HamSci Data Plane

HamSci Eye in the Sky

Michelle Thompson

Open Research Institute

For HamSci

January 15, 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.

A distributed receiver system for space weather research was proposed at TAPR DCC 2018.[1] 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

HamSci Space Weather Stations produce receiver data from transmitters associated with coordinated observations. Sensors range from ground magnetometers, to ionospheric sounders, to lightning detectors and more. The diversity of sensor types means a wide variety of radios can participate.

Slides 59 - 68 from Dr. Nathaniel Frissell's TAPR DCC Sunday Seminar slide deck[2] are reproduced below.

Personal Terrestrial WX Station

- Multi-instrument
- Internet Connected
- Easy Set-Up
- Reasonable Cost

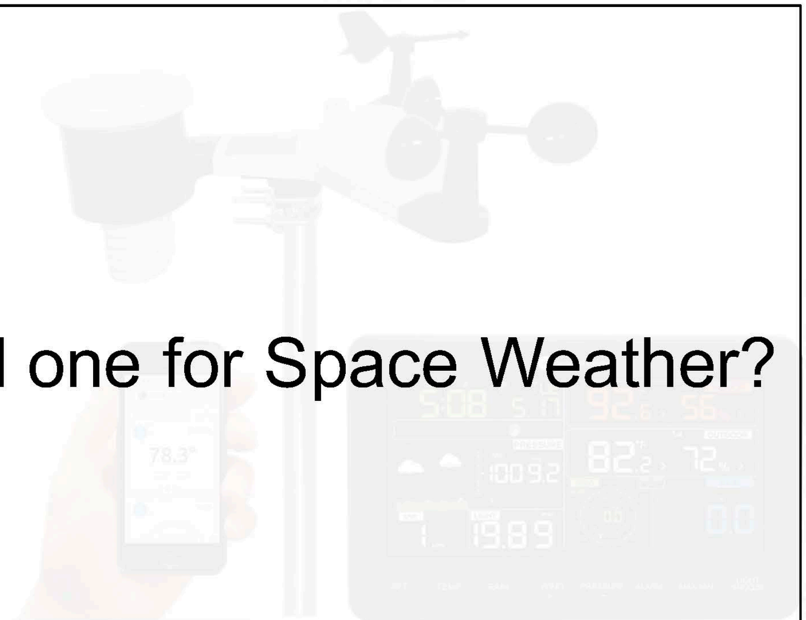
Ambient Weather WS-2902

HamSCI http://hamsci.org NJIT frissell@njit.edu

Personal Terrestrial WX Station

- Multi-instrument
- Internet Connected
- Easy Set-Up
- Reasonable Cost

Can we build one for Space Weather?



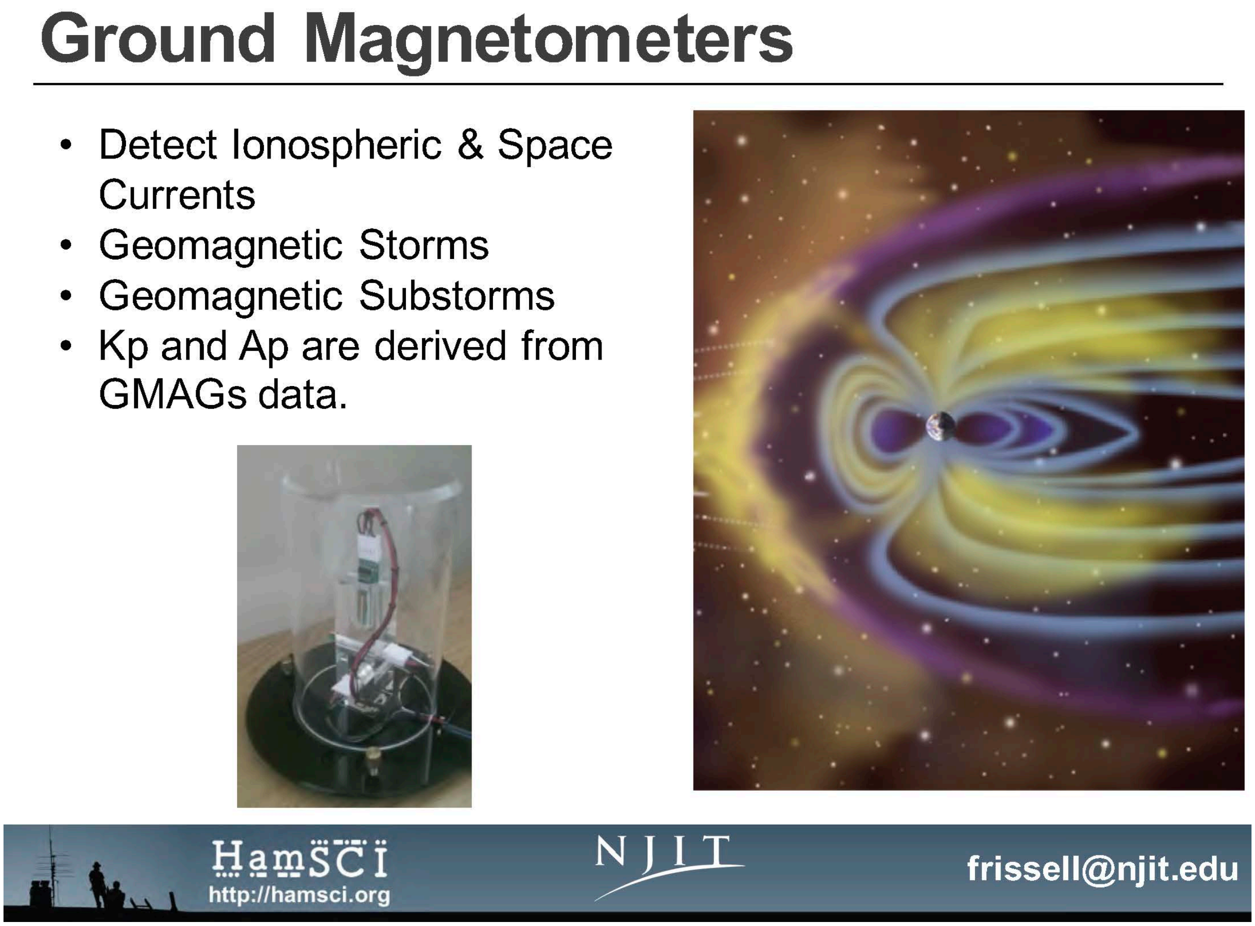
Ambient Weather WS-2902

HamSCI http://hamsci.org NJIT frissell@njit.edu

Instrument Possibilities

- Ground Magnetometer?
- GPS-TEC Receiver?
- Ionosonde?
- Riometer?
- WWV/Standards Station Monitor?
- RBN/PSKReporter/WSPR Receiver?
- Lightning Detector?
- Others?

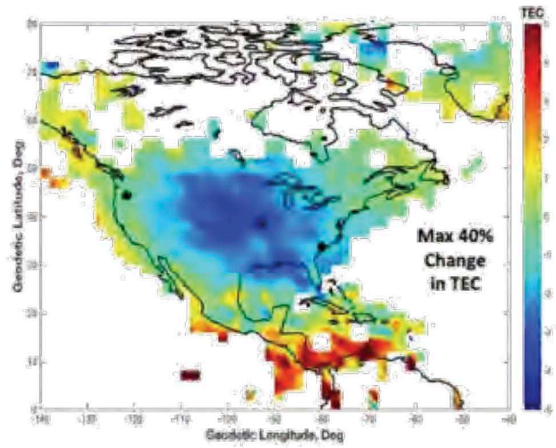
What makes sense for a personal, ground-based local station?


Ground Magnetometers
• Detect Ionospheric & Space Currents
• Geomagnetic Storms
• Geomagnetic Substorms
• Kp and Ap are derived from GMAGs data.
HamSCI
http://hamsci.org
NJIT
frissell@njit.edu

GPS Total Electron Content

- Total Number of electrons between ground and GPS Satellite
- Measured by examining delay between two GPS Frequencies
- Traveling Ionospheric Disturbances
- Storm Effects
- Ionospheric Scintillations

Solar Eclipse
GNSS Vertical Total Electron Content
21 August 2017
Difference in TEC at 18:15 UT from start of solar eclipse at 16:45 UT



Support: NSF AGS-1242204, NASA NNX17AH71G

Courtesy of Anthea Coster

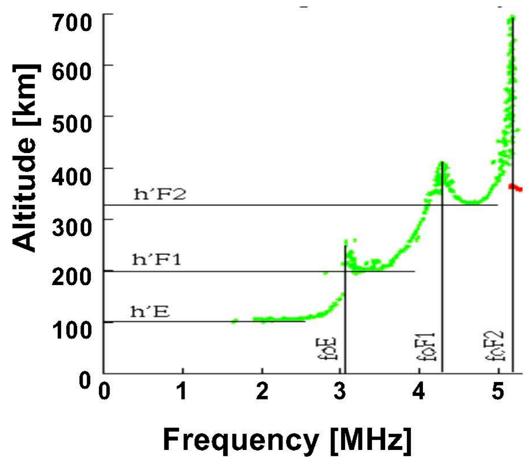
HamSCI
http://hamsci.org
NJIT
frissell@njit.edu

Ionosondes

- Vertical Incidence HF Radar
- Measure Plasma Density for bottomside Ionosphere

$$f_{pe} \approx 9\sqrt{n_e}$$

[Dr. Terry Bullett, W0ASP, U of Colorado]



HamSCI http://hamsci.org NJIT frissell@njit.edu

Riometer

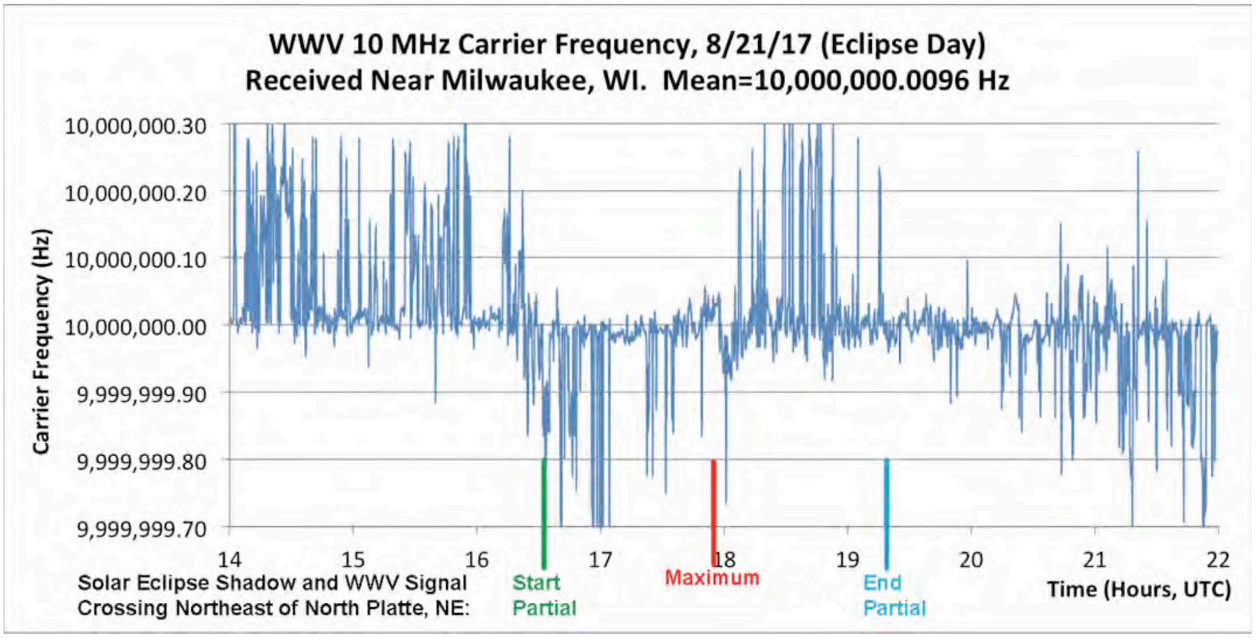
- **R**elative **I**onopheric **O**pacity **M**eter
- Directly measures absorption of cosmic rays
- Indirectly measures electron density, particle precipitation
- Typically passive instrument 30-50 MHz

IRIS - Imaging Riometer for Ionospheric Studies in Finland (http://kaira.sgo.fi/)

Photo: Derek McKay

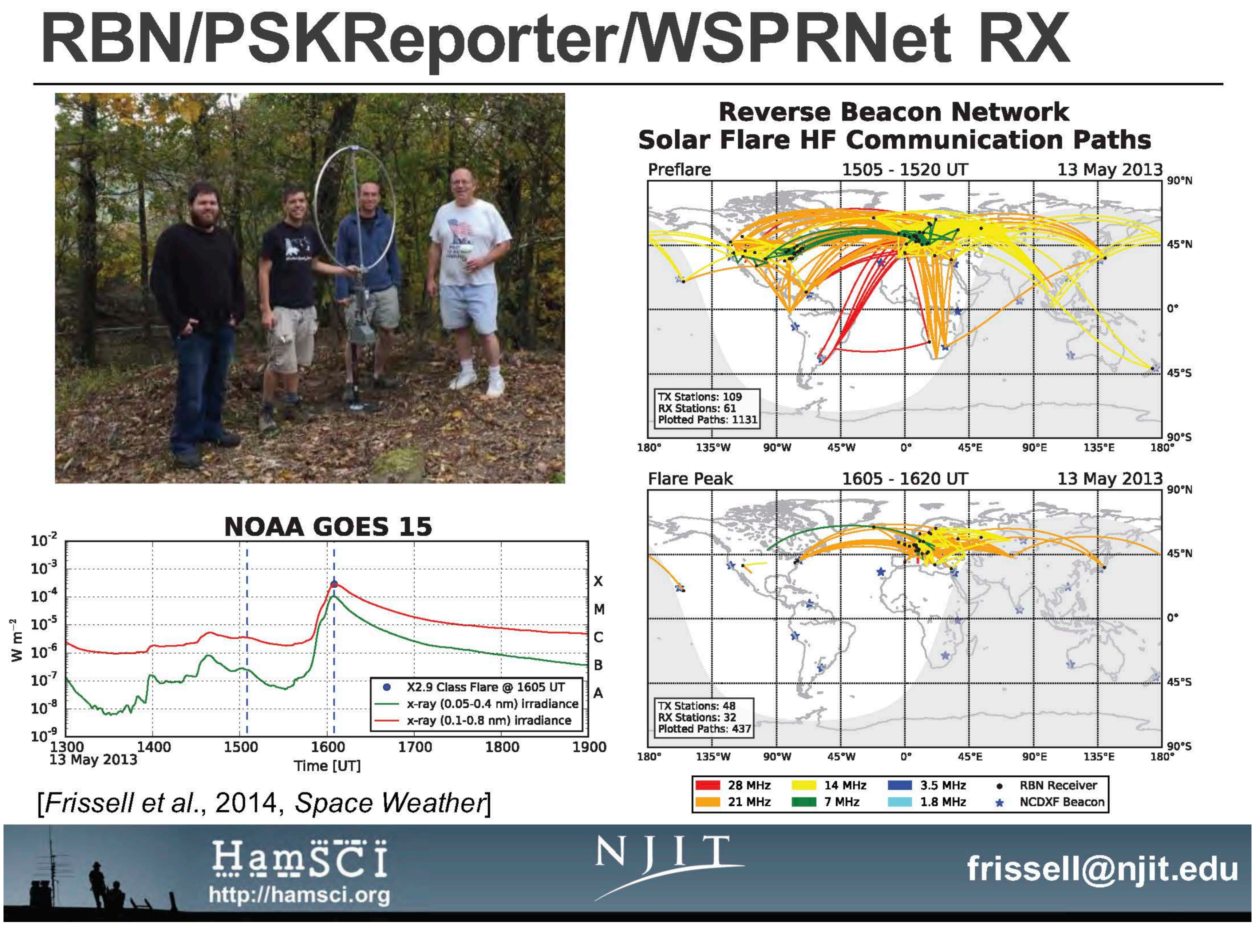
HamSCI http://hamsci.org NJIT frissell@njit.edu

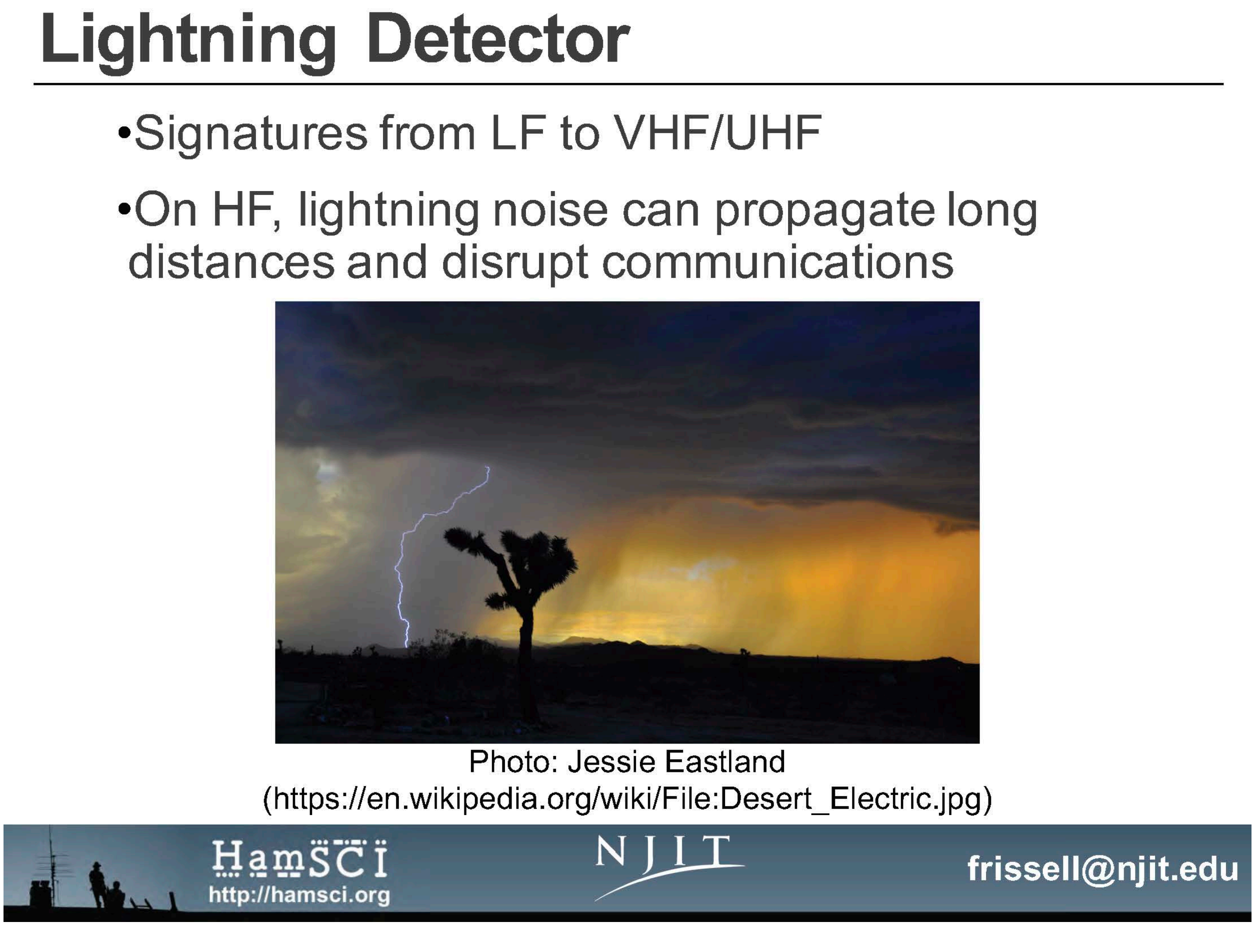
WWV/CHU Standards Monitor



Steve Reyer, WA9VNJ

HamSCI http://hamsci.org NJIT frissell@njit.edu


RBN/PSKReporter/WSPRNet RX
Reverse Beacon Network
Solar Flare HF Communication Paths
Preflare 1505 - 1520 UT 13 May 2013
TX Stations: 109
RX Stations: 61
Plotted Paths: 1131
Flare Peak 1605 - 1620 UT 13 May 2013
TX Stations: 48
RX Stations: 32
Plotted Paths: 437
90°N 45°N 0° 45°S 90°S
180° 135°W 90°W 45°W 0° 45°E 90°E 135°E 180°
28 MHz 14 MHz 3.5 MHz RBN Receiver
21 MHz 7 MHz 1.8 MHz NCDXF Beacon
NOAA GOES 15
X2.9 Class Flare @ 1605 UT
x-ray (0.05-0.4 nm) irradiance
x-ray (0.1-0.8 nm) irradiance
W m^{-2}
Time [UT]
13 May 2013
[Frissell et al., 2014, Space Weather]
HamSCI
http://hamsci.org
NJIT
frissell@njit.edu



Combining sensor data from disparate sources, when the end result has greater certainty, accuracy, or quality than if the data was used individually, is called sensor fusion. The HamSci Space Weather System, as proposed above, is an example of sensor fusion. For example, a dedicated lightning detector on a Raspberry Pi in Florida, USA can participate in this network with a USRP X310 sampling at highest rate and bandwidth in Madrid, Spain. The data from the lightning detector may enhance the data from the radiometer and increase scientific knowledge. Another example is a set of five radios configured as ionosondes. Each have different capabilities or characteristics. The data combined is better than any one station's individual contribution.

An 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. This 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.

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 Systems 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. Slides[3] 77-82 proposes a ring buffer to manage the observations. These slides are reproduced below. Using a ring buffer 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.

Where do we start?

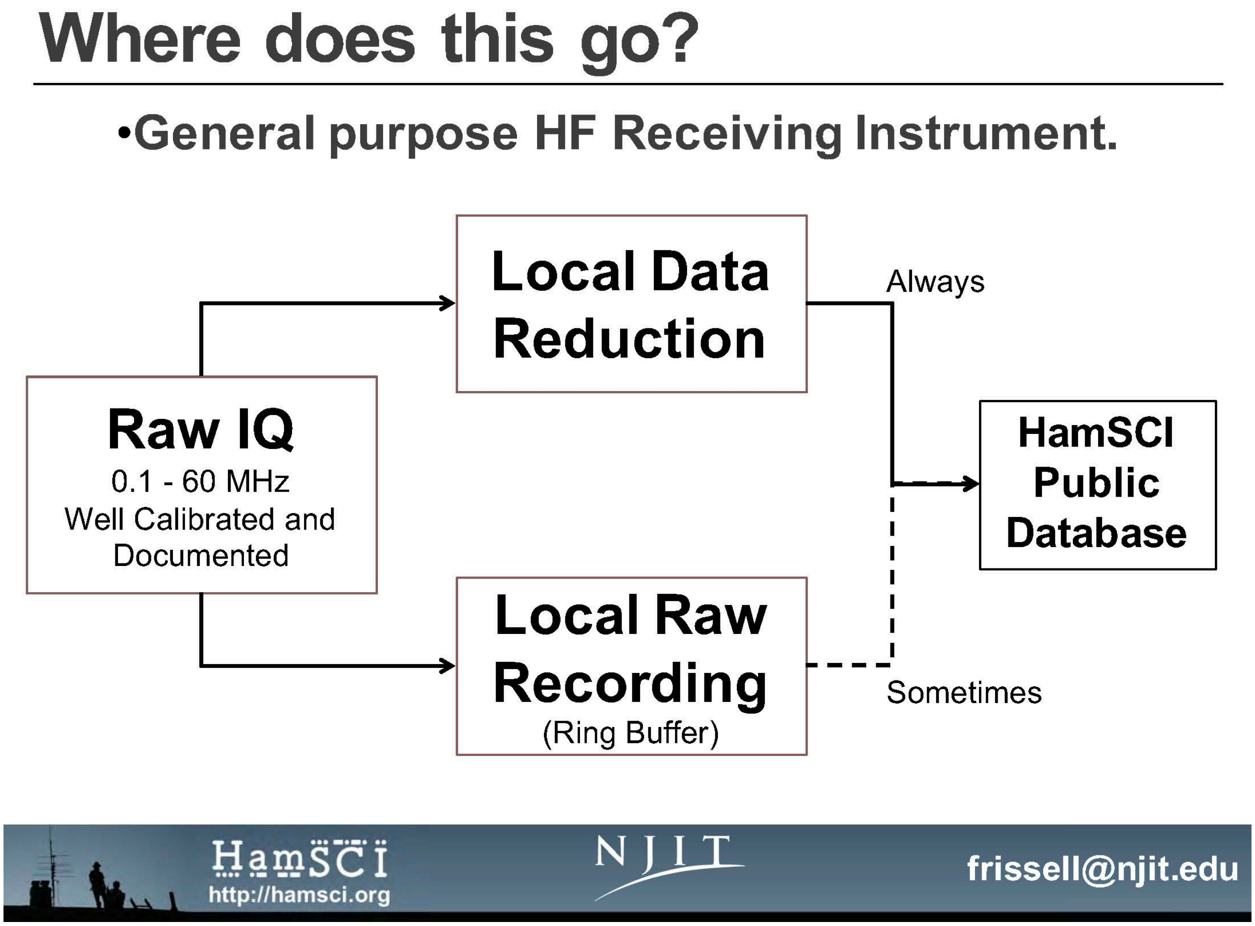
- **General purpose HF Receiving Instrument.**
- **Why?**
  - Few networks of widespread scientific HF radio receivers currently exist.
  - “Signals of opportunity” available.
  - Extremely flexible research tool.
  - Directly applicable to ham radio.
  - Radio is TAPR’s Bread and Butter ☺

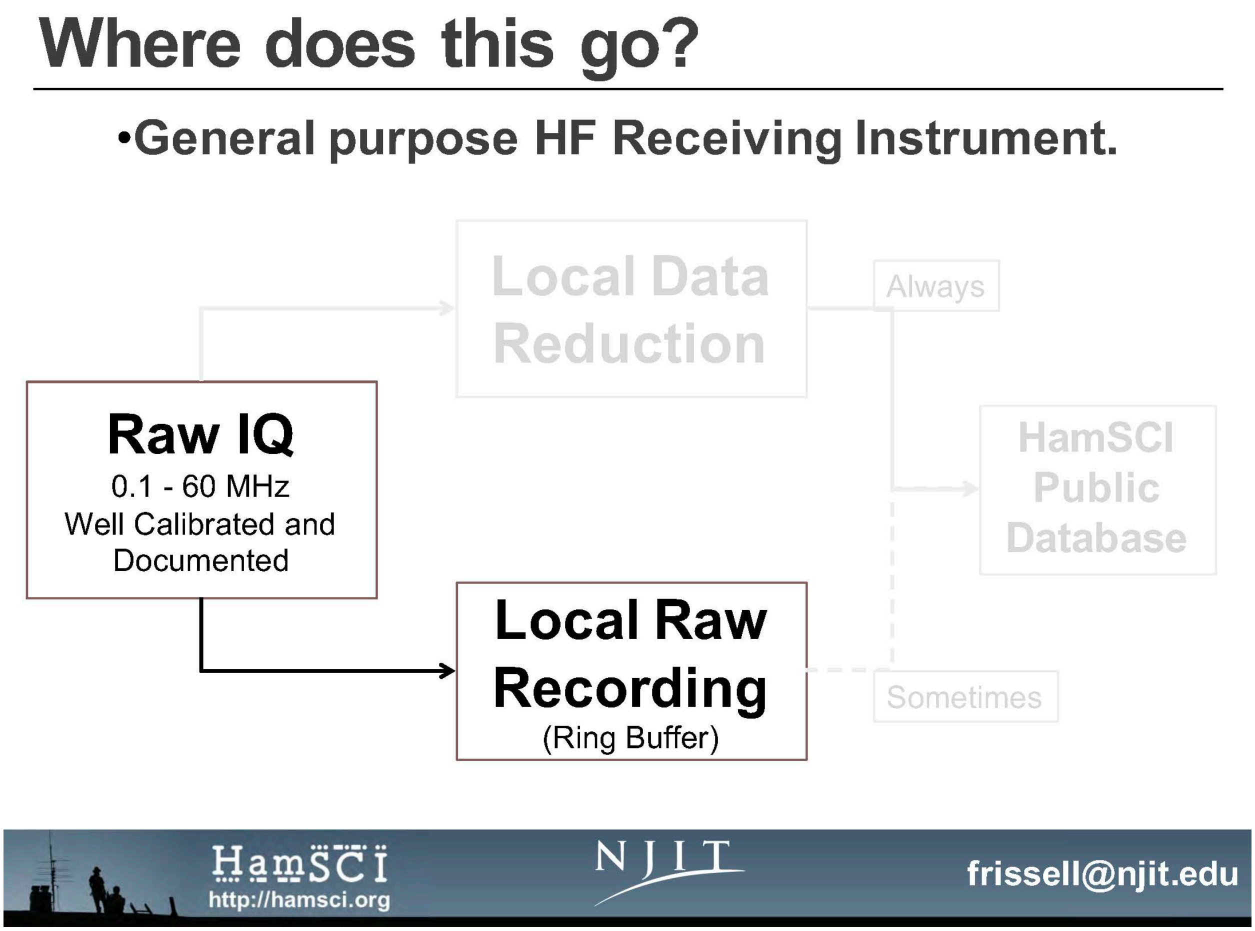
HamSCI
http://hamsci.org
NJIT
frissell@njit.edu

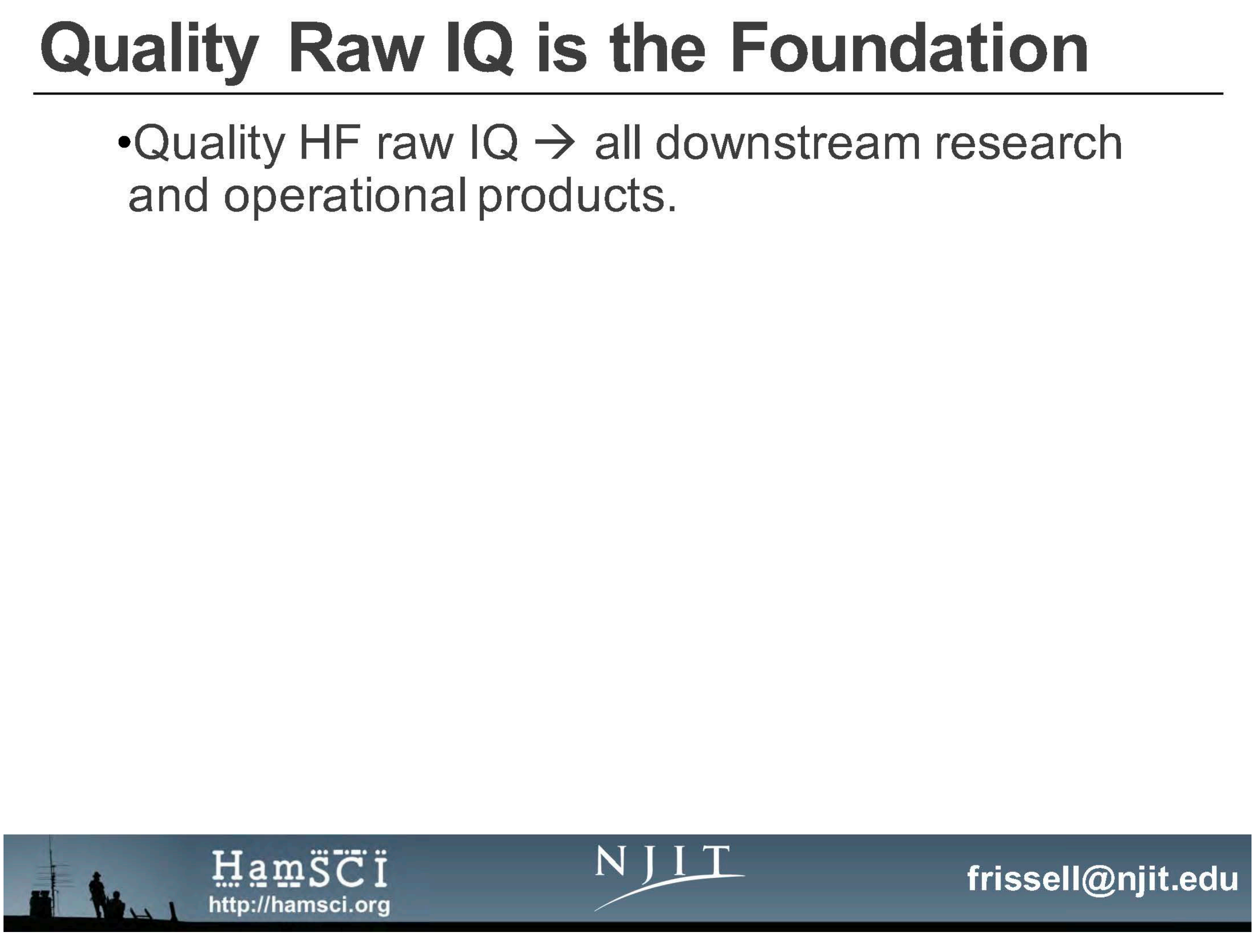
Where do we start?

- General purpose HF Receiving Instrument.

Raw IQ
0.1 - 60 MHz
Well Calibrated and Documented


Where does this go?
•General purpose HF Receiving Instrument.
Local Data Reduction
Always
Raw IQ
0.1 - 60 MHz
Well Calibrated and Documented
HamSCI Public Database
Local Raw Recording
(Ring Buffer)
Sometimes
HamSCI
http://hamsci.org
NJIT
frissell@njit.edu


Where does this go?
•General purpose HF Receiving Instrument.
Local Data Reduction
Always
Raw IQ
0.1 - 60 MHz
Well Calibrated and Documented
HamSCI Public Database
Local Raw Recording
(Ring Buffer)
Sometimes
HamSCI
http://hamsci.org
NJIT
frissell@njit.edu



Big Data

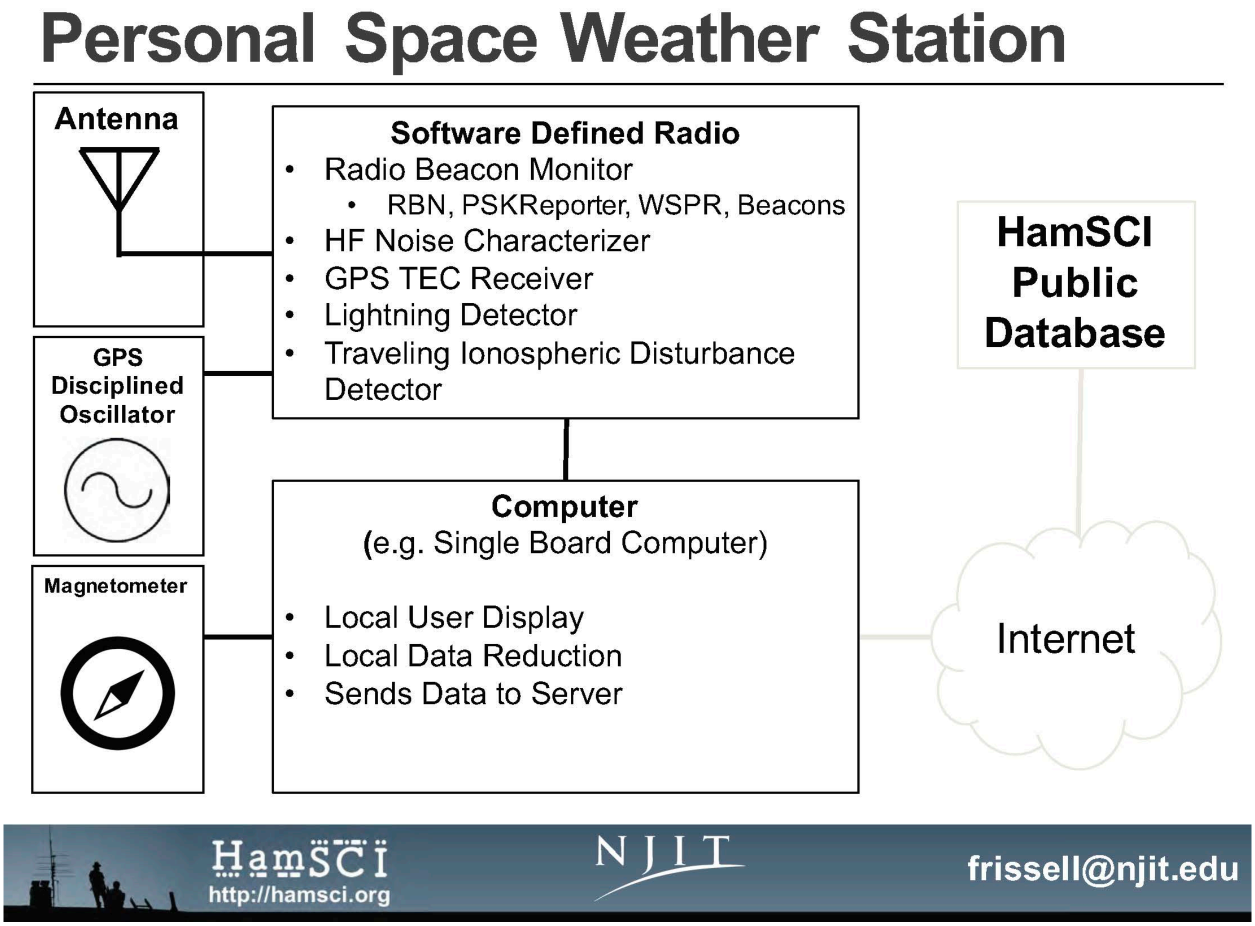
The size of the data that we are dealing with means we have a Big Data situation. Big Data groups together processing, collection, storage and visualization of large quantities of data. Data Science is the process of extracting knowledge from data. Data is collected, information is derived, and knowledge 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 machine readable and is also expected to be human readable. Structured data can be produced by humans. Structured data can be produced by hardware, circuits, or software.

HamSci expects 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. The data is machine and human readable.

Successful data fusion reduces the cost of the receiver network while increasing participation.

Slide[4] 69 is a very high-level view of the Data Plane. It is reproduced below.



The "Internet" and "HamSci Public Database" are the parts of the drawing that correspond to the Data Plane.

Internet

While the internet is the most obvious way to network a distributed receiver system together, very remote stations or stations that produce very large amounts of data could ship their data on physical media. There is a point where the cost to exfiltrate the data exceeds the cost of saving it to a local hard drive and periodically shipping that hard drive to the location where the data is processed.

HamSci Public Database

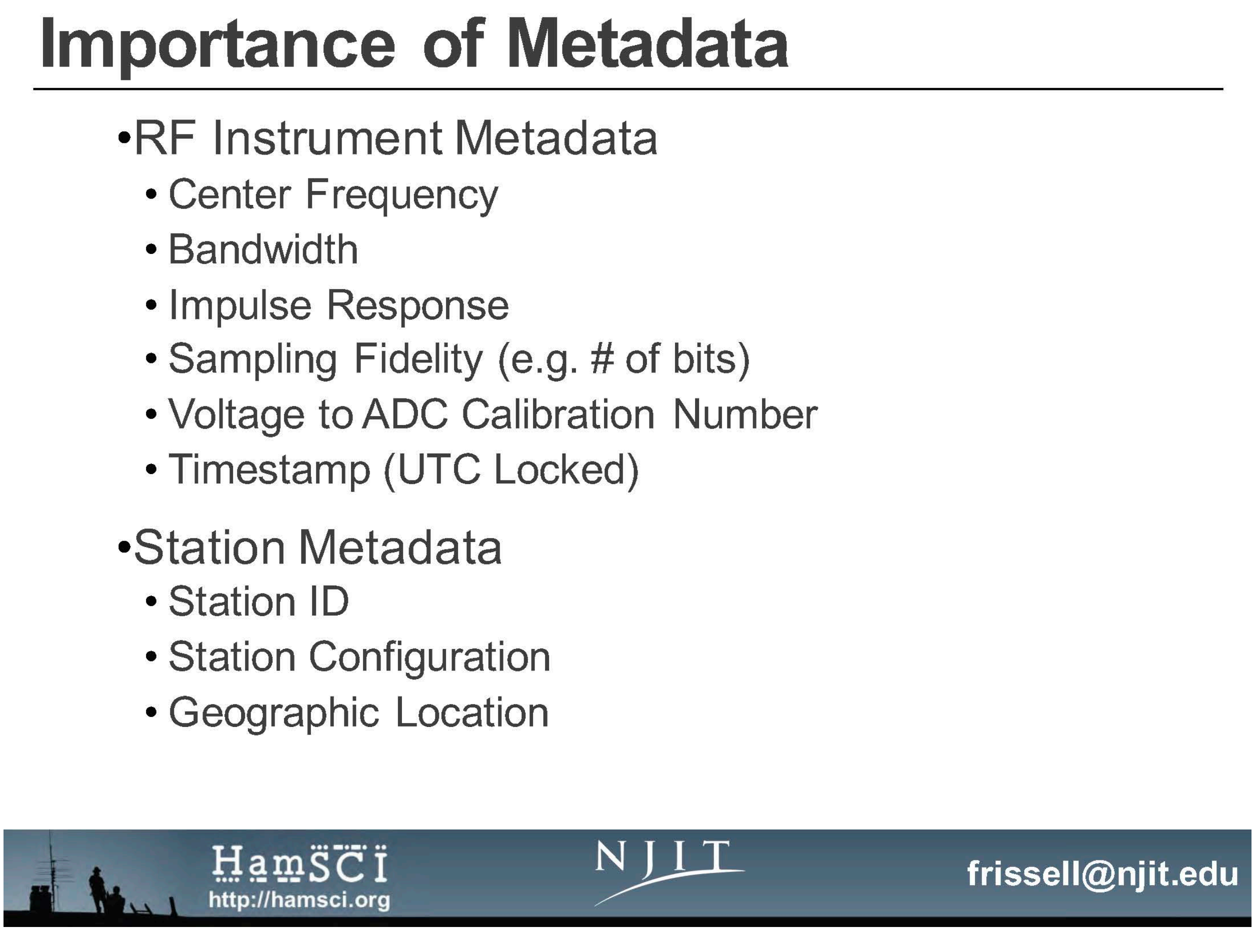
In order to get the full potential out of data, it must be available to anyone that has an interest in using it. The data must be available without unreasonable obstacles and at affordable prices. Data must be provided along with with models and equations so that interested users do not have to reinvent 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.

What isn't yet represented?

The internet and the public database are required elements. The models, equations, and computing resources are not yet represented.

Metadata

From slide[5] 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. The slide is reproduced below.



Metadata is required for successful sensor fusion. The identification and integration of an open source metadata protocol is one of the first steps in the Work Plan. Two candidate protocols are Haystack and SigMF. The selection of any particular metadata protocol has significant performance repercussions.

Visualization

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.

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 sensor types. What is the minimum viable product that will record accurate data, per sensor 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 need to be 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 development, deployment, operation, and maintenance of a distributed receiver system. The receiver system produces data that is collected and processed by a Data Plane. The output of the Data Plane is published in a HamSci Public Database.

Work Plan

This is arguably the most important part of the proposal as it shows how you will conduct your study. As such, it should be as realistic as possible and very detailed. For my proposal I broke it down by what I intended to accomplish each month.

January

8. Data analysis and write-ups
9. Begin writing thesis draft
10. Brown bag presentation to department

February - April

11. Continue writing thesis and revise when necessary
12. Submit draft to committee for review

May

13. Final draft complete by end of May with committee revisions

-=-=-=-=-=-=-=
June-July
-=-=-=-=-=-=-=

August

1. Determine survey questions

October

2. IRB request - 2 months
3. Meet with committee

November

4. Mail surveys
5. Data Entry
6. Submit abstract for professional conference paper or poster presentation

December

7. Data Entry

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
Storage	
Access	
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/

---

[1] Tucson Amateur Packet Radio's Digital Communications Conference was held 14-16 September 2018 in Albuquerque, New Mexico. https://tapr.org/dcc.html

[2] http://hamsci.org/sites/default/files/pages/swstation/20181116_TAPR_Sunday_Seminar_Frissell_W2NAF.pdf

[3] http://hamsci.org/sites/default/files/pages/swstation/20181116_TAPR_Sunday_Seminar_Frissell_W2NAF.pdf

[4] http://hamsci.org/sites/default/files/pages/swstation/20181116_TAPR_Sunday_Seminar_Frissell_W2NAF.pdf

[5] http://hamsci.org/sites/default/files/pages/swstation/20181116_TAPR_Sunday_Seminar_Frissell_W2NAF.pdf