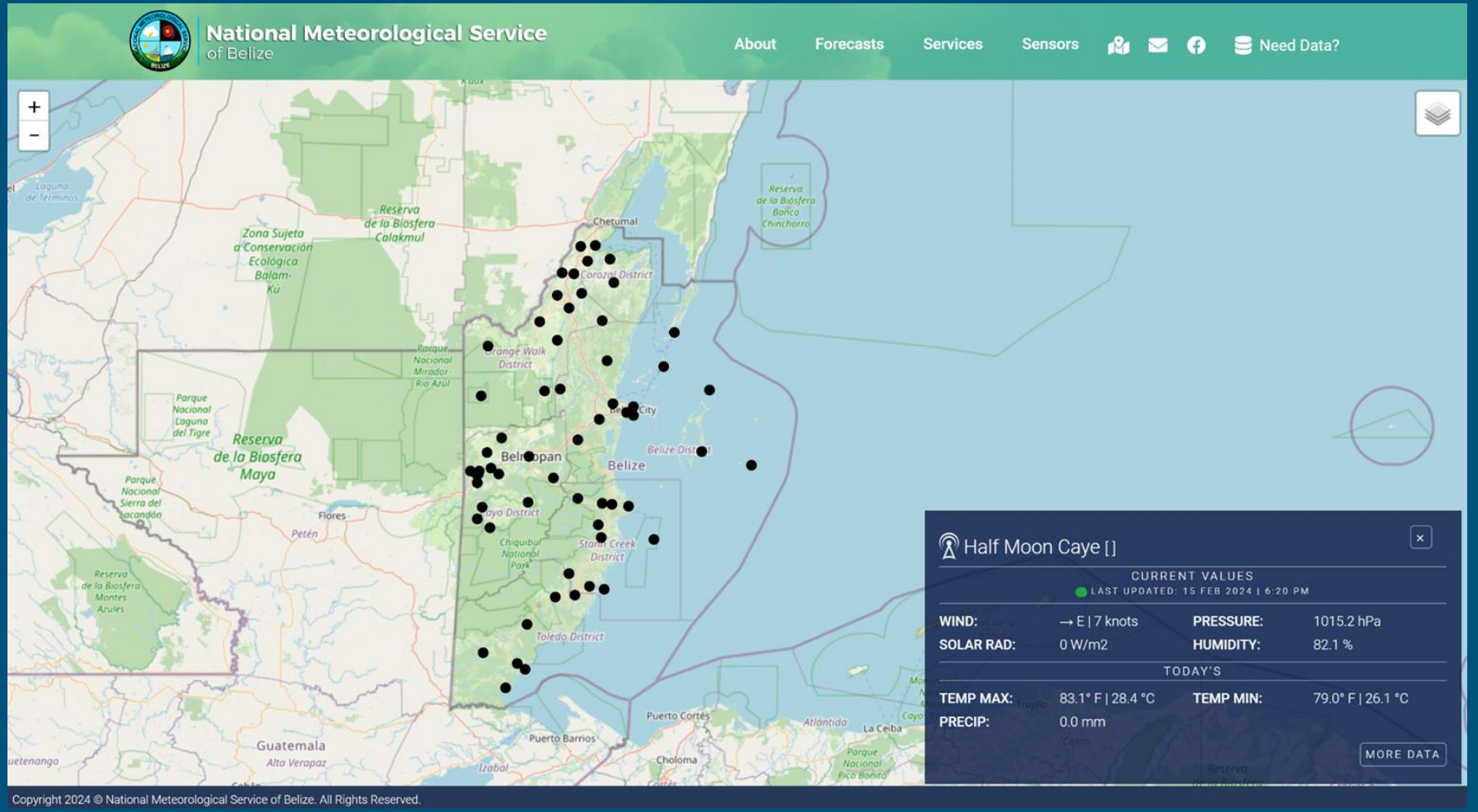
Standardization of First-mile Data Collection from Automatic Observing Platforms

Data collection from Automatic Weather Stations in Belize

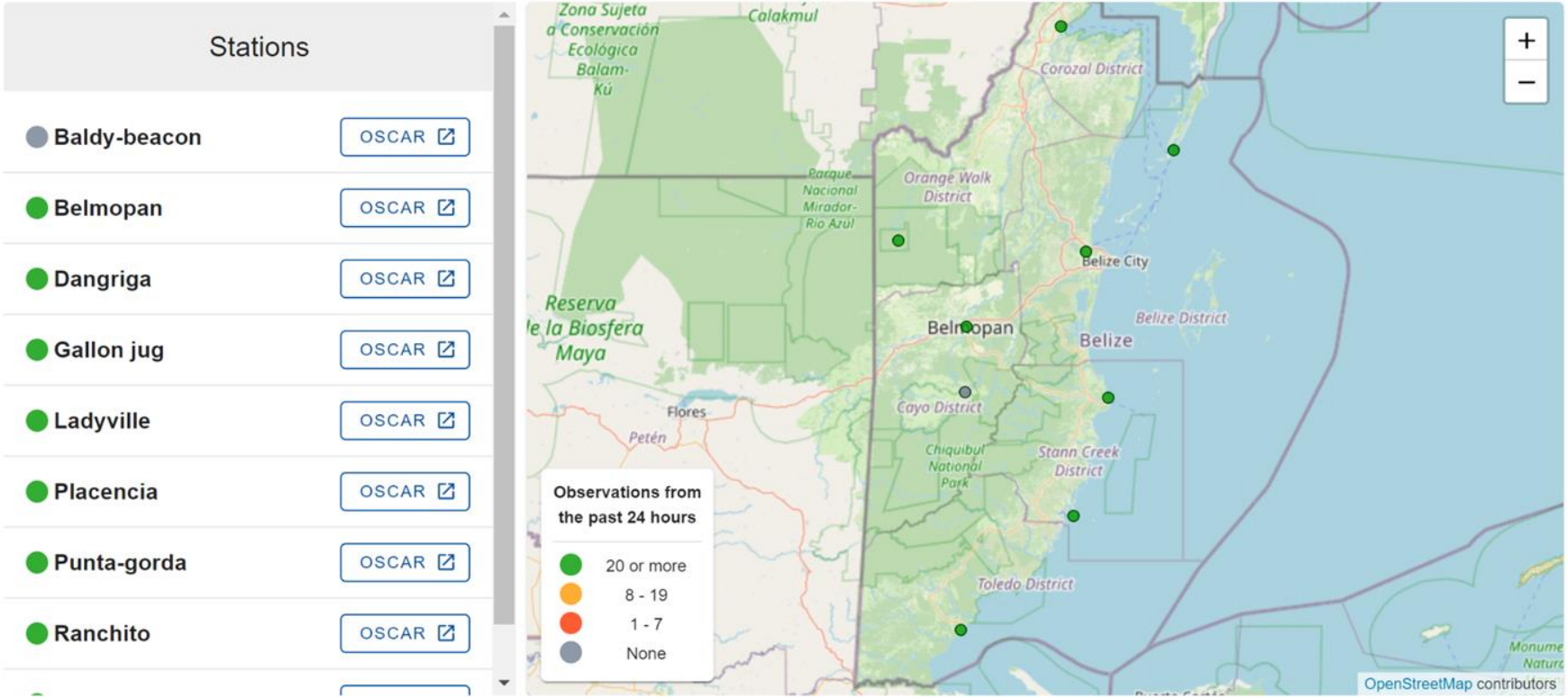
Dwayne Scott
dscott@nms.gov.bz
Electronic Technician

How is data displayed here? - https://nms.gov.bz



https://wis.nms.gov.bz

HOME | DOCUMENTATION

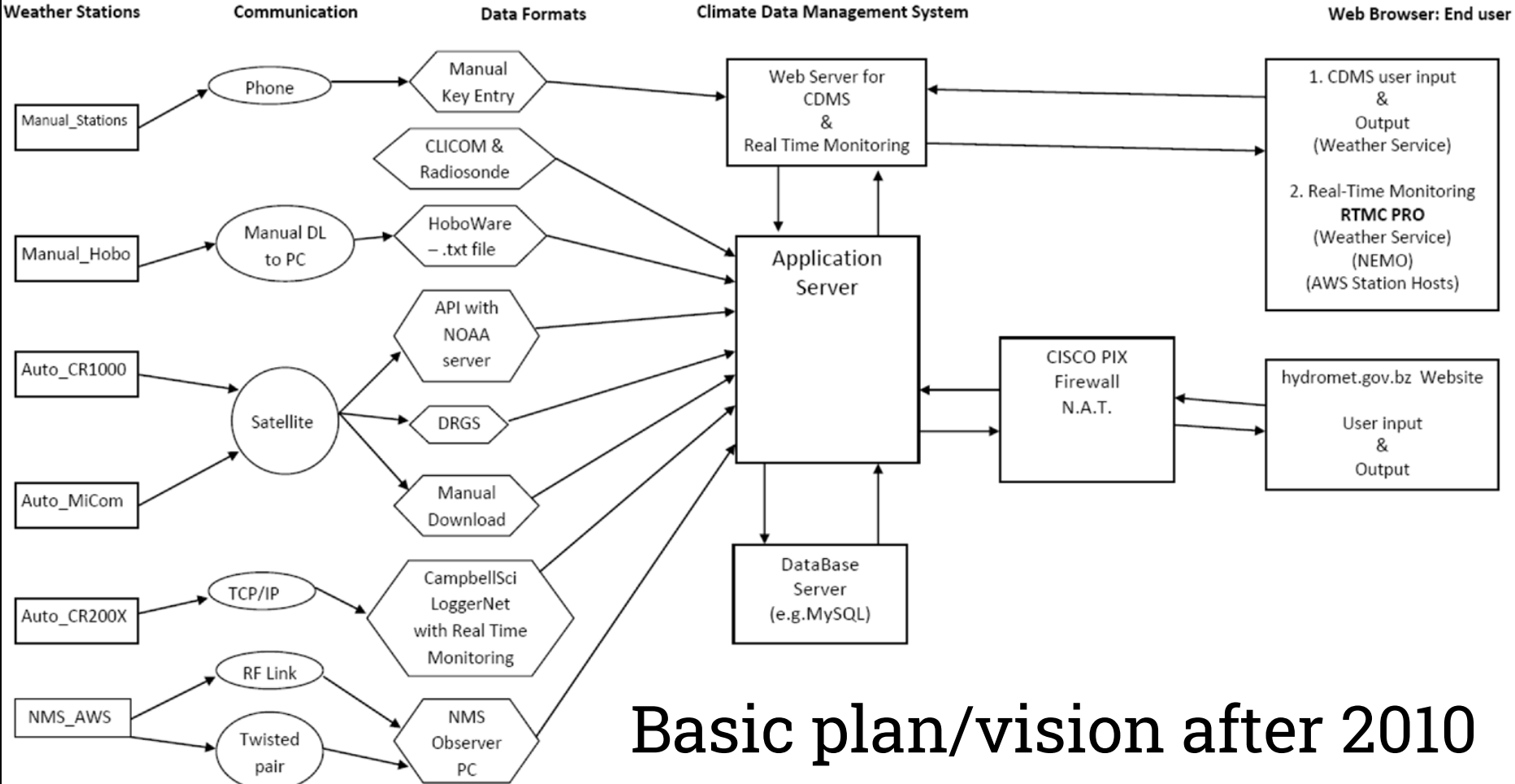


Background

- Started working at the NMS - Feb 2010
- Hong Kong - AWS network 1 week workshop - Dec 2010

Figure 1.1

Belize Weather Station Network (BWSN) Overview



Basic plan/vision after 2010

Figure 1.1

Belize Weather Station Network (BWSN) Overview

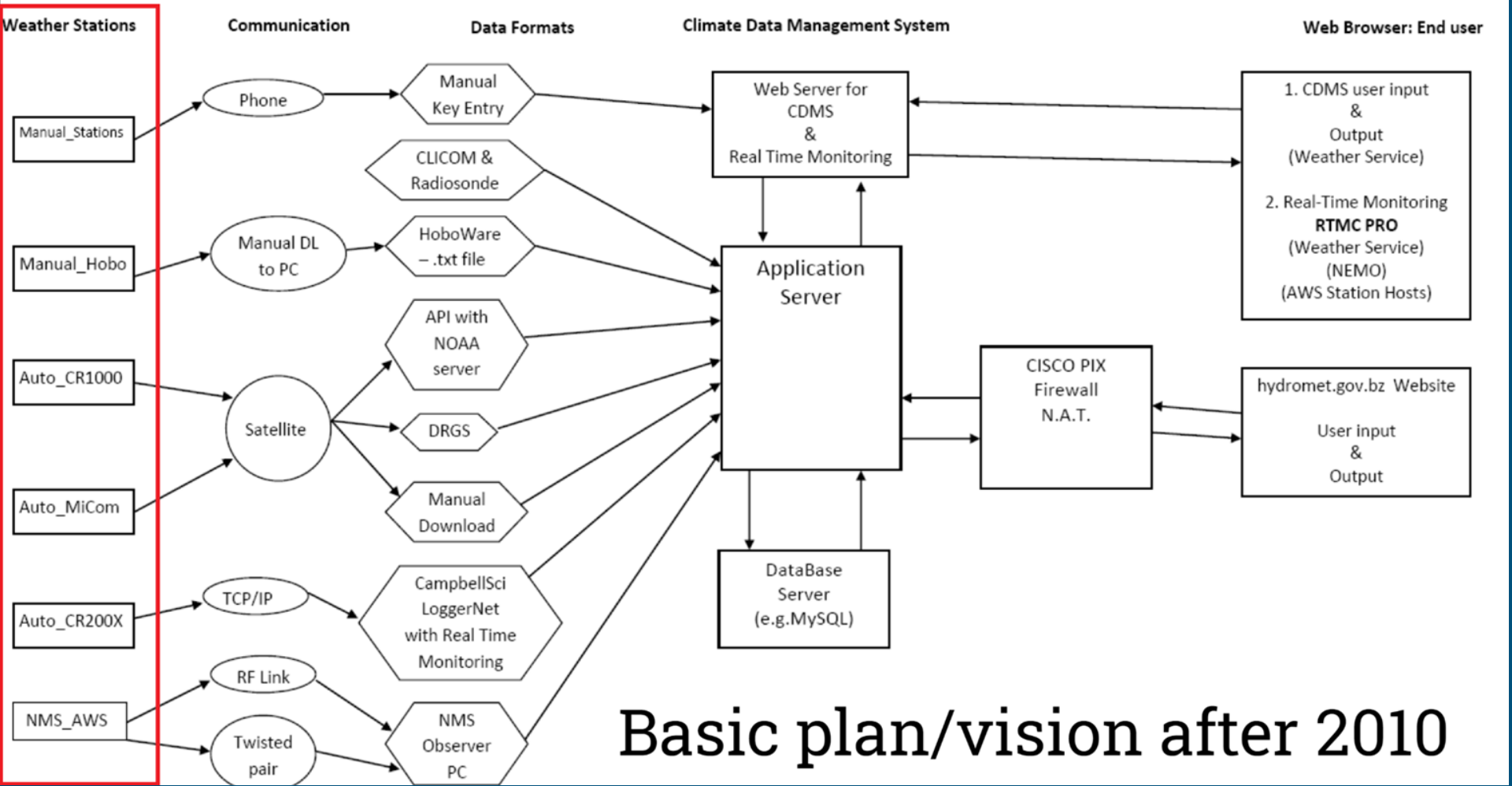
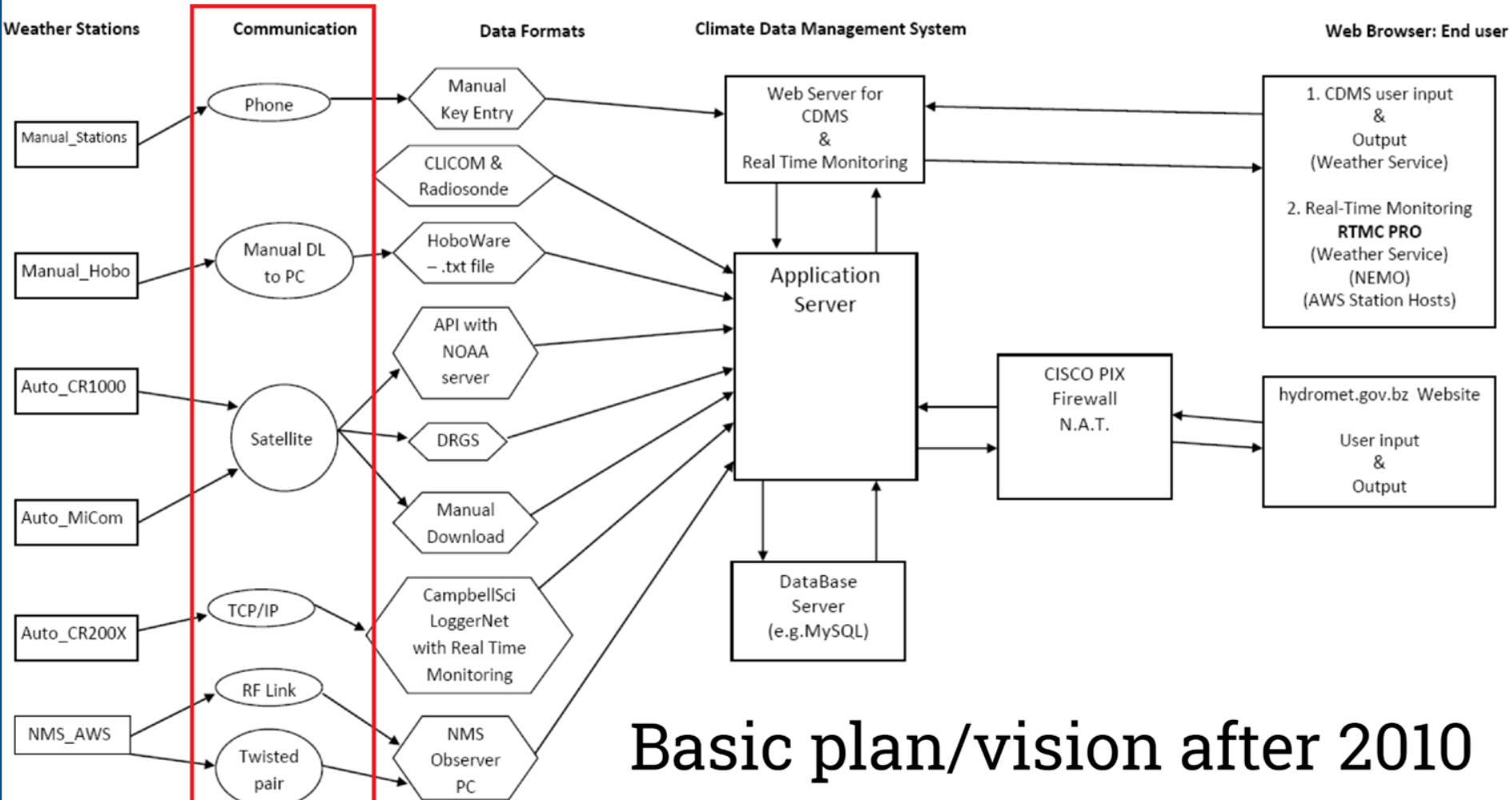


Figure 1.1

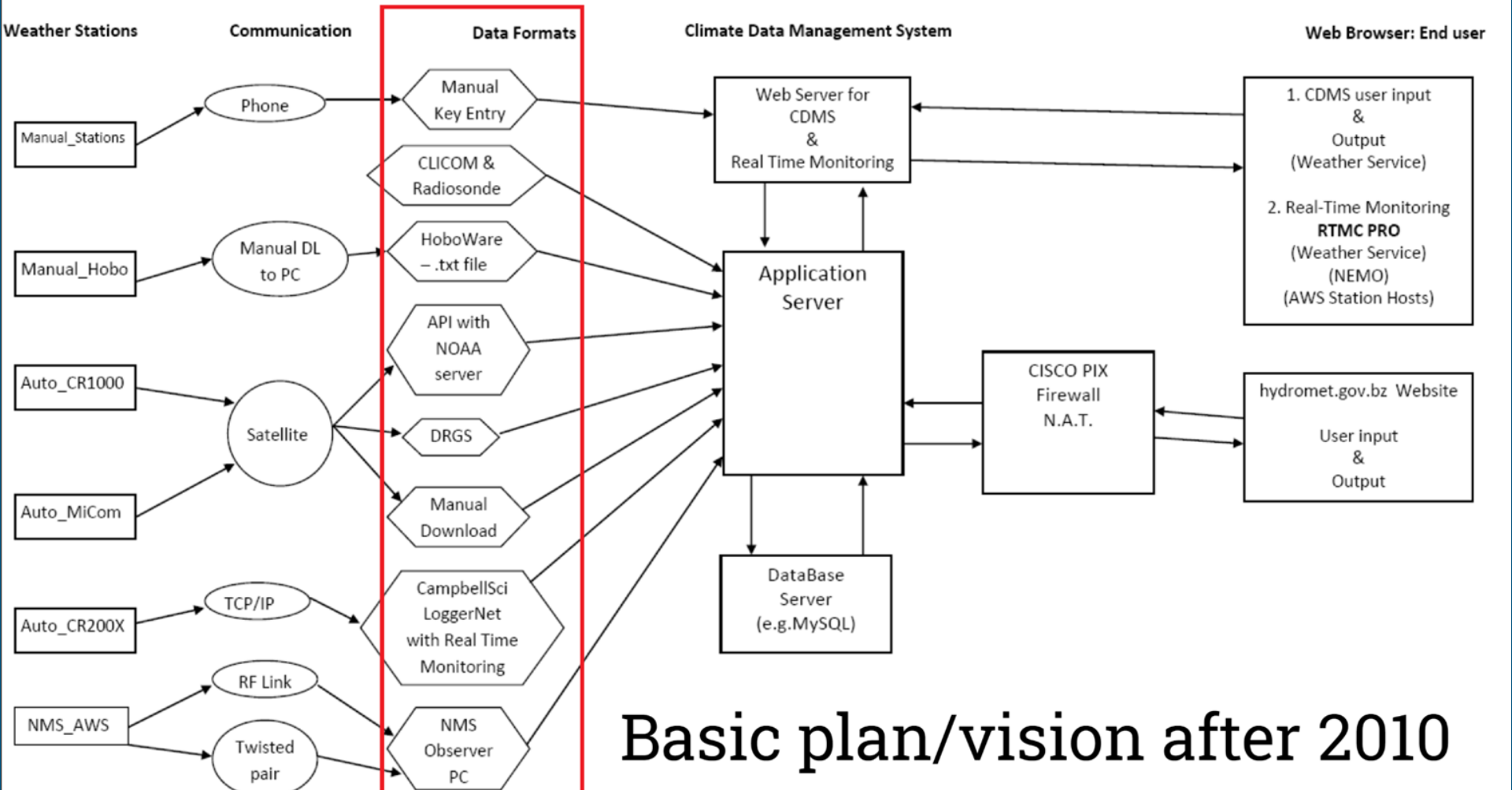
Belize Weather Station Network (BWSN) Overview



Basic plan/vision after 2010

Figure 1.1

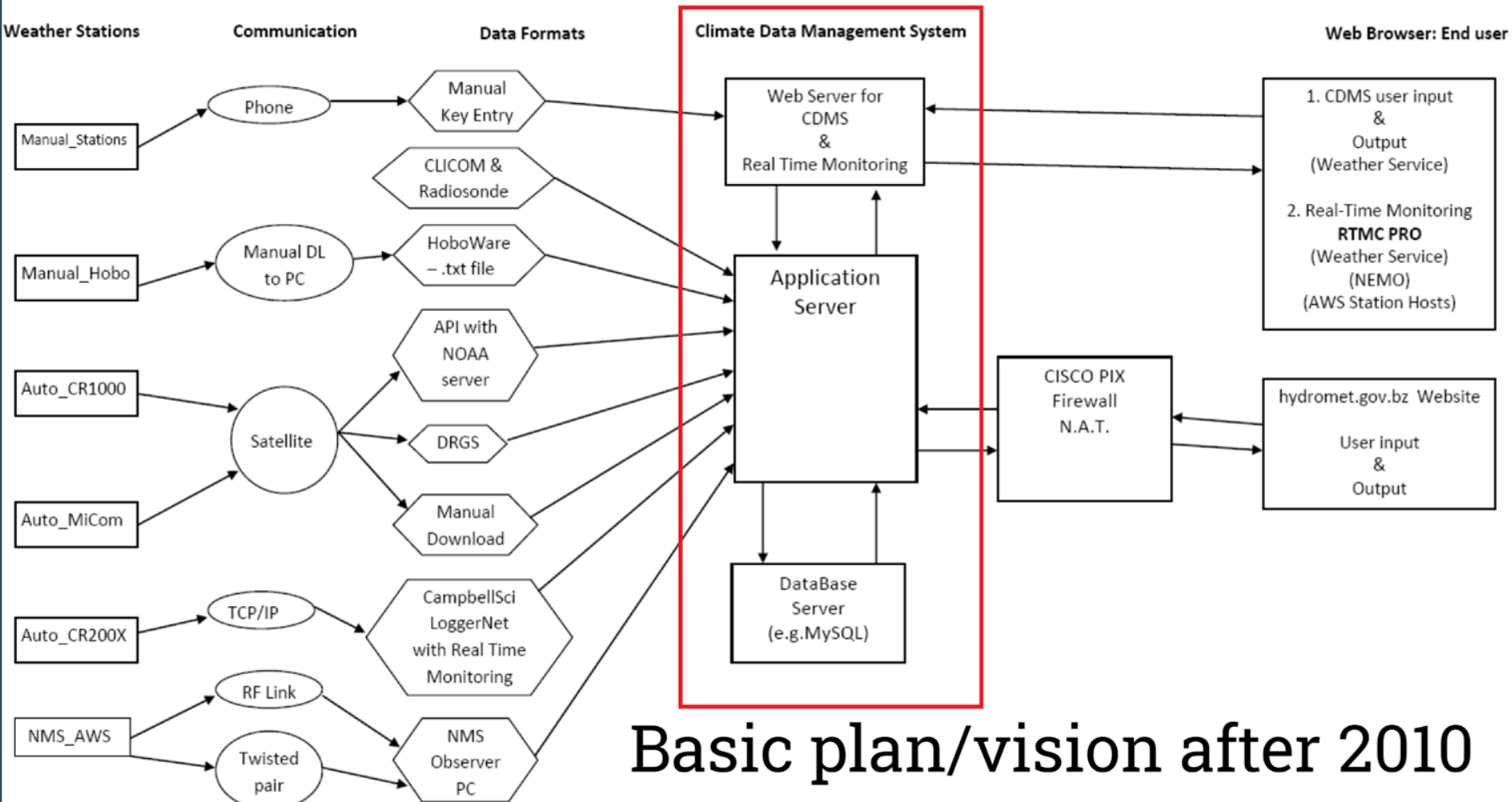
Belize Weather Station Network (BWSN) Overview



Basic plan/vision after 2010

Figure 1.1

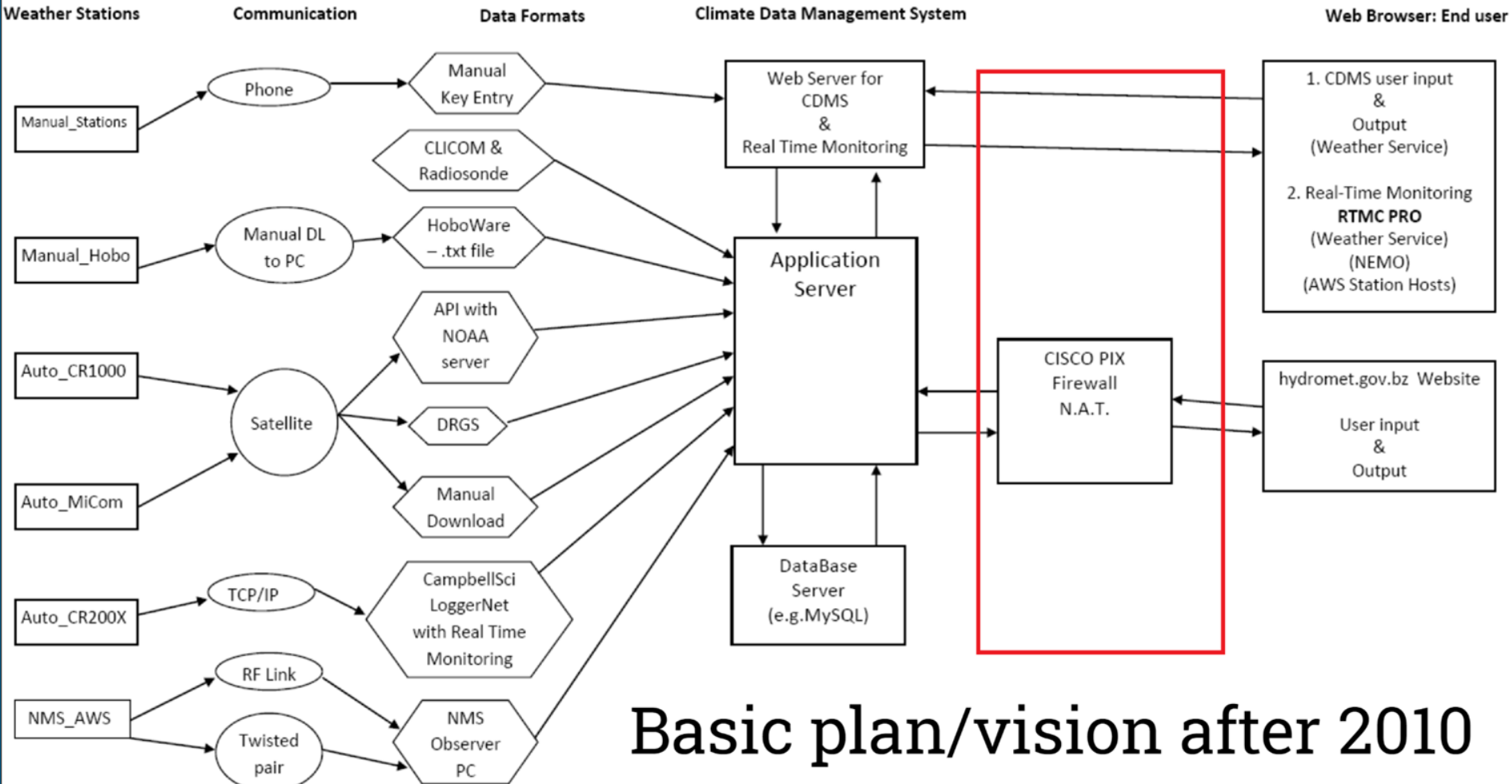
Belize Weather Station Network (BWSN) Overview



Basic plan/vision after 2010

Figure 1.1

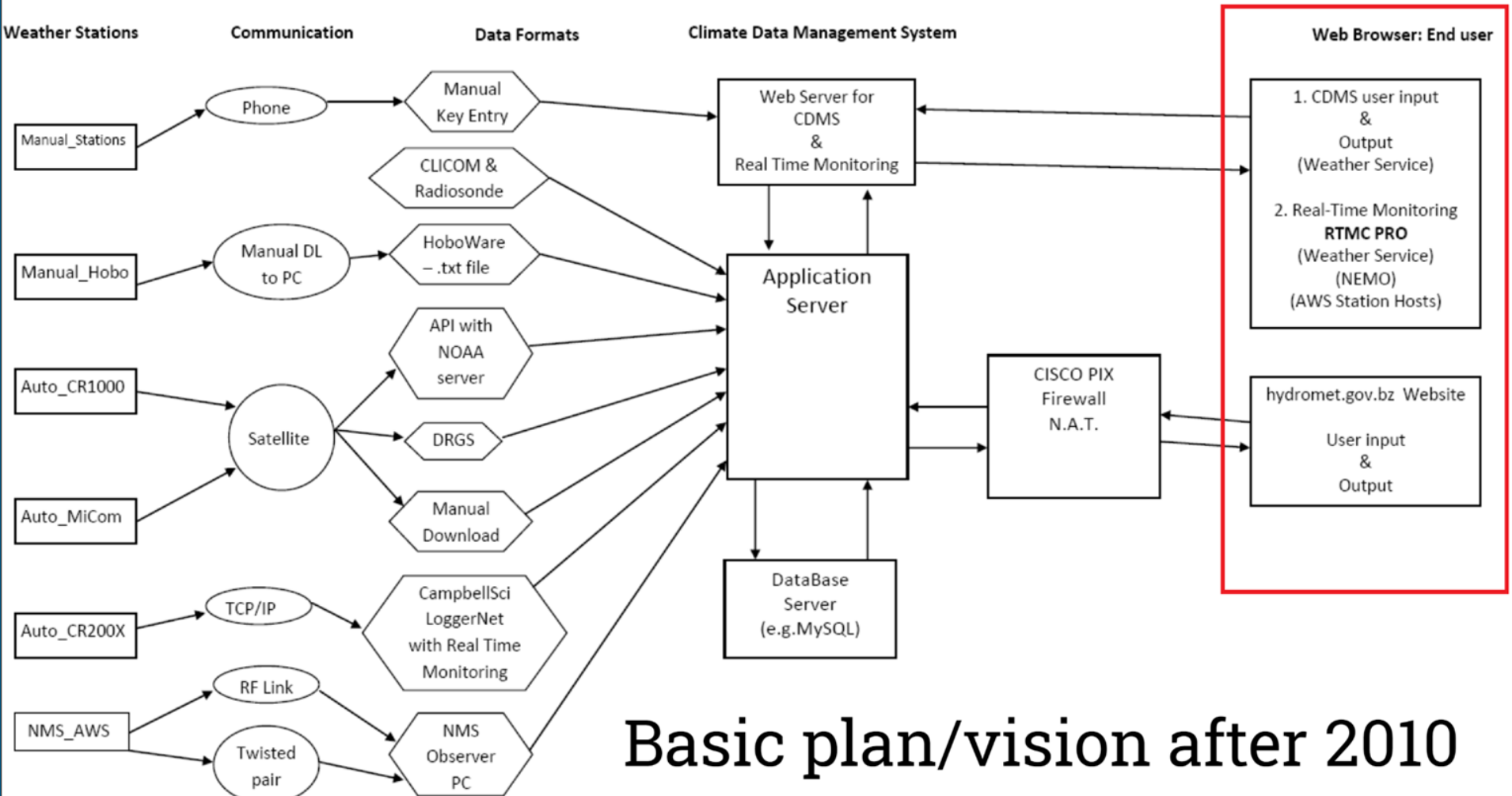
Belize Weather Station Network (BWSN) Overview



Basic plan/vision after 2010

Figure 1.1

Belize Weather Station Network (BWSN) Overview



Basic plan/vision after 2010

Main parts of Observation Network

1. In-Situ and Remote Sensing
   - *Weather stations(Manual and Automatic), Radar, Lightning & Satellite*
   - *Sensors & Other Accessories - Panels, batteries, mounts, masts etc*
2. Data Transmission
   - *Satellite DCPs, Mobile, TCP/IP, RF, Manual download*
   - *Modems & Accessories*
3. Data Reception
   - *How will raw data be collected, archived & sent to Database*
   - *Monitoring data reception*
4. Data Storage and processing
   - *Where will data be sent once received? How will it be processed?*
   - *QC checks; Metadata for stations and sensors*
5. Data output and display
   - *Message outputs - SYNOP, METAR, CLIMAT, ATIS*
   - *RTMC Pro, Climate Data Management System(CDMS), NMS Website*

Data Flow

AWS(Data logger) (Stage 1)

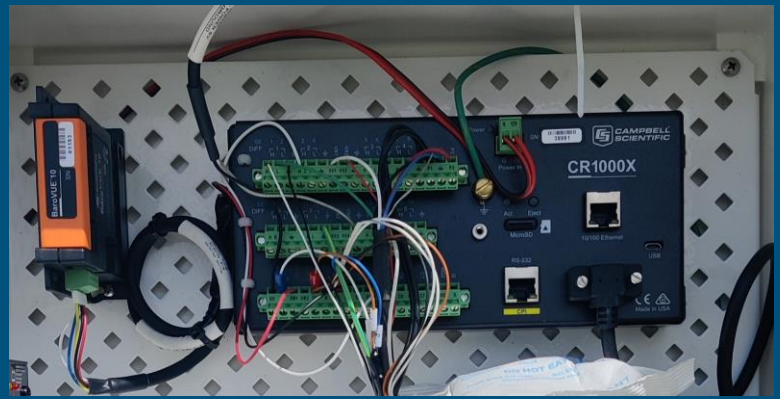
Raw Data Storage (Stage 2)

CDMS - Data Processing (Stage 3)

Data Output - Display/Access/Transmission(Stage 4)

Stage 1 - AWS (Data logger)

1. At each weather station, data loggers are programmed to take meteorological samples every 10 seconds(6 samples each min)
    - wind measurements are sampled each second(1hz minimum)
2. The Samples are held in a buffer and select computations from those samples are stored every 5 mins(data Tx) or 10 mins(Satellite Tx)
    a. Sample/Instantaneous - Temperature, Relative Humidity, Barometric Pressure, Solar Radiation
    b. Average - Wind Speed and Direction,
    c. Maximum - Temperature and Relative Humidity
    d. Minimum - Temperature and Relative Humidity
    e. Total/Sum - Rainfall and Solar Radiation
    f. Standard deviations - Wind Speed and direction
    g. Any other calculation that is deemed necessary



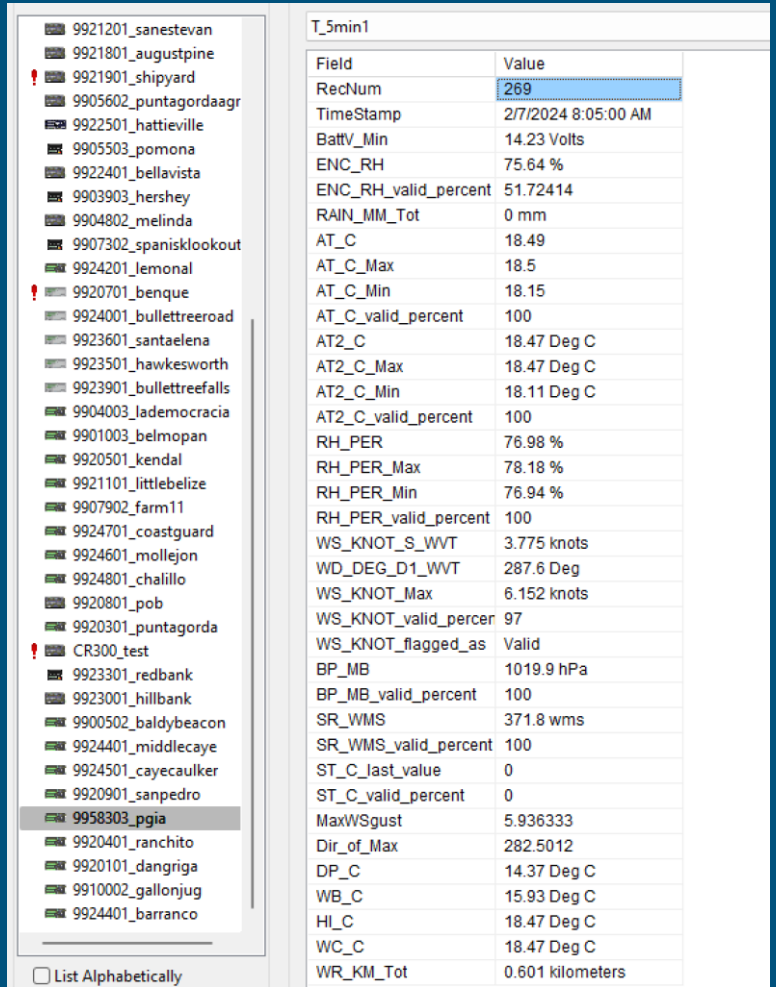
Stage 1 -AWS (Datalogger)

1. Currently testing basic QC procedures on datalogger with data quality value

This will give us an idea of the quality of the raw data calculated and stored in the 5 min tables

- Good data
- any skipped scans while making the computations?
- if there were skipped scans how many
- measurements failed QC checks?
- If so how many of those measurements were outside the range

Data Tx



Field	Value
RecNum	269
TimeStamp	2/7/2024 8:05:00 AM
BattV_Min	14.23 Volts
ENC_RH	75.64 %
ENC_RH_valid_percent	51.72414
RAIN_MM_Tot	0 mm
AT_C	18.49
AT_C_Max	18.5
AT_C_Min	18.15
AT_C_valid_percent	100
AT2_C	18.47 Deg C
AT2_C_Max	18.47 Deg C
AT2_C_Min	18.11 Deg C
AT2_C_valid_percent	100
RH_PER	76.98 %
RH_PER_Max	78.18 %
RH_PER_Min	76.94 %
RH_PER_valid_percent	100
WS_KNOT_S_WVT	3.775 knots
WD_DEG_D1_WVT	287.6 Deg
WS_KNOT_Max	6.152 knots
WS_KNOT_valid_percer	97
WS_KNOT_flagged_as	Valid
BP_MB	1019.9 hPa
BP_MB_valid_percent	100
SR_WMS	371.8 wms
SR_WMS_valid_percent	100
ST_C_last_value	0
ST_C_valid_percent	0
MaxWSgust	5.936333
Dir_of_Max	282.5012
DP_C	14.37 Deg C
WB_C	15.93 Deg C
HI_C	18.47 Deg C
WC_C	18.47 Deg C
WR_KM_Tot	0.601 kilometers

Satellite Tx

```
T1220,12,78,0,209G,208G,769G,37G,3384G,135G,10184G,0B,0B,NAN,
T1230,12,78,0,209G,207G,771G,29G,3351G,84G,10184G,51B,0B,NAN,
T1240,12,78,0,209G,207G,774G,37G,3426G,113G,10186G,223G,0B,NAN,
T1250,13,78,0,211G,208G,782G,31G,3384G,90G,10187G,630G,0B,NAN,
T1300,14,77,0,212G,211G,812G,28G,3272G,87G,10188G,1013G,0B,NAN,
T1310,14,74,0,212G,210G,829G,37G,3147G,90G,10188G,1380G,0B,NAN,
T1120,12,78,0,210G,209G,736G,49G,3529G,139G,10177G,0B,0B,NAN,
T1130,12,78,0,210G,209G,741G,45G,3451G,142G,10178G,0B,0B,NAN,
T1140,12,78,0,211G,209G,754G,46G,3427G,126G,10179G,0B,0B,NAN,
T1150,12,78,0,211G,210G,757G,46G,3398G,113G,10181G,0B,0B,NAN,
T1200,12,78,0,210G,209G,759G,37G,3420G,110G,10182G,0B,0B,NAN,
T1210,12,78,0,209G,209G,771G,35G,3365G,97G,10183G,0B,0B,NAN,
T1020,12,78,0,211G,210G,745G,60G,35G,142G,10173G,0B,0B,NAN,
T1030,12,78,0,210G,208G,736G,55G,20G,116G,10174G,0B,0B,NAN,
T1040,12,78,0,209G,208G,733G,54G,3596G,132G,10175G,0B,0B,NAN,
T1050,12,78,0,209G,208G,728G,59G,3598G,139G,10175G,0B,0B,NAN,
T1100,12,78,0,209G,208G,737G,50G,3578G,119G,10176G,0B,0B,NAN,
T1110,12,78,0,210G,208G,738G,55G,3564G,119G,10176G,0B,0B,NAN,
T0920,12,81,0,209G,208G,810G,42G,3427G,106G,10172G,0B,0B,NAN,
T0930,12,81,0,210G,209G,817G,42G,3421G,129G,10172G,0B,0B,NAN,
T0940,12,80,0,210G,209G,818G,44G,3453G,110G,10172G,0B,0B,NAN,
T0950,12,80,0,211G,209G,790G,51G,0G,132G,10172G,0B,0B,NAN,
T1000,12,79,0,211G,211G,774G,57G,53G,126G,10172G,0B,0B,NAN,
T1010,12,78,0,211G,210G,759G,57G,22G,126G,10173G,0B,0B,NAN,
```

Raw Data Storage

AWS(Data logger) (Stage 1)

Raw Data Storage (Stage 2)

Stage 2 - Raw Data Storage

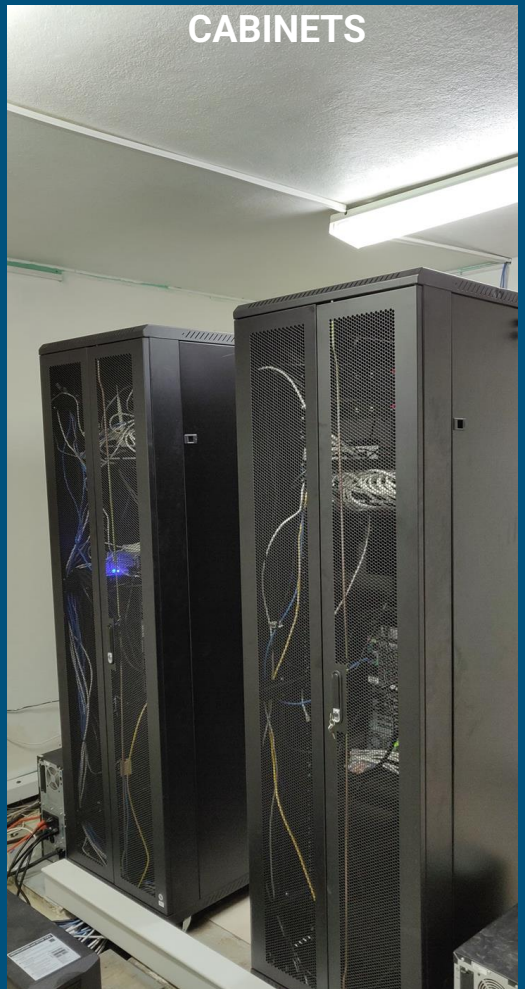
1. Once the 5/10 minute data is measured and stored on the datalogger it is transmitted to NMS servers or directly to our CDMS via NOAA's LRGS(GOES DCPs).
2. Data files transmitted to the NMS using - RF, Wifi, Mobile, Satellite Modems.
3. Data files are received and stored on a data collecting server which also acts as an FTP server - called our "Data Store"
    a. If communication is broken with the weather station, data is stored on the data logger until communication is re-established
4. The "Data Store" stores a wide range of raw data from
    a. Weather Stations
    b. Radar files
    c. Lightning Detection network
    d. Database backup files
    e. Observation Network configuration files
5. The NMS chose this repository so that computer systems and select stakeholders wishing to get data can "Pull" data from the repository instead of the NMS bearing the responsibility of "pushing" data

Stage 2 - Raw Data Storage

Raw Data storage Computer

- Workstation PC
- Raid configured with redundant drives
- Windows 11
- AWS data reception software - manages our network of AWSs
- FTP Server software
- Python and code editor
- Necessary ports and software must be configured
- Hosts and runs important scripts
  - Hourly Data ingestion from Satellite stations
  - Hourly CSV generation for WIS2box
  - Daily Backups to cloud storage

SERVER CABINETS



WORKSTATION PCs (Collect)s Data

DATABASE SERVERS

C.D.M.S.

AWS(Data logger) (Stage 1)

Raw Data Storage (Stage 2)

CDMS - Data Processing (Stage 3)

STAGE 3 - Data Processing

CLIMATE DATA MANAGEMENT SYSTEM

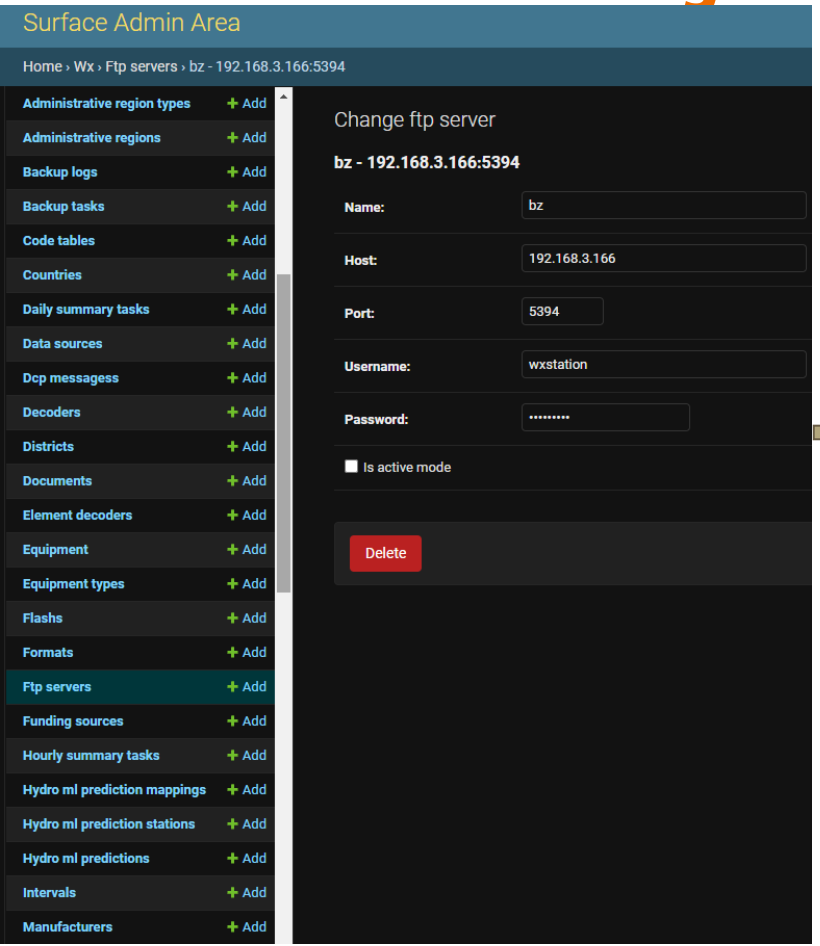
S.U.R.F.A.C.E. CDMS

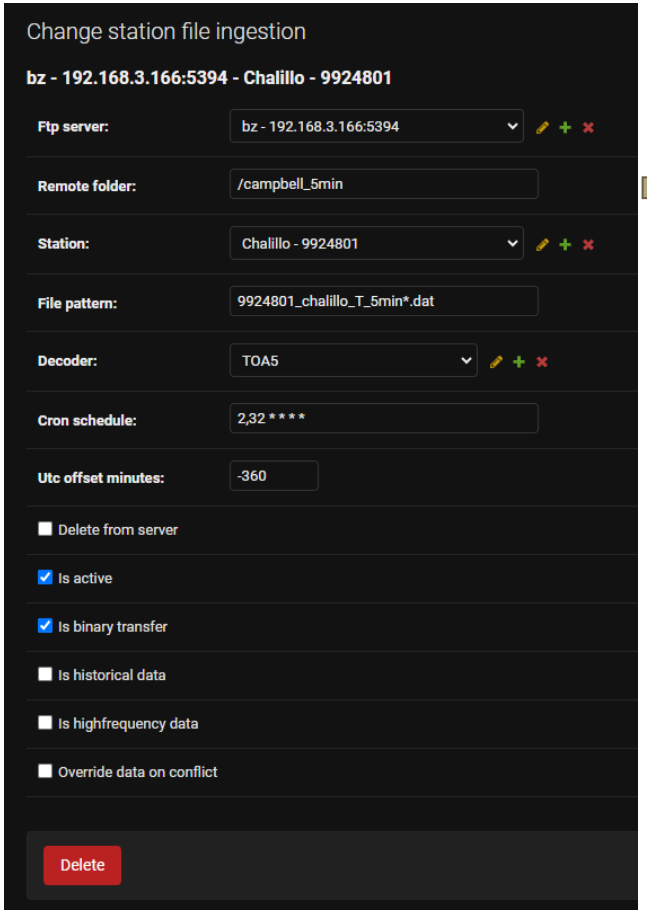
SURFACE

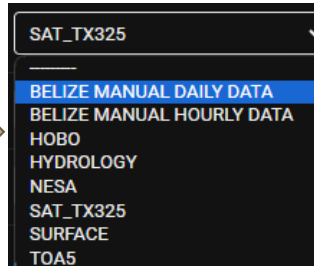
System for **U**nified **R**eal-time monitoring and **F**orcasting of **A**tmospheric and **C**limatic **E**vents

- Ingests multiple data formats from AWSs
- Aggregates AWS raw data(Sub-hourly)
  - Hourly, Daily, Monthly, Yearly
- Monitors Station performance & data flow
- Stores AWS and instrument Metadata
- Manual & Automatic QC Procedures
- Postgres with TimescaleDB
- Python/Django Backend
- VueJS frontend
- Restful API
- Docker stack
- Open Source/OpenCDMS
  - Project - https://github.com/opencdms/surface
  - Test site - http://surface.opencdms.org/

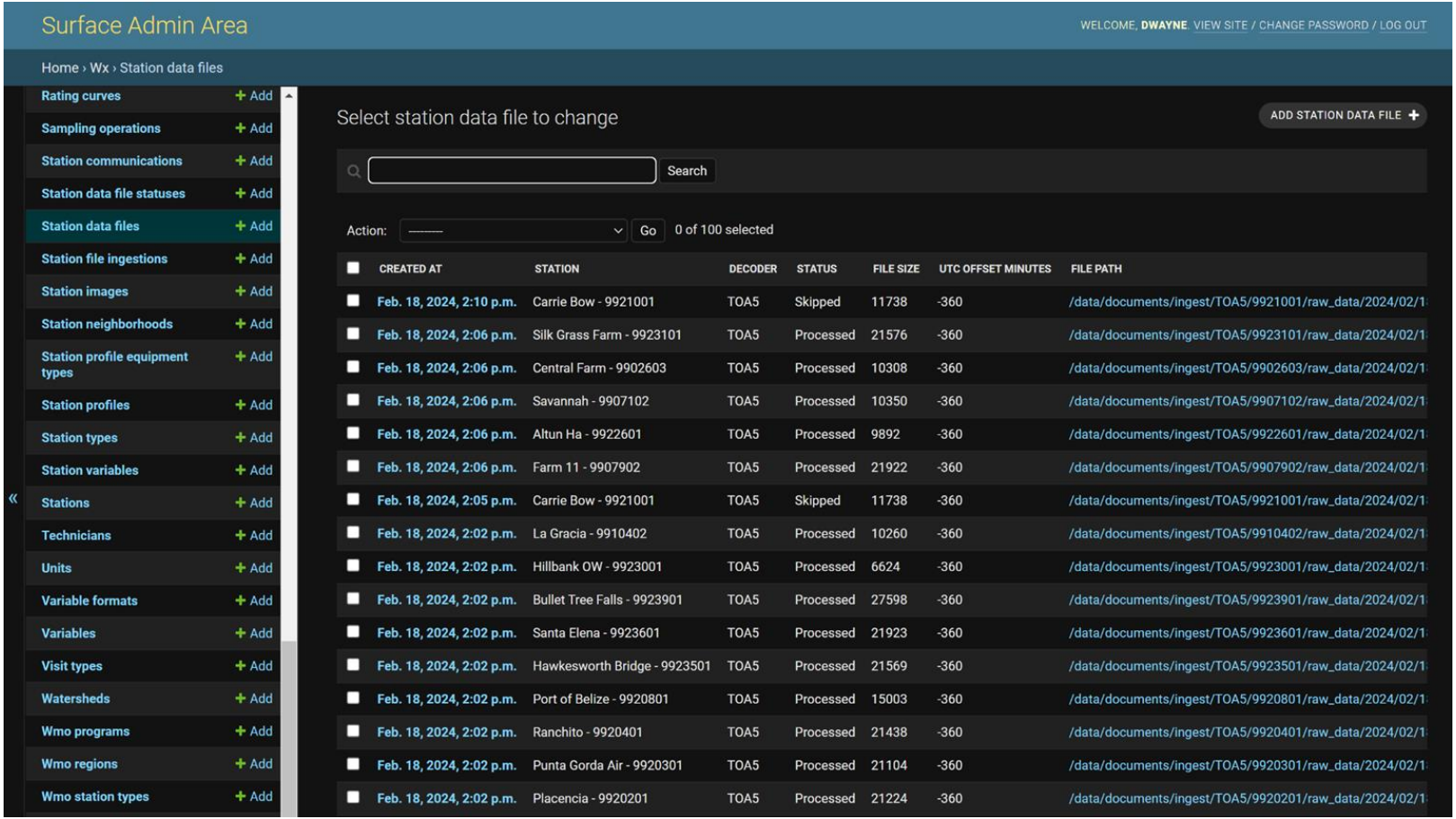
How is raw data ingestion setup? (moving from data store to CDMS)



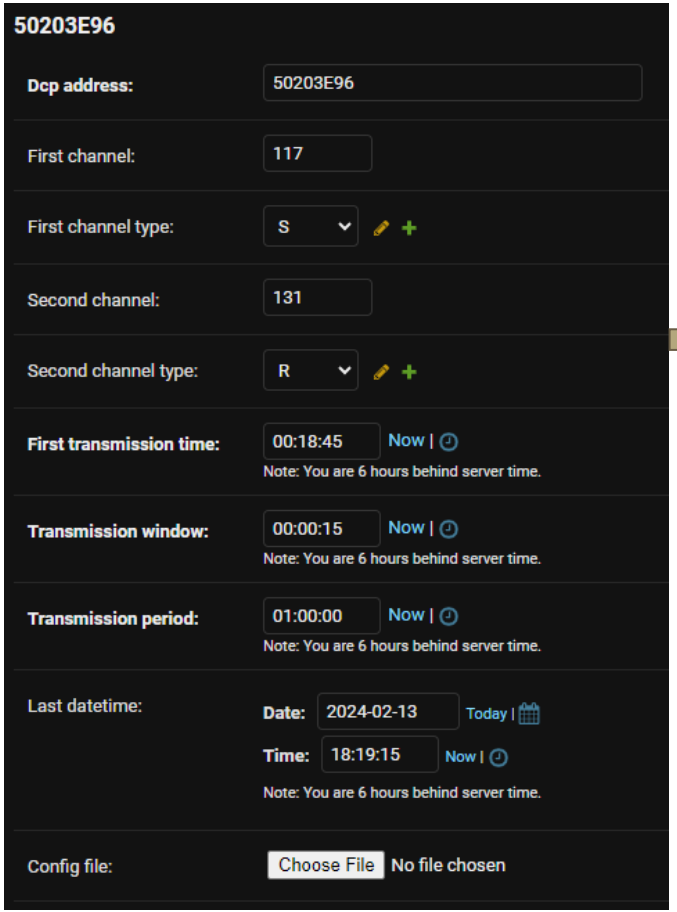


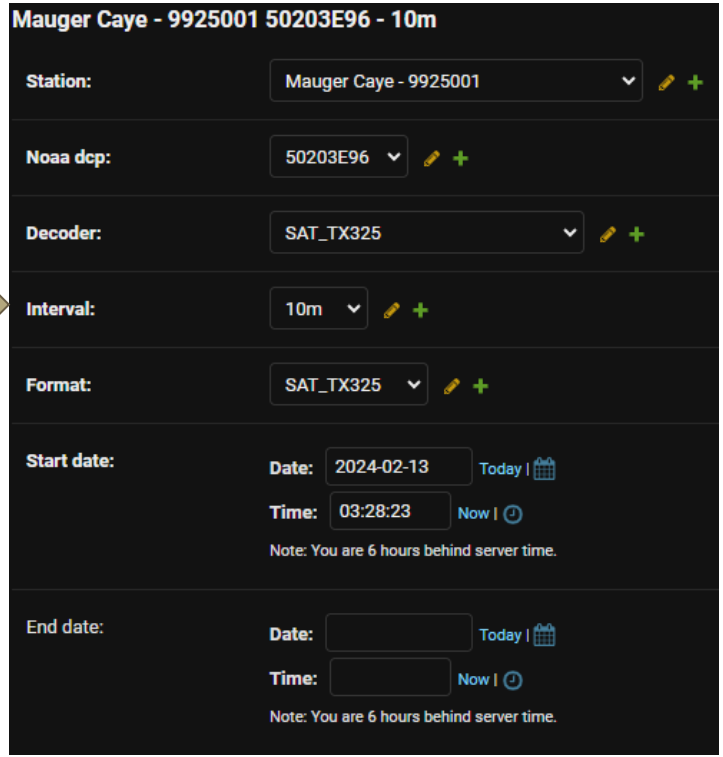


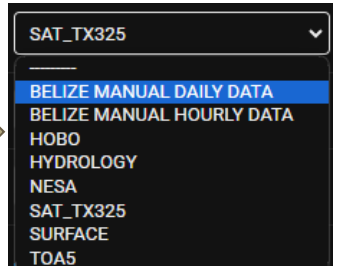
Checking Data file ingestion…



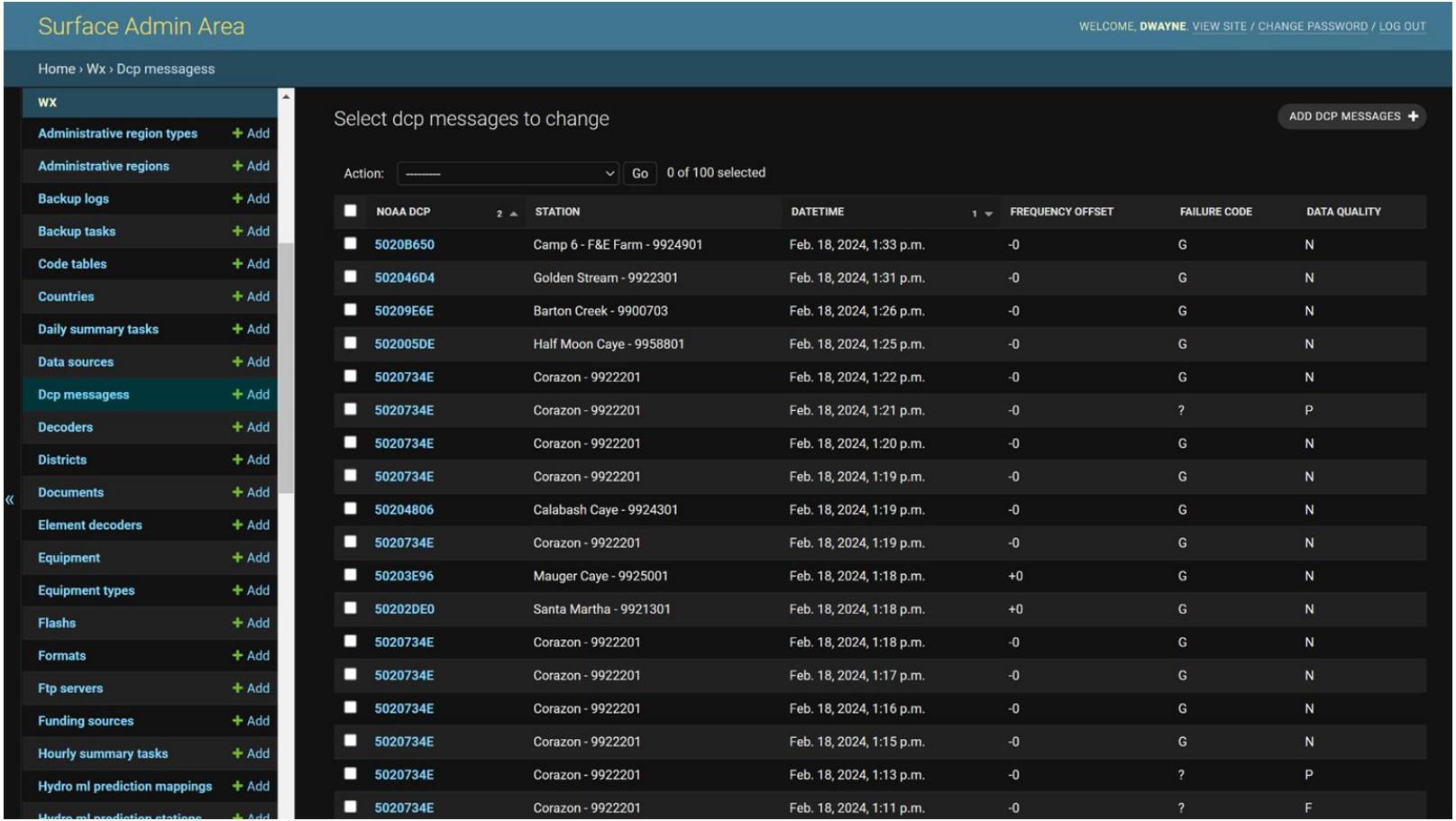
GOES /LRGS/ Satellite ingestion setup





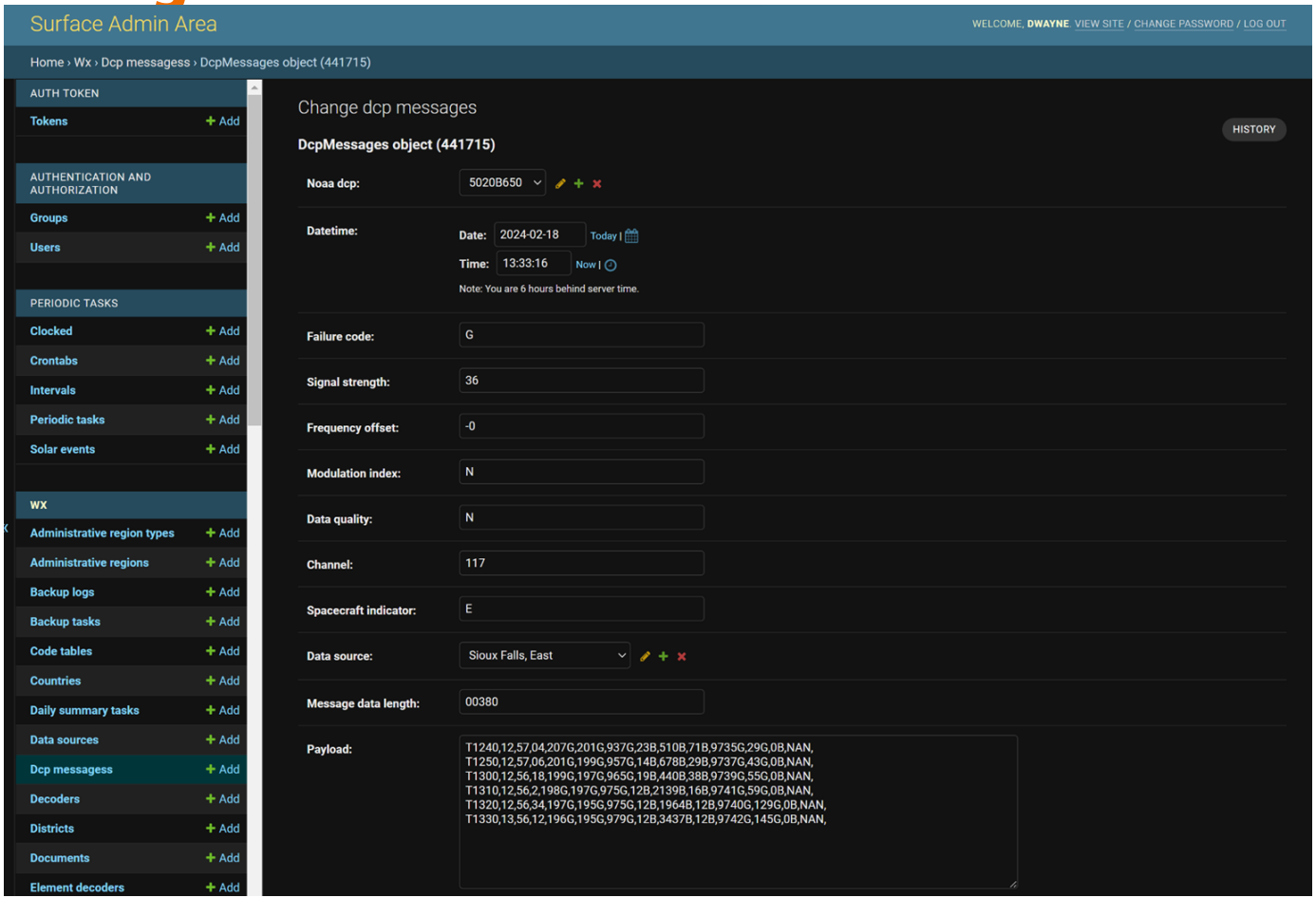


DCP message check…



NOAA DCP	STATION	DATETIME	FREQUENCY OFFSET	FAILURE CODE	DATA QUALITY
5020B650	Camp 6 - F&E Farm - 9924901	Feb. 18, 2024, 1:33 p.m.	-0	G	N
502046D4	Golden Stream - 9922301	Feb. 18, 2024, 1:31 p.m.	-0	G	N
50209E6E	Barton Creek - 9900703	Feb. 18, 2024, 1:26 p.m.	-0	G	N
502005DE	Half Moon Caye - 9958801	Feb. 18, 2024, 1:25 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:22 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:21 p.m.	-0	?	P
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:20 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:19 p.m.	-0	G	N
50204806	Calabash Caye - 9924301	Feb. 18, 2024, 1:19 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:19 p.m.	-0	G	N
50203E96	Mauger Caye - 9925001	Feb. 18, 2024, 1:18 p.m.	+0	G	N
50202DE0	Santa Martha - 9921301	Feb. 18, 2024, 1:18 p.m.	+0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:18 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:17 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:16 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:15 p.m.	-0	G	N
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:13 p.m.	-0	?	P
5020734E	Corazon - 9922201	Feb. 18, 2024, 1:11 p.m.	-0	?	F

DCP message check…



Data Flow

AWS(Data logger) (Stage 1)

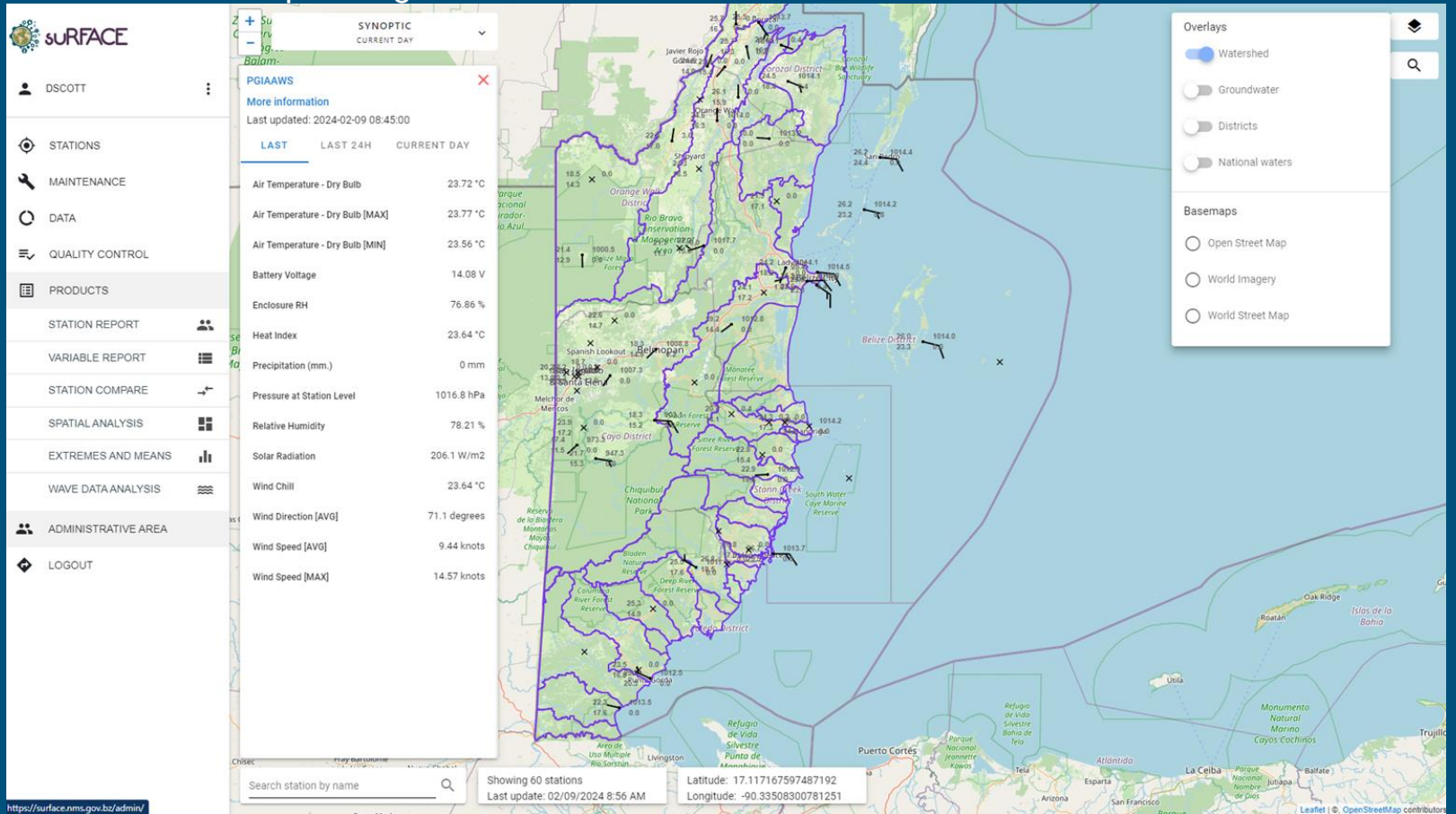
Raw Data Storage (Stage 2)

CDMS - Data Processing (Stage 3)

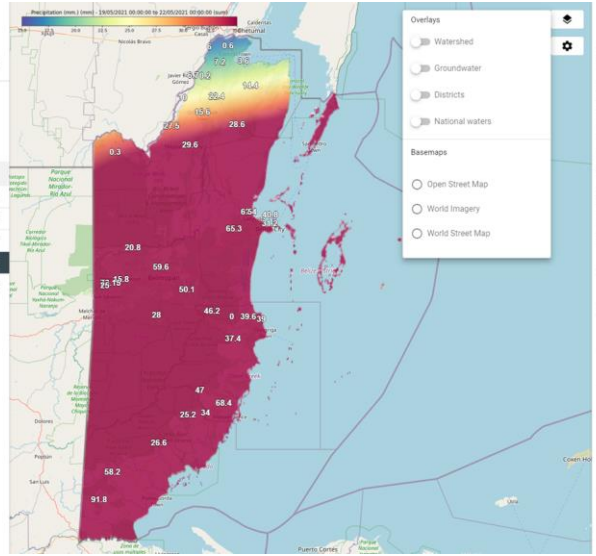
Data Output - Display/Access/Transmission(Stage 4)

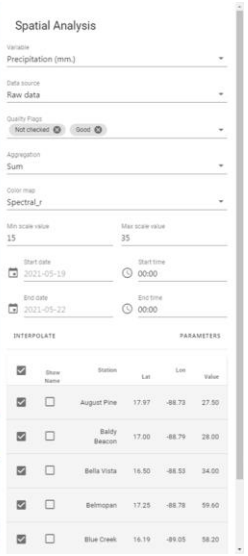
Stage 4 - Data output

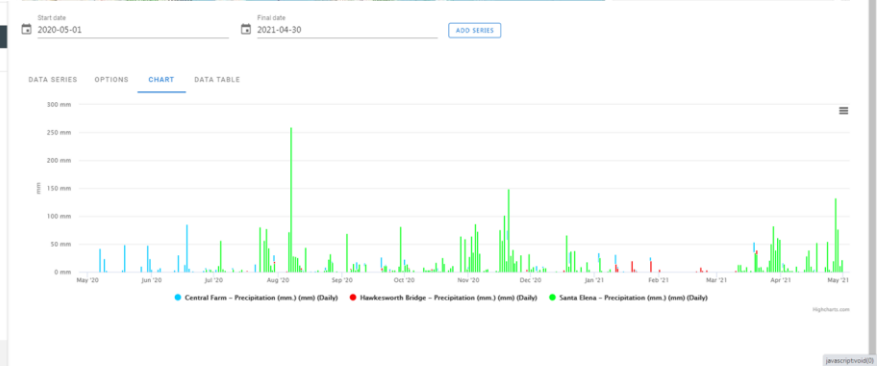
CDMS - Stakeholders can request login access

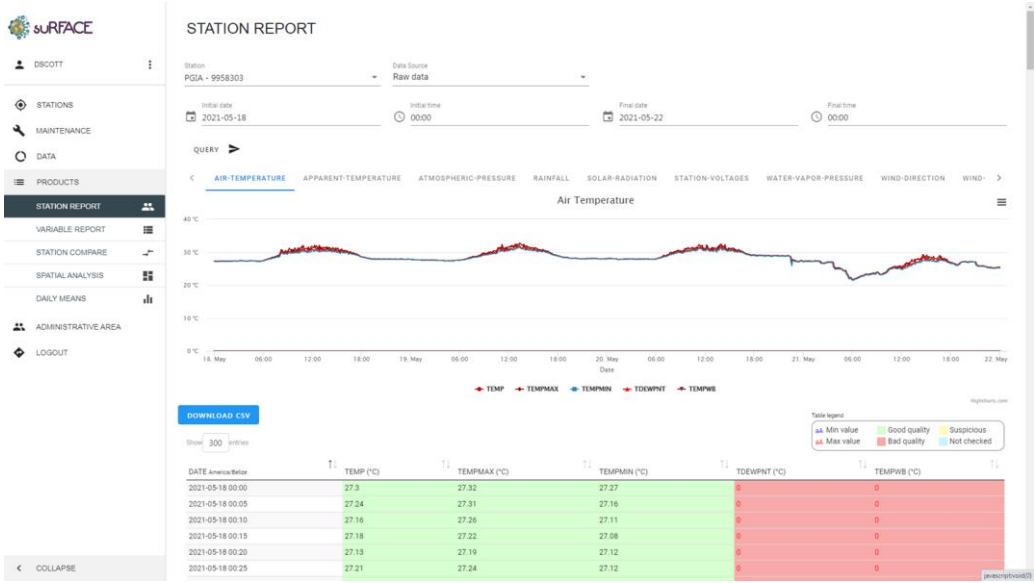


Various reports and outputs









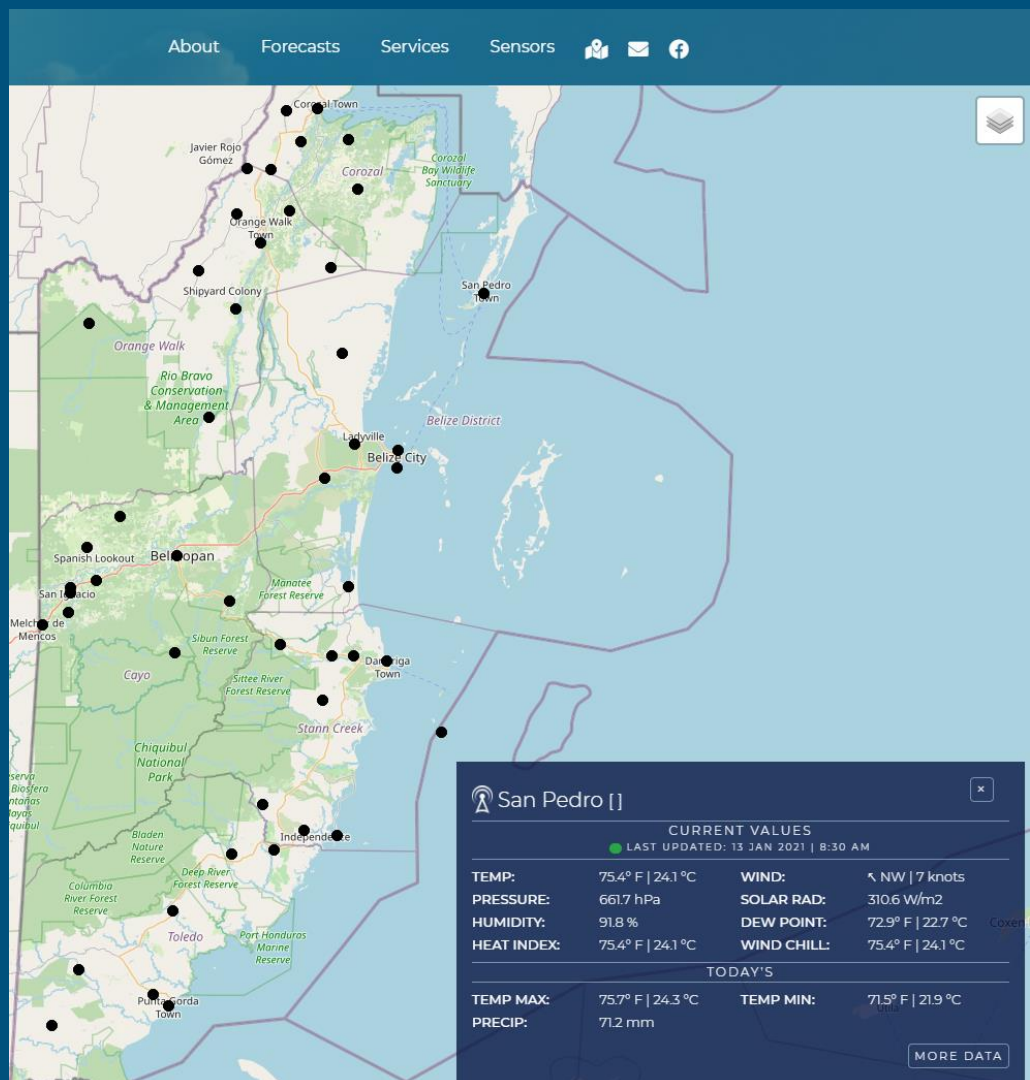
Stage 4 - Data output

CDMS

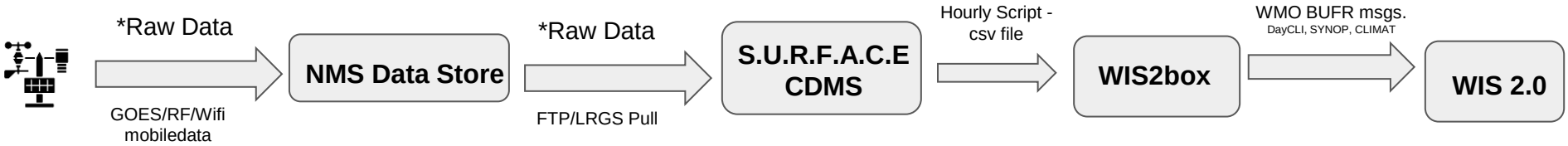
- Stakeholders can request login access

Django REST API - (select endpoints available)

- Facilitates automated requests for data
  - C.I.M.H.
  - San Ignacio Early Warning/Taiwan
  - C.A.F.F.G. - Central American Flash Flood and Guidance
  - CCRIF/CATIE - Rainfall analysis
  - NMS WEBSITE
- WIS2box
  - Hourly Script queries database to generate csv file



Belize WIS 2.0 Implementation



Challenges and Recommendations

WMO members do not have access to PRACTICAL, TECHNICAL ASSISTANCE, GUIDANCE and RECOMMENDATIONS.

E.g. wis2box

Stage 1(AWS) - Challenges/Recommendations

Currently there is no recommended standard for data logger operation - While there are multiple manufacturers, users and stakeholders of weather data; for the exchange of weather data among WMO members there should be a recommended standard for data logger operation.

a. Standard Sample rate for each variable
b. Standard computations done on each variable
c. Standard QC checks done on data logger
d. Standard Storage intervals
e. Standard way to check if recommended standards are being met
f. ALL THIS IS MANUFACTURER/DEVICE AGNOSTIC

Manufactures should meet these basic standards by either shipping data loggers pre-programmed or providing end users with the program - e.g. Onset(Hobo), Sutron, MicroCom, Vaisala, Campbell all these data loggers are able to do these things. Each manufacturer does it differently, but the raw data stored as a data file would have been created the same way!

E.g. NMSB stores multiple storage intervals and computations for various stakeholders but we have a standard for our data collection and storage

- 1 min wind data, 1 hour data for general public and stakeholders - only raw files not CDMS
- However we do have our standard sample rate, computation and storage intervals for CDMS ingestion

Stage 2 - Challenges/Recommendations
Data Reception and Storage

There are many technological ways to handle data reception and storage.
It would be nice to encourage more established services to help smaller NMHS's who have not really thought through these issues.

For example for us in Belize…

- As we face a problem we try to find a solution. Are we following best practices?
- Is there a more efficient/secure method?
- What are other NMHSs doing?
- How can we share experiences, help and learn each other?
- What is the most cost effective tech stack to handle this kind of workload
- Do AWS manufacturers provide a way for data files to be sent to NMHSs?
    - Free, 1 time payment, continual subscription
    - Proprietary or non-Proprietary
    - What Protocols are used for data transfers?
- Do NMHS's know how to ingest AWS satellite data via NOAA LRGS or NOAA DADDS

Stage 3 CDMS - Challenges/Recommendations

1. NMHSs should have a CDMS where AWS raw data files are ingested and where additional quality control checks can be done before any distribution of the data.

1. The CDMS should be the only source of truth, from which all other data flows and is accessed.
    - WIS2box
    - Hourly messages
    - Daycli
    - CLIMAT messages
    - Stakeholders
    - Websites
    - Etc

1. The CDMS should:
    - Be a modern web application with standard API
    - Be able to Ingest from multiple data sources
    - Conduct Automatic QC procedures
    - Allow for user friendly setup and configuration
    - Should be able to generate the necessary WMO messages from stored data and have it sent on GTS/WIS 2.0

Stage 4 Data Access - Challenges/Recommendations

1. WMO messages should be generated from the CDMS
2. CDMS should be able to push data unto WIS 2.0 or wis2box
3. Documented API for data sharing and access
4. Data sharing policies should be setup by NMHSs

SURFACE CDMS USER AGREEMENT

Please read the following before being granted access to SURFACE and its API

1. Access to the SURFACE CDMS's dataset is a courtesy being extended to stakeholders of the National Meteorological Service of Belize (NMS) who make an official request to the Agro-Climatic Section of the NMS for historic climatic In-Situ weather station data.
2. Any user granted access to this data source agrees not to share or profit from this dataset. Furthermore users are asked to properly acknowledge and cite the data source if used in publications or reports and to ensure that copies of any such documents are given to the NMS for record keeping. Any breach of this agreement will result in the immediate termination of access and users will be prohibited from gaining future access in any subsequent requests.
3. Please note that the data retrieved from the SURFACE after 2010 **HAS NOT** undergone the NMS's quality control procedures, therefore users of these data sets do so **AT THEIR OWN RISK!**
4. Quality Control Checks (QC) are applied to some AWS raw data variables after 2010. In these cases, any data that has been flagged as erroneous or suspicious by these checks will not be part of the aggregated dataset. Please see page 2 details on rawdata and the generated summaries.
5. ALL requests for official, validated data should be sent by email to the NMS Climate Services at agroclimat@nms.gov.bz

I (Name of User) _____ using (email address)_____

from(Name of Organization)_____hereby agree to the NMS SURFACE CDMS user agreement as outlined above.

Signature:_____ Date:_____

SURFACE API user feedback…

"…overall, we were very pleased with the ease of accessing the API and working with the data!

Not only was the provided documentation very clear and informative, we found the available functionality for accessing the data to be intuitive and appropriate for our operational application. Additionally, the information/data available through the API was both well-defined and easily understandable. All of these factors led to a very clear path forward for our operational intent…"

Central America Flash Flood Guidance System(C.A.F.F.G.)
Hydrologic Research Center, San Diego, CA

Questions?