HamSci Data Plane

HamSci Eye in the Sky

Michelle Thompson

Open Research Institute

For HamSci

January 14, 2019

Abstract

HamSci, or Ham Radio Science Citizen Investigation, advances scientific research and understanding through amateur radio activities. Primary cultural benefits include the development of new technologies along with providing excellent educational opportunities for both the amateur community and the general public.

At TAPR DCC 2018 (Tucson Amateur Packet Radio's Digital Communications Conference), a distributed receiver system for space weather research was proposed. The HamSci Space Weather Project home page with complete description and extensive project resources is located here: http://hamsci.org/basic-project/personal-space-weather-station

Space weather station receivers are envisioned as having one or more of a wide variety of sensors. Sensors range from ground magnetometers, to ionospheric sounders, to lightning detectors and more. The opportunity for a diversity of radios participation is immediately obvious and very exciting. The addition of a open source cubesat to coordinate observations from above is proposed as part of the network. Observing from ground and space simultaneously provides substantial additional scientific value. A receiver network that can be coordinated to make scheduled observations that align with satellite passes is possible and can be modeled and enabled with SatNOGS open source software. See https://satnogs.org/ for more information about this open source satellite network on the ground.

From slides 58 - 71 of the Sunday Seminar slide deck at http://hamsci.org/sites/default/files/pages/swstation/20181116_TAPR_Sunday_Seminar_Frissell_W2NAF.pdf, it appears that a dedicated lightning detector on a Raspberry Pi can participate alongside a USRP X310 sampling at highest rate and bandwidth. This type of distributed network opens up enormous possibilities.

We stipulate and expect that the radio hardware currently available and the radios under discussion for development by TAPR will be of sufficient capability to support

useful data collection. The central challenge of the HamSci space weather program is not the radio hardware. The central challenge is how the radios are interconnected and how the large quantity of data is handled. The central challenge is how to produce, transport, process, share, and publish the data, datasets, information, and knowledge. That challenge is what this paper is about.

Introduction

HamSci Space Weather Stations produce radio receiver data. In some cases, the data stream may be small in size or volume. Data on lightning strikes or for particular, narrowly-defined, or infrequent atmospheric observations can be exfiltrated over normal retail internet channels without requiring any special equipment or services. In other cases, the data stream may be very large in size and volume. For example, taking measurements up to 60MHz wide with any degree of precision in resolution and time will require substantial resources in transport, storage, and processing. Slide 77-82 shows how large the data stream can potentially be and proposes a ring buffer to manage the observations. This means that there's a window of opportunity to identify a set of desired observations before they circulate out of the ring buffer and are lost. Rapidly identifying interesting "unknown unknowns" is a difficult requirement.

The size of the data that we are dealing with means we have a Big Data situation. Big Data groups together processing, collection, storage and visualization of large quantities of data. Data Science is the process of extracting knowledge from data. Data is collected, information is derived, and knowledge is gained. Knowledge comes from analysis and synthesis of information. Information comes from classifying, organizing, and interpreting the data.

Data is produced and recorded as measurements or samples that can be either unstructured or structured. Unstructured data is generally considered to be the sort of data we get from humans. For example, current mood, a year's worth of daily diary entries, opinions on a product from a survey, posts to a forum, height and weight linked to gym check-ins, and so on. Structured data is generally considered to be what we get from machines or circuits and software. In our case, we expect to be dealing with some variety of structured digital samples and information from the radio spectrum. The amount of spectrum, the resolution, the time accuracy, and the signal to noise ratio are the essential characteristics of these radio measurements. The characteristics will vary for each sensor type. The characteristics will vary within a population. There are radio techniques that can produce a body of high quality radio data from a population of

radios that do not have enough individual performance to produce quality data individually. Interferometry, phasing, averaging and other digital signal processing techniques allow for low cost simpler gear to be used to create data that is higher quality in aggregate. This could dramatically reduce the cost of the receiver network while dramatically increasing participation.

The reduction of cost and increase in participation is purchased through effort and care in the complexity contained within the elements first encountered on slide 69. These include the "internet cloud" and the HamSci Public Database.

In order to get the full potential out of data, it must be available to anyone that has an interest in using it. The data must be available without unreasonable obstacles and at affordable prices. Data must be provided along with with models and equations so that interested users do not have to reinvent well-understood processes of deriving information from the data. If a user is interested in deriving new information from the data, then the new models and equations can be published alongside the existing ones. Over time, this increases the value of what we call a dataset. Data by itself isn't a dataset. A dataset is a documented searchable set of data that includes models, tools, and equations. A database is an excellent template for a dataset.

From slide 85, a list of RF Instrument Metadata is given. Metadata is information added to samples to record sample characteristics, increase findability, and define station identity, configuration, and location.

Visualization is the art of representing data in a visual manner. Graphs, diagrams, maps, animations and videos rapidly communicate essential knowledge from the dataset. The users of the datasets, visualizations, and knowledge are at the opposite end of the ecosystem from the station operators, who are producing and shipping data. In the middle is the processing, transporting, storage, classifying, informatics, analysis, and synthesis. This middle part is the HamSci Data Plane.

Purpose or Hypothesis

The purpose of the HamSci Data Plane is to process the data produced from HamSci Space Weather Receivers into datasets.

Research Questions

1) Is there an existing open source RF metadata protocol that fits our needs?

Proposed options include but are not limited to:

Haystack
https://github.com/MITHaystack/digital_rf

GNU Radio SigMF
https://github.com/gnuradio/SigMF

2) Exactly what information needs to be derived from the data in order to guarentee a minimum standard of quality for the science mission?

Receiver specifications need to be derived from the form of the scientific observations. What is the minimum viable product that will record accurate data, per observation type?

3) Exactly what information needs to be derived from the data in order to support heterogeneous radio receiver participation in aggregate?

If digital signal processing techniques are used to create quality aggregate data from less-capable receivers, what characteristics are required as the inputs to the objective functions? What objective functions are defined in the Data Plane?

Importance of the Project

The HamSci Space Weather Project is important because it enables basic research into atmospheric science.

Method of Approach

The research methodology is the deployment, operation, and maintenance of a distributed receiver system, a Data Plane that collects, processes, and publishes data, and an open data system that serves and maintains a HamSci Public Database.

Work Plan

Budget

This should be as detailed as possible but simple to understand. A simple table with line items should suffice. Any descriptions or justifications can be done below the table.

Line Item	Amount
Total	

References

S. Sedkaoui & JL Monino. (2016). *Big Data, Open Data and Data Development.* Wiley-ISTE.

https://github.com/MITHaystack/digital_rf

https://github.com/gnuradio/SigMF

https://satnogs.org/