HamSci Data Plane + Satellite

Michelle Thompson

Open Research Institute

For HamSci

January 16, 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

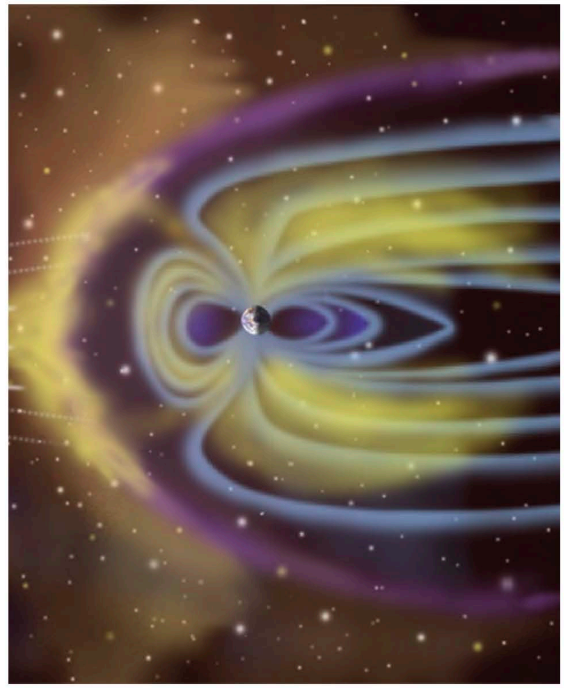
Instrument Possibilities

- 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.

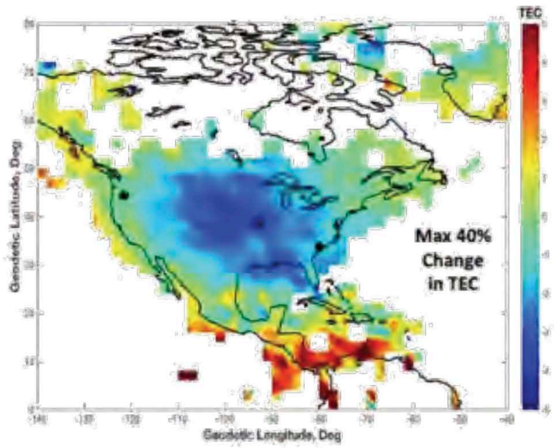


HamSCI http://hamsci.org NJIT frissell@njit.edu

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

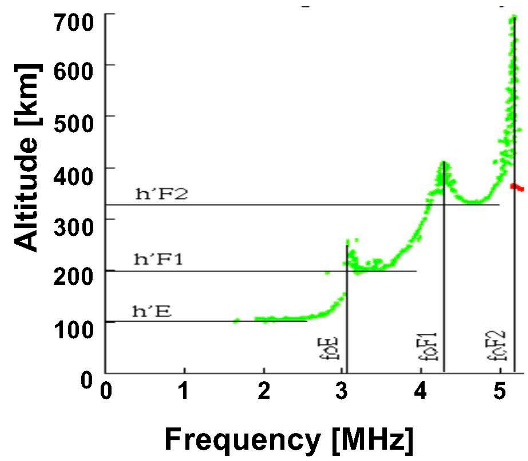
HamSCI http://hamsci.org NJIT frissell@njit.edu

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

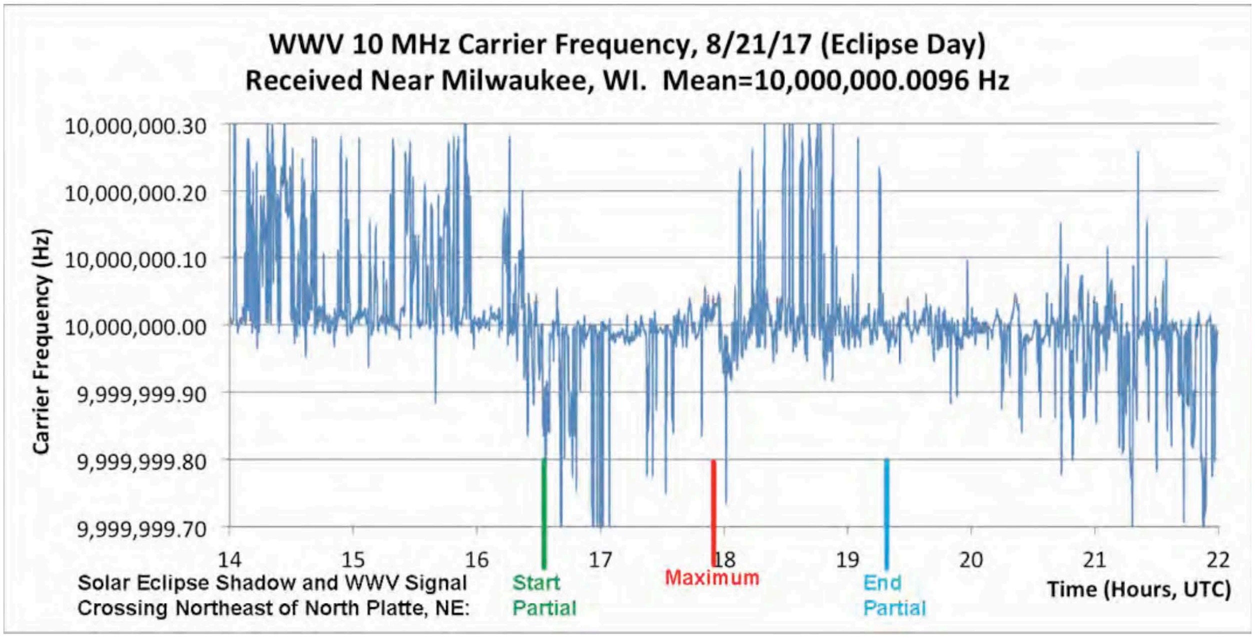
- **R**elative **I**onopheric **O**pacity **M**eter
- Directly measures absorption of cosmic rays
- Indirectly measures electron density, particle precipitation
- Typically passive instrument 30-50 MHz

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

Photo: Derek McKay

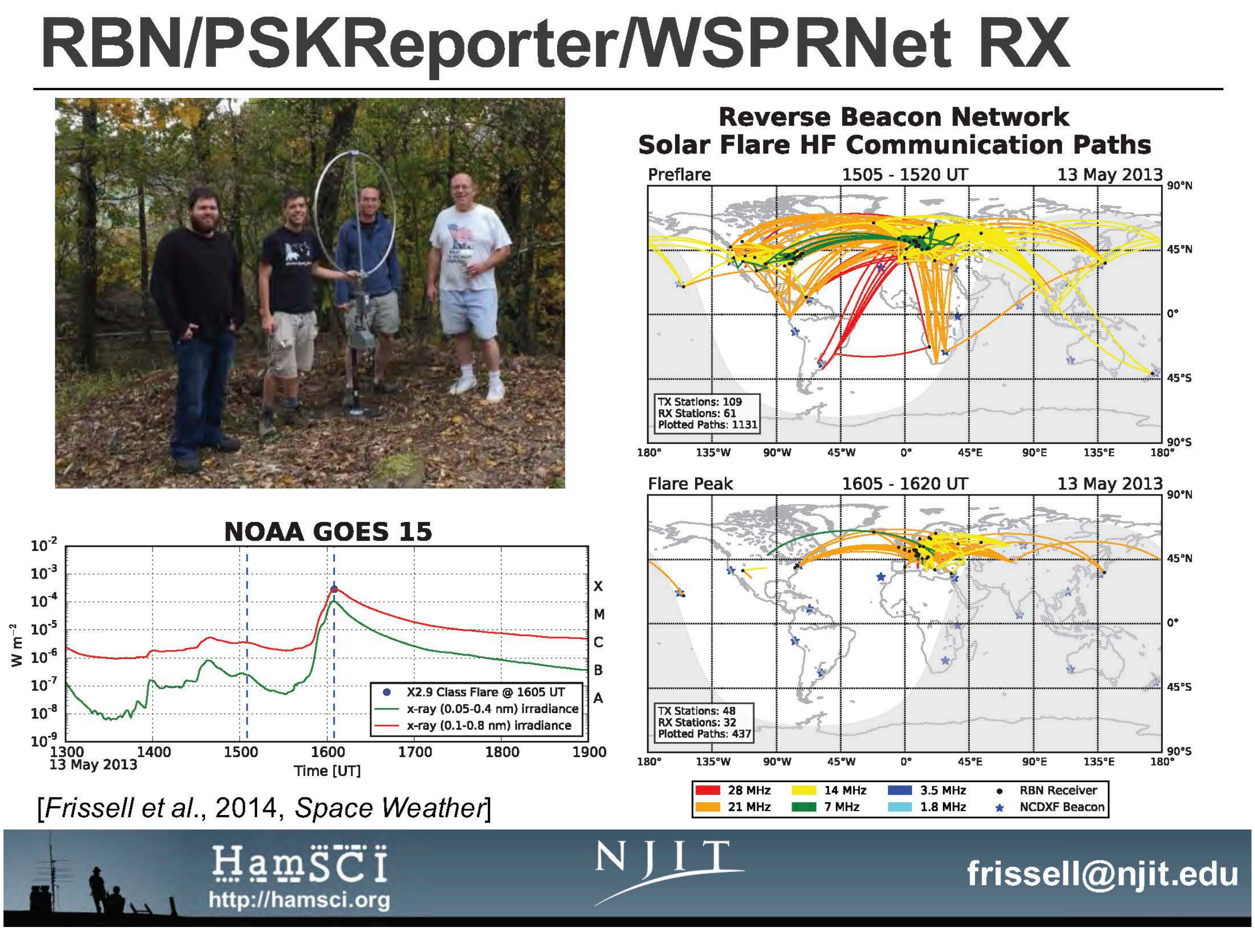
HamSCI http://hamsci.org NJIT frissell@njit.edu

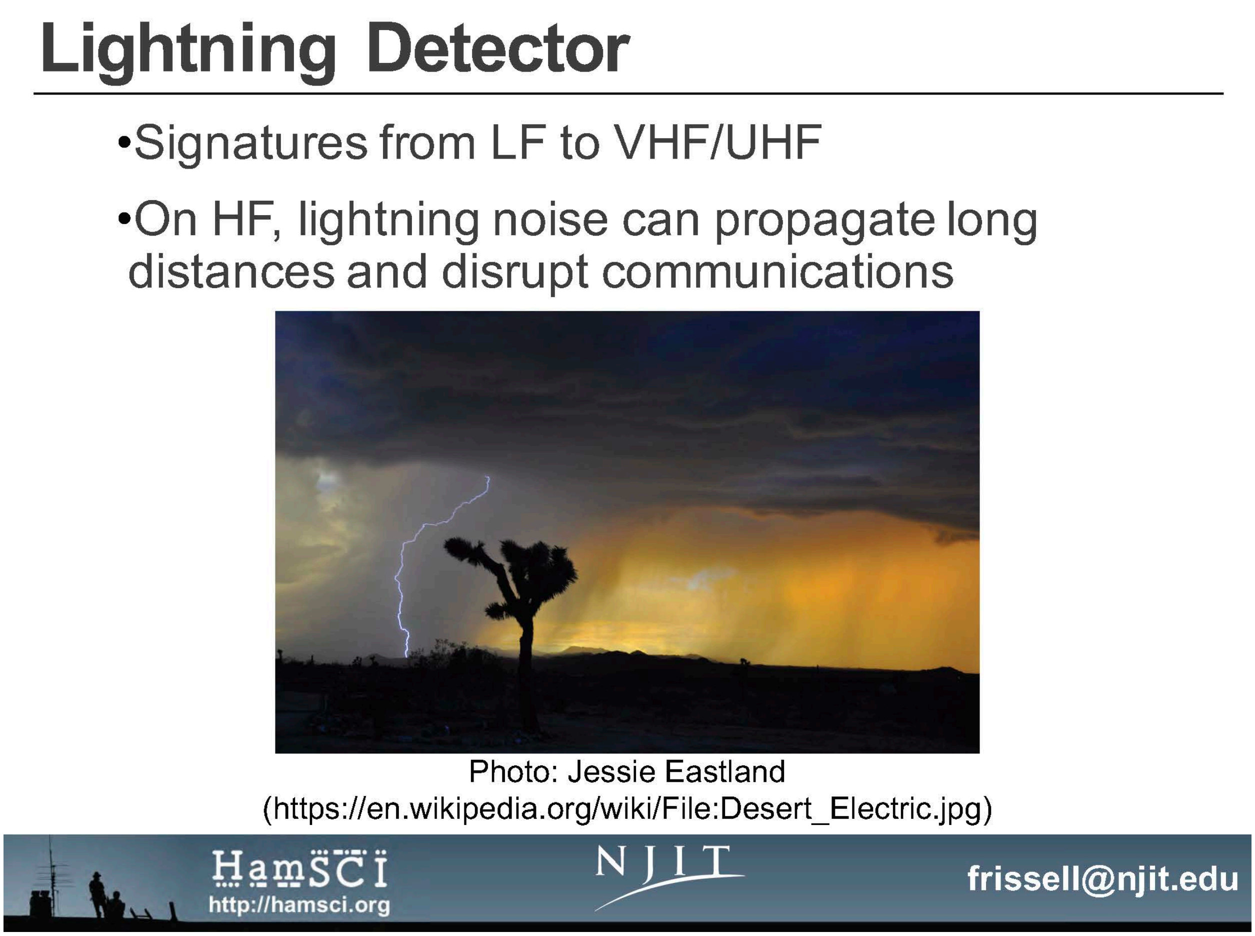
WWV/CHU Standards Monitor



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
28 MHz
21 MHz
14 MHz
7 MHz
3.5 MHz
1.8 MHz
RBN Receiver
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
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?

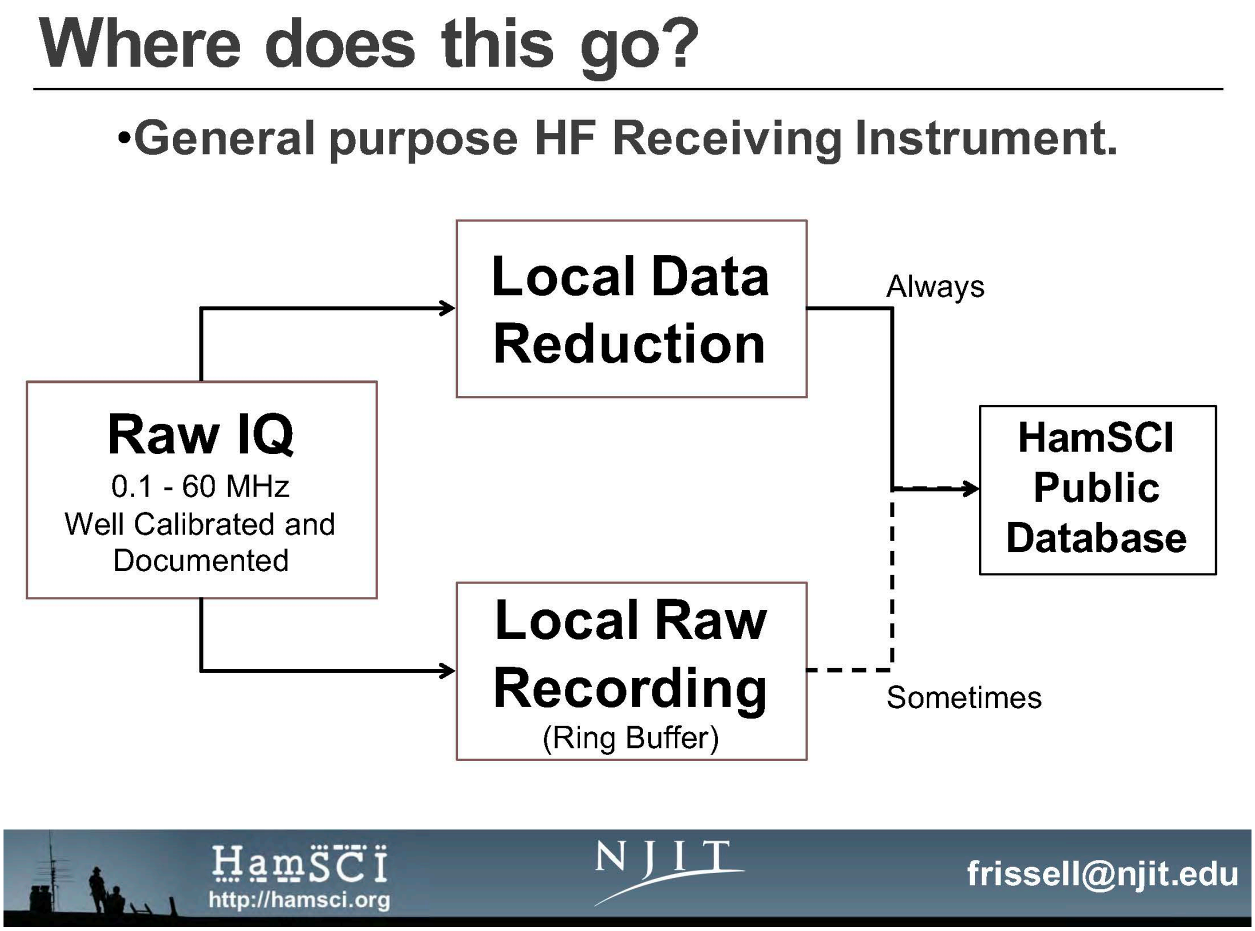
- **General purpose HF Receiving Instrument.**
- **Why?**
  - Few networks of widespread scientific HF radio receivers currently exist.
  - “Signals of opportunity” available.
  - Extremely flexible research tool.
  - Directly applicable to ham radio.
  - Radio is TAPR’s Bread and Butter ☺

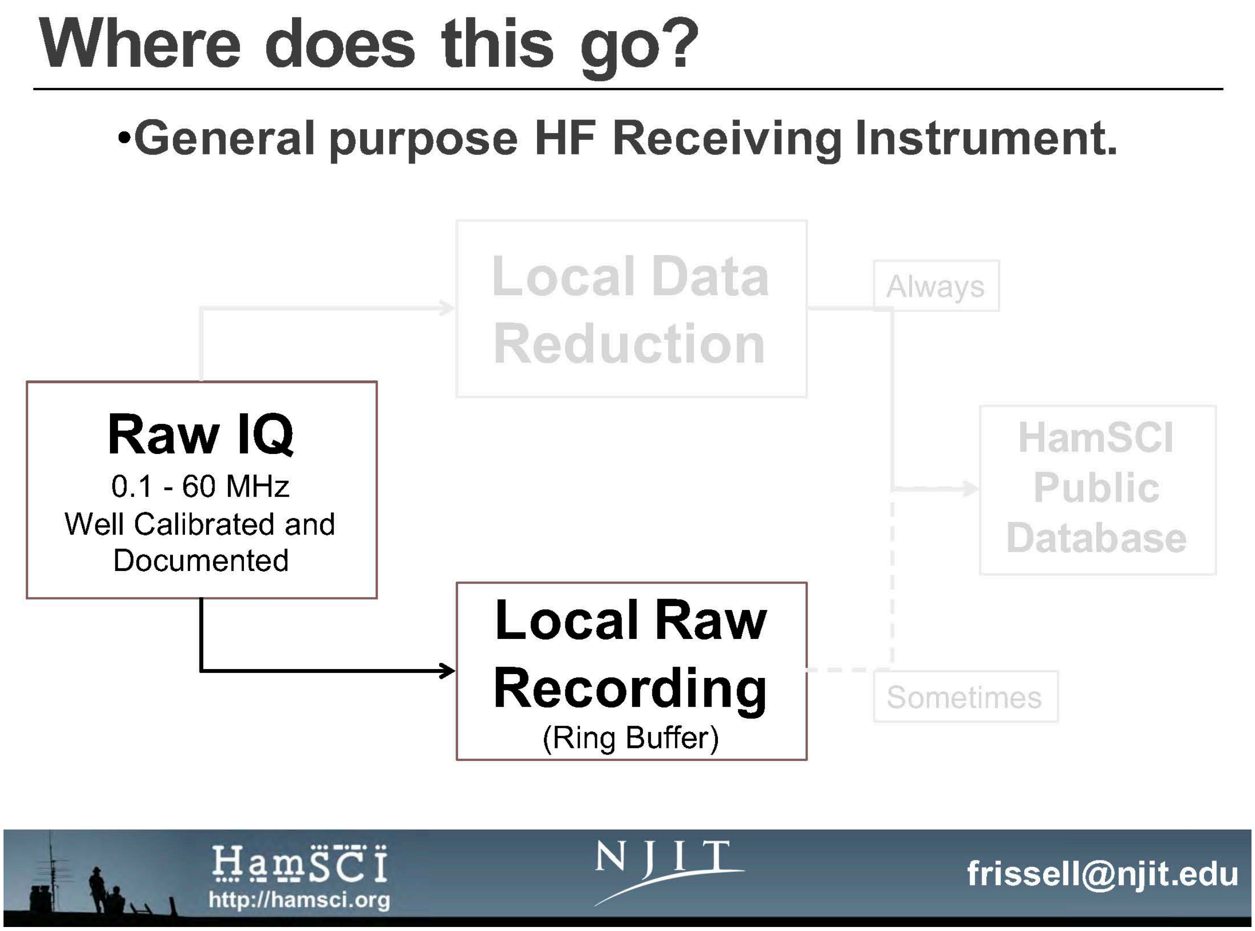
HamSCI http://hamsci.org NJIT frissell@njit.edu

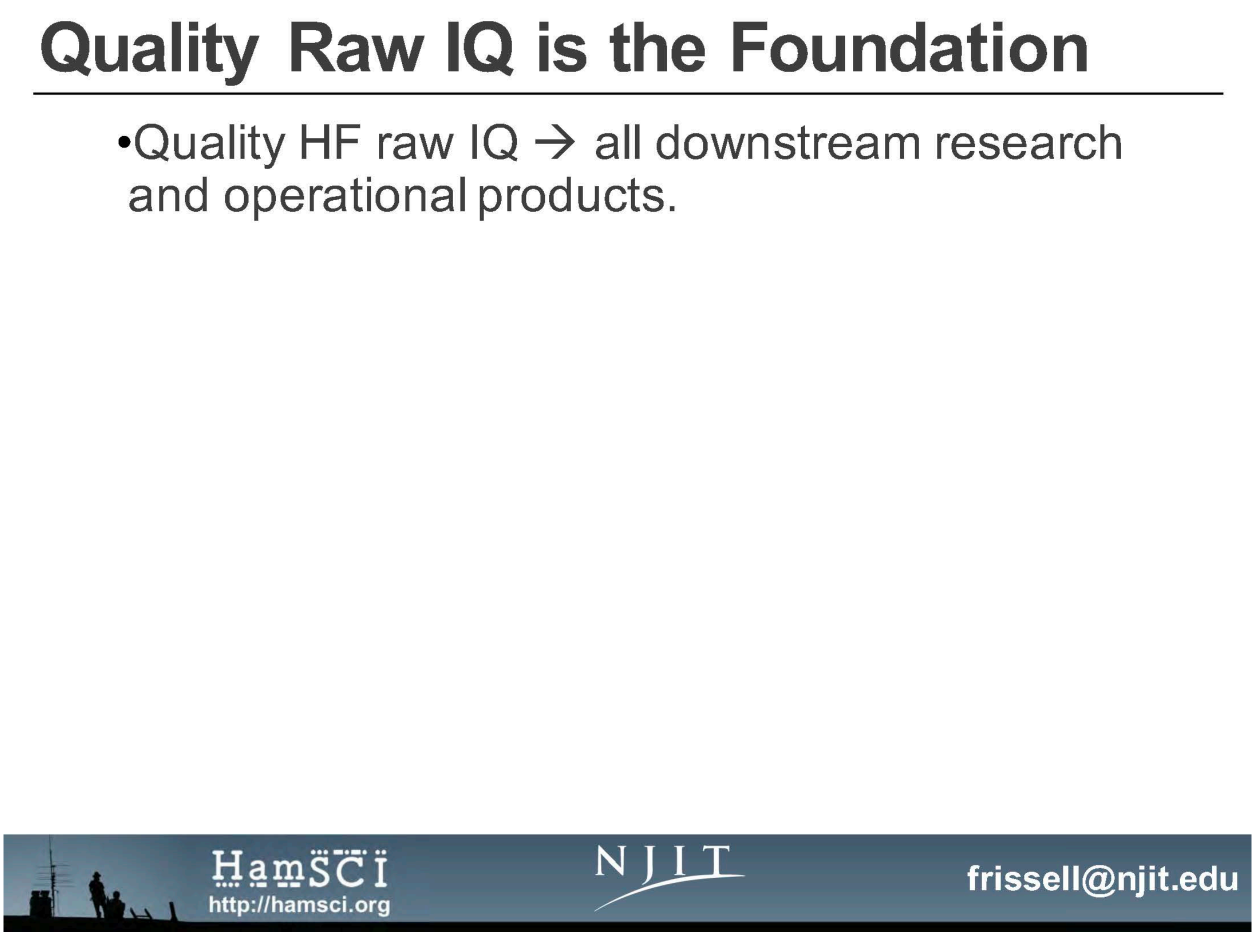
Where do we start?

- General purpose HF Receiving Instrument.

Raw IQ
0.1 - 60 MHz
Well Calibrated and Documented


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


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



Big Data

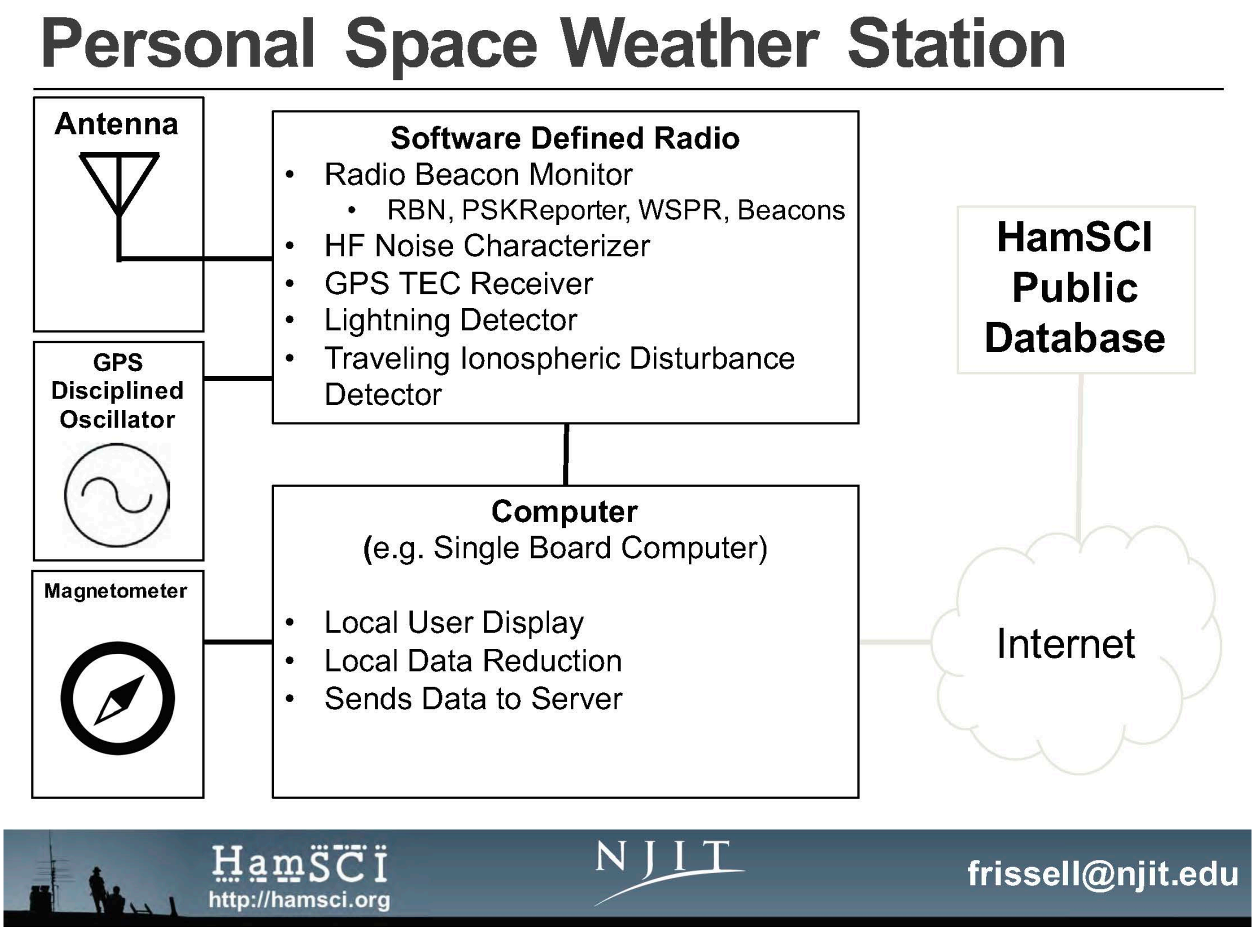
The size of the data that we are dealing with means we have a Big Data situation. Big Data groups together processing, collection, storage and visualization of large quantities of data. Data Science is the process of extracting knowledge from data. Data is collected, information is derived, and knowledge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

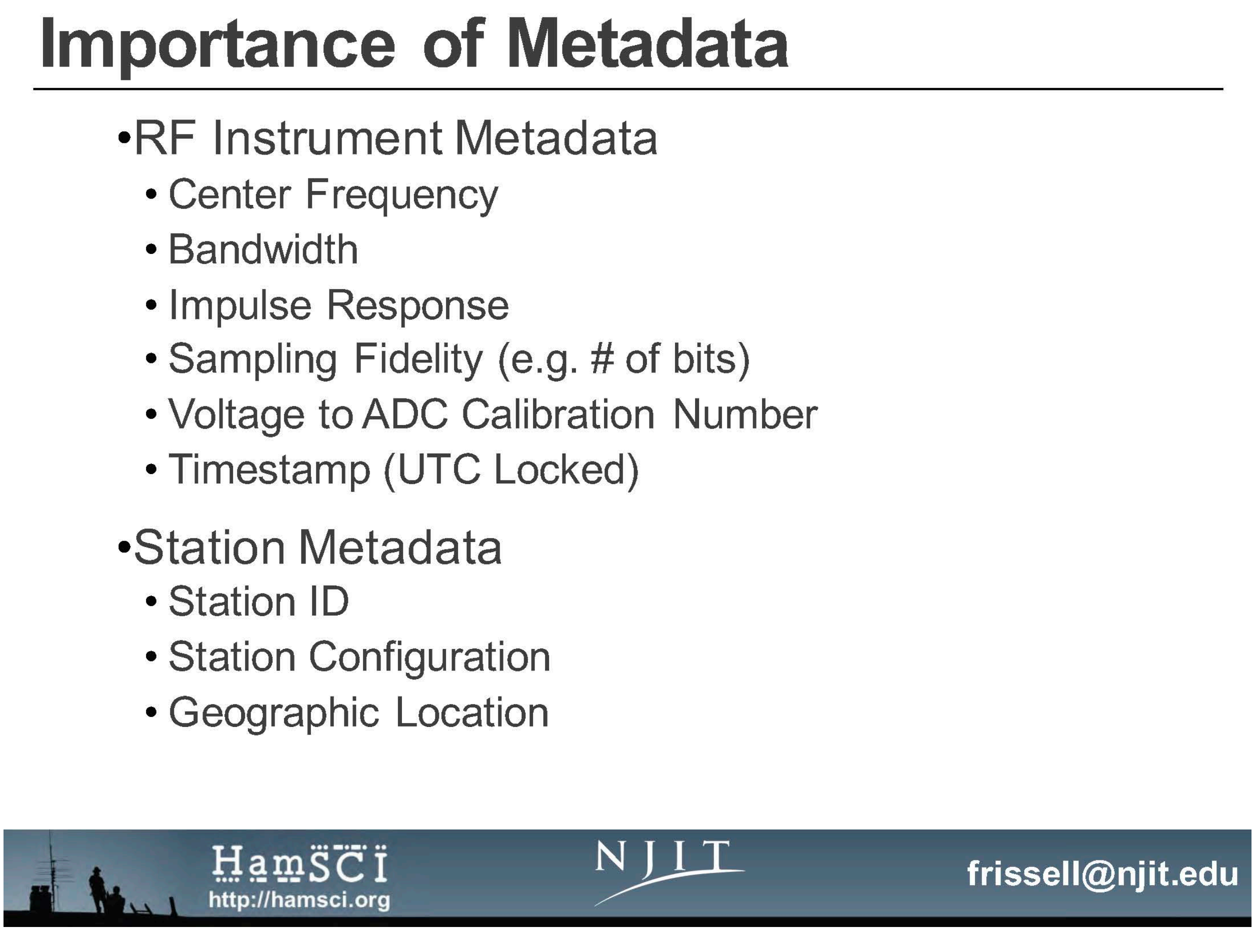
In order to get the full potential out of data, it must be available to anyone that has an interest in using it. The data must be available without unreasonable obstacles and at affordable prices. Data must be provided along with with models and equations so that interested users do not have to reinvent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

4) What type of public database is necessary for the HamSci Space Weather System?

A relational database may meet HamSci data needs. If the data has clearly documented relationships, is modeled, is less than or equal to single digit petabytes in size, has fewer than 300k writes per second, then a relational database works well.

Hadoop should be considered if HamSci has a lot of data that users do not know what to do with. In other words, if the relationship between the data is unknown. If relationships between the data cannot easily be modeled, then a relational database is not a good fit.

> "Suffering-oriented programming can be summarized like so: don't build technology unless you feel the pain of not having it." -Nathan Marz

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 Data Plane includes a HamSci Public Database. The research work product is datasets and scientific results.

Work Plan

This is arguably the most important part of the proposal as it shows how we would conduct the study.

February - March

8. **Metadata protocol research, development, integration, test, and documentation using off the shelf radio gear. Getting firsthand experience with each protocol's API is very important in order to make the best possible selection.**
9. **Individual sensor identification. Define the minimum viable product radio characteristics per sensor type. Each sensor type is listed with the minimum required individual performance specifications. These include but are not limited to bandwidth, resolution, timing, frequency range, signal to noise radio, transmit power, exfiltration bandwidth (internet upload speed), uptime, power consumption, and cost.**

February - April

11. **Prototype individual examples of each sensor. Connect each of them to the internet. Measure upload and download (for the command link) bandwidth requirements, reliability (uptime), and latency in standalone operation.**
11. **Prototype a variety of local storage options for collected data (SD card, hard disk, tape drive). Measure performance (latency, size, reliability, cost, maintenance interventions).**
11. **Prototype cloud storage options for collected data (AWS, dropbox, scaled NAS, Hadoop nodes, cloudera, etc.). Measure performance. Getting firsthand experience with each service's API is very important in order to make the best possible selection.**

May - August

11. **Begin the sensor population study. Identify and implement algorithms that enable sensor fusion. Measure performance. Adjust the minimum viable radio characteristics per sensor type. The goal is to relax individual radio specifications while preserving the quality of the aggregated data. Relaxing the radio specifications lowers the cost of the stations, increases the number of stations that can participate, and improves coverage.**

12. **Document and classify the features that begin to emerge from sensor fusion population study.**

13. **Design review for local and cloud storage options. Output of the design**

review is rank-ordered list of recommended options.

August

13. Begin the command and control functional study. The science depends on coordinated observations. Coordinated observation implies that each station is remotely controlled. Remote control implies a central authority. Functions of the central authority must be defined and documented. Central authority must have reasonable security, authorization, and authentication.

14. Incorporate satellite inputs into the model. Sensors in space can drive scheduling. Orbital pass-oriented command and control inputs currently exist as open source products.

14. Define the schedule model. The schedule model comes from the sensor fusion feature set. As observations are made, the sensor fusion feature set is modified. Models, equations, and data are all recorded as datasets in the HamSci Public Database. The schedule model produces the commands to the receivers. Example commands include: Stations are ordered to report particular tranches of data from their local ring buffer storage. Stations will point a directional antenna in a particular direction at a particular time and upload the received data. Stations will change frequency or polarization. Stations will transmit a probe. Etc.

September

11. Design Review for the Space Weather Station specification. Output of the design review is a document that can be used by and individual or group to create a compliant space weather system on their own.
13. Begin policy development for marketing, sales, customer service, technical support, delivery, and returns.
13. Begin policy development for sustaining engineering and data security.

October

11. Publish version 1.0 of individual Space Weather Station.
12. Commence productizing for custom Space Weather Station hardware (BOM, layout, SDR architecture, RF, antenna, networking, software, user guide, technical documentation, software development manual, regulatory compliance documentation, accessibility and accommodations).
13. Design Review for marketing, sales, customer service, technical support, delivery, and returns policy. Outcome of review are published policies. Design Review for sustaining engineering and data security policy. Outcome of review is sustaining engineering plan and a data security policy.
13. Identify candidates for manufacturing prototype build.

November

4. Manufacturing review for Bill of Materials. Outcome of the review is a decision on manufacturability and a cost estimate.

December

4. Design review for layout. Outcome of the review is prototype layout.
5. Decide who will prototype the various sensors and how many will be made.
6. Place order for prototypes.

January-March

2. Prototype build, test, delivery.
3. Verify and validate Space Weather Station specification.
4. Prototype Design Review. Outcome of this review is manufacturing approval for Friends and Family build.
5. Identify candidates for manufacturing Friends and Family build.

April-July

3. Friends and Family build. Limited sales to Friends and Family.
5. Friends and Family deployment. System is operational. Performance measured.
5. Friends and Family Design Review. Outcome of the review is gating item for mass manufacturing.
5. Marketing, sales, customer service, technical support, delivery, returns, sustaining engineering commences.

August

5. Mass manufacturing.
5. Marketing, sales, customer service, technical support, delivery, returns, sustaining engineering continues.

Budget

Human Resources

Engineering and support costs for research and development of a sensor fusion big data system are considerable.

Big data and sensor fusion engineering salary range is US$90,000-$130,000, according the the IEEE.

Customer support salary range is US$35,000-$43,000 a year, according to World at Work.

Technical support salary range is US$50,000-$65,000 a year, according to Salary.com.

Components engineering salary range is US$86,000-$106,000 a year, according to World at Work.

Software engineering salary range is US$80,000-$115,000 a year, according to World at Work.

RF engineering salary range is US$79,000-$92,000 a year, according to World at Work.

Data scientist salary range is US$50,000-$200,000 a year, according to World at Work.

Depending on how many human resources are required for the project, and how long they are needed, the donated time could easily exceed US$1,000,000 per year.

Data Storage

Big data storage expenses go up dramatically if the data needs to be retrieved quickly. It's not possible to scope this cost at this time due to the very large ranges of estimates.

Hosted Hadoop nodes cost approximately $4000 a year.

Station Cost

Costing a complex electronics project such as a software defined radio generally requires components engineering, development of a bill of materials, management of the logistics of a complex parts order, and working with a manufacturer throughout the process to contain and reduce costs. While a rough estimate can be made looking at comparable radio systems and drawing on experience in the field, it's not possible to accurately predict the cost for something that has not yet been engineered.

The most expensive sensor is the most capable. This assumes the maximum RF bandwidth requirement of 60MHz, an ADC resolution of ~16bits, GPSDO timing, fast and reliable data storage, fast and reliable remote control, durable RF interface, power supply, cooling, enclosure, a broadband high-performance antenna, and a high performance low noise amplifier. If local visualization is required, then an additional processor and graphics driving capability would be added. As of January 2019, cost estimates range from $1000 to $1500 for the highest performance station.

In the relatively low volumes quoted in the slide deck, a radio with the highest performance sensor is substantially more expensive than the target range of US$100 to US$500 presented at DCC.

The budget depends on the mix of sensors that are chosen for manufacture. Sensors range from the highest performance (and most expensive) to very modest (and still very useful) sensors. This mix is unknown at this time. The total number of stations given in the DCC presentation (1000) is much higher than any coordinated amateur or open source distributed network with similar capabilities in existence to date. As of January 2019, SatNOGS had 300 coordinated satellite observing stations after two years of growth. All SatNOGS stations are built or integrated by the operator.

Some simple budget models for the cost of 1000 stations are listed below.

100% High Performance
1000 of the highest performing stations, at a cost of US$1500, is US$1,500,000.

Uniform Distribution
A mix of ten sensor types, ranging from US$100 to US$1000, with uniform distribution among 1000 stations, is US$550,000

Bimodal Favoring Less Expensive
A mix of two sensor types, where 90% are inexpensive multi-sensor station at US$300, and 10% are the more expensive high performance type at US$1500, is US$420,000

Heterogenous Populations
Open specifications allow for operators to construct their own stations. Relaxed requirements due to sensor fusion and Data Plane algorithms reduce the cost of the custom hardware and allow for off-the-shelf options, as long as the radio complies with whatever is necessary to access the Data Plane. If half of the stations were constructed or purchased off the shelf, the cost goes down. It won't go down by half, due to volume pricing for some of the radio components. However, the cost for a DIY-oriented heterogenous receiver network could be low six figures.

Maintenance Costs

Data scientist as manager salary range is US$110,000-130,000 a year, according to the IEEE.
Network engineer salary range is US$78,000-$88,000 a year, according to the IEEE.

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