that the signal spectrum has been inverted by the net analog and digital translation process. Spectral flips in analog (high side versus low side injection) and digital (aliasing) mixing are common. This indicates the sign of the Doppler shift is opposite of the normal convention (i.e. upshift moving towards receiver, downshift away). This can be useful when correcting the sign of the data is not desired for some reason.
- Approximate -3dB bandwidth of the RF signal after any digital downconversion and filtering.

[1] (www.openradar.org/cgi-bin/openradar.cgi/Documents/OpenRadarRFSignalObject)

- Approximate -60 dB ultimate stop bandwidth of the RF signal after any digital downconversion and filtering.
- Ratio of the sampling bandwidth to the RF bandwidth of the signal.
- Sampling period of the RF signal at this stage in seconds.
- The estimated net transfer function response of the channel prior to this signal. Represented as an autocorrelation function with a finite number of lags and a fixed lag spacing. A one with all zeros if unknown.
- The complex amplitude and phase error associated with the transfer function response estimate. All zeros if unknown or a mathematical estimate.
- The time spacing of lags in the impulse response estimation.
- Description of the transfer function and how it was determined.

2.3. Signal Processing Chain

- Software stage number in the overall signal chain which produced this object. The count starts at zero which represents the data acquisition program.
- Names of the software associated with each stage in the signal chain up to the production of this object.
- Descriptions of the signal processing stages for each stage in the signal chain up to the production of this object.
- Revision tags of the software associated with each stage in the signal chain up to the production of this object.
- Configuration file names for the software used at each stage in the signal chain up to the production of this object.
- Configuration file revision tags for software used at each stage in the signal chain up to the production of this object.

2.4. Conversion information

- Name of the converting device that translated from the analog to digital domain.
- Type of the converting device that translated from the analog to digital domain. (e.g. "Echotek GC314FS digital receiver", "Ettus USRP2", etc...
- Model information for the converting device.
- Name of the analog path prior to A/D conversion.
- Description of the analog path prior to A/D conversion.
- Gain in dB of the analog path prior to A/D conversion.
- Compression point in dBm of the analog path.
- Third order intercept point in dBm of the analog path.
- Name of the A/D converter chip used in the conversion.
- Number of bits provided by the A/D converter.
- Expected spur free dynamic range of the A/D converter.
- Full scale clipping point of the A/D converter. This is the clipping point of the ADC card and not the analog path prior to it.
- Sampling rate at the A/D converter for the RF signal. This may or may not be the sampling rate of the data. With a digital downconverter this is the input A/D converter clock rate which is then downconverted and output at a lower final rate.

2.5. Coherence Information

- String representation of the UTC second for the first sample of the first frame. Must conform to a compliant ISO8601 representation.

- Unix long second representation of the UTC for the first sample of the first frame. From the epoch 1 January 1970 at 00:00 UTC.
- Alignment from the synchronization boundary on which the first sample of the first frame of data starts. For UTC alignment this would be relative to the UTC PPS mark. Other alignments should be described. The offset may not be zero for some kinds of alignment. For example if a frame of samples starts at a fixed delay from a UTC PPS signal boundary.
- Expected error of the alignment information. Zero if unknown.
- Description of the selected alignment boundary for the data.
- Description of the type of synchronization used (i.e. UTC, trigger, approximate, etc.)
- Descriptive synchronization status (e.g. locked, unlocked, semi-coherent, etc)
- Name of the reference source used for the UTC second alignment.
- Description of the source used for the UTC second alignment.
- Name of the A/D converter clock source.
- Description of the A/D converter clock source.

2.6. Signal Description

- A string describing the signal and its representation. (e.g. PRI frame radar data, CW passive radar data, etc).
- Description of RF signal normalization (e.g. self, physical, arbitrary).
- Factor to normalize a sample by to obtain physical units of volts as a multiplicative constant. Defaults to 1.0 for non-physical normalizations. Assumes a linear normalization model.
- Factor in volts to offset a sample by to obtain zero DC offset in volts. Defaults to 0.0 for non-physical normalizations.

2.7. Signal Data

- A number indicating the relative position of this RF signal object in the overall sequence of such objects for a particular RF object stream. This is a relative value useful for detecting sequence errors or gaps when traversing a sequence of RF signal objects.
- Real samples describing the data as a vector of IQ sample data that have been converted to floating point values. This is the real portion of that vector which represents the I portion of the complex signal. The data are left in sample value units and the normalization information should be used for conversion to volts where needed. A series of frames of a fixed size are provided. All frames organized in a single RF signal object have equal size and the same type. Thus they are effectively a fixed sized table of data in a given object. NaN is valid for partial frames, fields, or rasters where data is not available.
- Imaginary samples describing the data as a vector of IQ sample data that have been converted to floating point values. This is the imaginary portion of that vector which represents the Q portion of the complex signal. The data are left in sample value units and the normalization information should be used for conversion to volts where needed. A series of frames of a fixed size are provided. All frames organized in a single RF signal object have equal size and the same type. Thus they are effectively a fixed sized table of data in a given object. NaN is valid for partial frames, fields, or rasters where data is not available.
- Timestamp in seconds of the start sample of each data frame relative to the coherence boundary. This is to allow gaps between frames and computation of the time within a frame. Gaps between frames may vary to allow for different sampling intervals to be stored that are separated in time by different amounts.

- Timestamp fraction in picoseconds of the start sample of each data frame relative to the coherence boundary.

3. Data formats

We have also discussed an extension to go deeper into the format for the future sampling type of data, in with the data are a continuous stream regardless of the radar transmission. Here the VITA 49 protocol[2] seems to become a future standard. This development will be closely followed. We are now on the way to develop a format fit for distribution for our present HF radar, and will use that as an example for either or both of the above formats.

For the highest level of data, the ionospheric physical quantities contained in the joint Madrigal database, we will work closely to ease the data access for a broader community. Here, we plan a new release in 2014, with a new HDF-5 data format (MIT), a simpler and better user interface (Jicamarca/Cornell) and added extensions to other portals (EISCAT).

We have started some data interchange activities of low levels of data between SRI and EISCAT, and are investigating the possibilities to adopt the EISCAT analysis software to be used together with the SRI HDF-5 format. We are planning a face-to-face meeting with SRI in December 2013 to discuss the preliminary outcome of these activities.

The AMISR hdf-5 interchange format (see Table 1) has in principle all of the components needed by generic incoherent scatter analysis programs, like the one used at EISCAT, GUISDAP. It is one hdf structure, with 9 main groups for coding, integration, receiver, signal, experiment setup, site, time and transmitter parameters. Now, in Europe there are properties that are used for the analysis, which the US radars has other parameters for. Examples of this are the ambiguity functions, which in GUISDAP is mainly the Frequency Ambiguity Function, whereas the US data provides Lag Ambiguity Functions, being the transforms of each other. The radar 'constant' is something GUISDAP calculates from the antenna parameters, but the AMISR radar have different constants for different beam directions. So, the 'constant' is not really a constant anymore, and it's not feasible to calculate is in the analysis phase. The AMISR data provides the constant for each measurement profile, and in the EISCAT_3D case, this is something also the European side have to do.

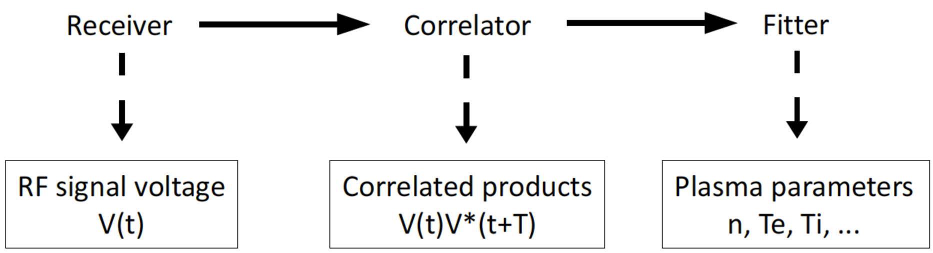

Figure 1: Typical flow from the signal of ISR observations shown and discussed in the gap analysis report (COOPEUS Deliverable 2.2). The considerations of the present report now refer to the intermediate level (Correlated Products). This figure is from the Gap Analysis Report.

---

[2] (wdv.com/Electronics/Reference/SDROoverWEB-VITA.pdf)

Table 1: Example - HDF5 d0311385.dt0.h5

```
Group '/'
   Group '/IncohCodeFl'
      Group '/IncohCodeFl/Data'
         Dataset 'Ambiguity'
         Dataset 'Beamcodes'
         Dataset 'PulsesIntegrated'
         Dataset 'Pulsewidth'
         Dataset 'TxBaud'
         Group
'/IncohCodeFl/Data/Acf'
            Dataset 'Data'
            Dataset 'Lagind'
            Dataset 'Lagmat'
            Dataset 'Lags'
            Dataset 'Range'
         Group
'/IncohCodeFl/Data/Power'
            Dataset 'Data'
            Dataset 'Range'
      Group '/IncohCodeFl/Mode'
         Dataset 'Dll'
         Dataset 'Name'
         Dataset 'Type'
         Dataset 'Version'
         Dataset 'VersionDate'
   Group '/Integration'
      Dataset 'MissedPulses'
   Group '/Rx'
      Dataset 'AeuRx'
      Dataset 'Bandwidth'
      Dataset 'CalTemp'
      Dataset 'FilterDelay'
      Dataset 'FilterRangeDelay'
      Dataset 'FilterTaps'
      Dataset 'Frequency'
      Dataset 'SampleRate'
      Dataset 'SampleSpacing'
      Dataset 'SampleTime'
      Dataset 'TuningFrequency'
   Group '/S'
      Group '/S/Cal'
         Dataset 'Beamcodes'
         Dataset 'PulsesIntegrated'
         Dataset 'Pulsewidth'
         Dataset 'TxBaud'
         Group '/S/Cal/Power'
            Dataset 'Data'
            Dataset 'Range'
      Group '/S/Mode'
         Dataset 'Dll'
         Dataset 'Name'
         Dataset 'Type'
         Dataset 'Version'
         Dataset 'VersionDate'
      Group '/S/Noise'
         Dataset 'Beamcodes'
         Dataset 'PulsesIntegrated'
         Dataset 'Pulsewidth'
         Dataset 'TxBaud'
         Group '/S/Noise/Power'
            Dataset 'Data'
            Dataset 'Range'
      Group '/S/ZeroLags'
         Dataset 'Ambiguity'
         Dataset 'Beamcodes'
         Dataset 'PulsesIntegrated'
         Dataset 'Pulsewidth'
         Dataset 'TxBaud'
         Group '/S/ZeroLags/Power'
            Dataset 'Data'
            Dataset 'Range'
   Group '/Setup'
      Dataset 'BeamcodeMap'
      Dataset 'Beamcodefile'
      Dataset 'Experimentfile'
      Dataset 'Program'
      Dataset 'RadacInfo'
      Dataset 'RxConfigfile'
      Dataset 'RxFilterfile'
      Dataset 'Systemfile'
      Dataset 'Tufile'
      Dataset 'TxConfigfile'
      Group '/Setup/Ambiguity'
         Group
'/Setup/Ambiguity/A16_30_10'
            Dataset 'Delay'
            Dataset 'Lags'
            Dataset 'Range'
            Dataset 'Wlag'
            Dataset 'WlagSum'
            Dataset 'Wrange'
            Dataset

'WrangeSum'
         Group
'/Setup/Ambiguity/Lp_30_10'
            Dataset 'Delay'
            Dataset 'Lags'
            Dataset 'Range'
            Dataset 'Wlag'
            Dataset 'WlagSum'
            Dataset 'Wrange'
            Dataset 'WrangeSum'
   Group '/Site'
      Dataset 'AeuTotal'
      Dataset 'Altitude'
      Dataset 'Code'
      Dataset 'Latitude'
      Dataset 'Longitude'
      Dataset 'Name'
   Group '/Time'
      Dataset 'MatlabTime'
      Dataset 'RadacTime'
      Dataset 'RadacTimeString'
      Dataset 'Synchronized'
      Dataset 'UnixTime'
   Group '/Tu'
      Dataset 'FrameTime'
      Dataset 'TrDuty'
   Group '/Tx'
      Dataset 'AeuTx'
      Dataset 'Frequency'
      Dataset 'Power'
      Dataset 'TuningFrequency'
```