Users Committee Virtual Meeting Agenda

Friday, May 7th

(Times are Mountain Daylight Time)

10:00 - 10:05 Convene and outstanding items

10:05 - 10:25 Status Reports (Committee)

10:25 - 10:30 Triennial Planning Committee (Josh Young)

10:30 - 10:45 AWIPS Update (Shay Carter/Tiffany Meyer)

10:45 - 11:00 AWIPS asynchronous training (Nicole Corbin/Committee)

11:00 - 11:10 Users Committee Community Survey (Josh Young/Committee)

11:10 - 11:15 Unidata DEI Committee (DEI Committee Members)

11:15 - 11:30 MSI Engagement Update (Jeff Weber)

11:40 - 11:55 IDV Update (Yuan Ho)

11:55 - 12:00 Close Meeting

Status Report: AWIPS and GEMPAK

October 2020 - April 2021

Tiffany Meyer, Shay Carter

Areas for Committee Feedback

We are requesting your feedback in a google form that will be sent out.

Activities Since the Last Status Report

AWIPS

Unidata's Jetstream production EDEX server is the primary host for serving real-time weather and geographic data to CAVE clients and through the python-awips data access framework (API). By offloading the processing of one very large data feed (NEXRAD3) to a separate EDEX Ingest Node, the current implementation of edex-cloud is now capable of processing and serving more data than ever before. The distributed architectural concepts of AWIPS allow us to scale EDEX in the cloud to account for the size of incoming data feeds. With this distributed architecture, we were able to assist Texas A&M in setting up their very own EDEX server. The server they built is actually larger and consists of more machines than the current Unidata one. We helped them configure a main EDEX machine, and two ancillary machines to deal with radar and model data separately. We are also looking into adding an additional machine for our own EDEX system, to help handle large volumes of requests (for instance, during severe weather, or at the start time of a classroom lab when students are connecting simultaneously to EDEX).

Since the last update, several EDEX changes have been tested on our backup server. We have been testing and acquiring several new datasets including GOES RGB products produced from CIRA, additional GOES derived wind and products from NOAAPort, GOES E GLM products produced by Eric Bruning, impact based warnings, and additional MRMS data products from NCO. With the addition of all the new GOES products, the Satellite menu has been revamped for easy access to new products and removal of entries with no data. All of these new datasets are also available via Unidata's IDD.

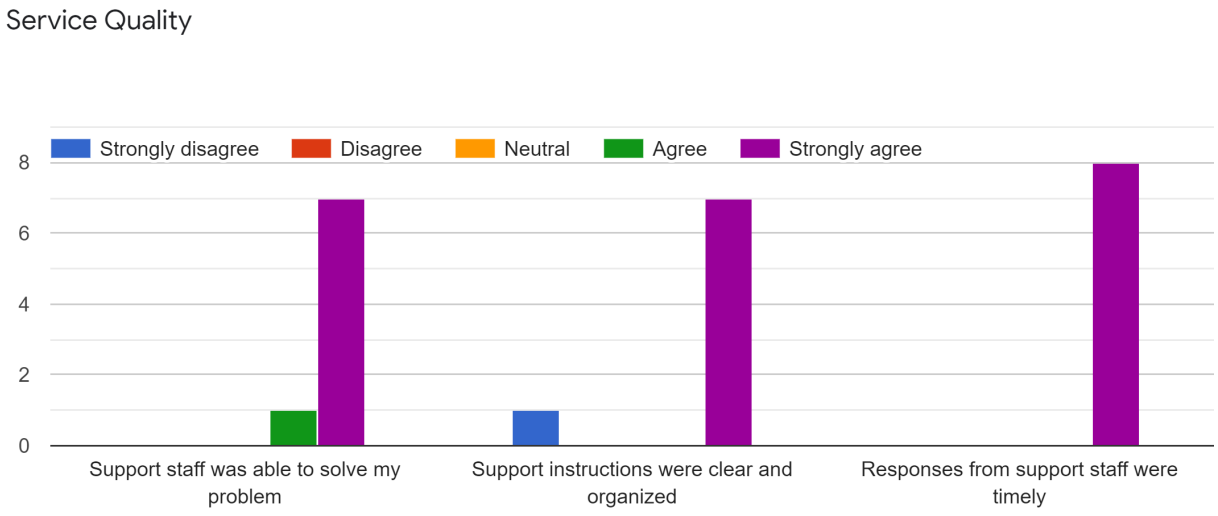
We are in the process of developing and releasing a new version of AWIPS in which we've migrated some of the configuration from CAVE to EDEX. This means we will have more flexibility to update menus for easier product access and understandability, without requiring users to download and install a new version of CAVE. With the ever changing landscape of data and data products, this switch makes it so Unidata's AWIPS can be more responsive and

reflective of what current data is available, and does so with no effort on the users' behalf. All the new data products we've been testing will also be available to the public with this release.

In order to support CAVE for Mac installation, we are in the process of officially signing and notarizing all of our packages through Apple. This will allow users to easily install CAVE without having to go through any system overrides or requiring administrative privileges. Additionally, with this new release of CAVE we have added previously inoperable functionality for Mac with the ability to capture and save the CAVE display. This functionality has always been available for our Linux and Windows distribution, but not for the Mac.

A significant portion of our documentation both for CAVE and EDEX, and for python-awips has been modified for easier understanding and comprehension. With the help of our new Educational Designer, we created and released a detailed video for the installation of CAVE (and all it's necessary prerequisite software) on Windows machines. These changes were partially fueled by our new blog series: AWIPS Tips. The first week of this month (April 2021), we officially released our first blog. The series is intended to highlight different functionality and capabilities of all three parts of AWIPS: CAVE, EDEX, and python-awips. The blogs will be released on a bi-weekly basis, and will be announced on our mailing list (awips2-users@unidata.ucar.edu) as well as our social media accounts (Facebook, Twitter, and LinkedIn).

In the past several months the AWIPS team created a new support evaluation survey that is advertised in our support email signatures. We have averaged about one response per month since the form went live and have received overwhelming positive feedback. The graphic below is from the results of the multiple choice questions we ask:



On a separate, but related topic, we have transitioned our Elasticsearch/Kibana functionality from an Atmosphere Jetstream instance to an OpenStack Jetstream instance. This has allowed our IT department to have more control over access and security of the machine. We have updated the versions of that technology stack and are still in the process of developing the system further to help support our EDEX machines.

The annual AMS conference in January was hosted virtually. We presented on Unidata's AWIPS current status and a look toward the future. The abstract and recording are available online.

Software Releases

We are nearing the point of putting out a new AWIPS release. This would be labeled version 18.2.1, and would include several EDEX configuration changes, CAVE updates, and access to new datasets as well. We have been able to successfully build CAVE installers for all platforms, although we are in the process of officially notarizing and certifying the Mac installer. In the more extended future, we aim to merge our Unidata AWIPS back with one of Raytheon's more recent builds (most likely version 20.3) which includes transitioning from Python 2 to 3 and Java 8 to openJDK 11.

GEMPAK/NAWIPS

At this time, GEMPAK development at the Unidata Program Center has ceased, and support through Unidata is very limited. In March 2021, we released a statement about the state of GEMPAK and the Program Center's hopes of handing off ongoing maintenance of the package to a community-based team. Response to this statement was favorable, and Daryl Herzmann of Iowa State University agreed to serve as the initial point of contact for those interested in contributing to this project. In order to support a smooth transition to community-based maintenance, we are taking the following steps:

- Providing commit access to Unidata's GEMPAK repository on Github to a new community-based maintenance team.
- Working with the community-based team to determine where GEMPAK documentation should be hosted. Documentation will remain available on Unidata's web site until a suitable replacement is available.
- Working with the community-based team to determine the future role of the gembud mailing list. The mailing list will remain in operation under Unidata's auspices until a suitable replacement is available.

We envision the community-based GEMPAK open source effort as a self-sustaining program, with Unidata acting as a community member rather than as the software's official maintainer.

Activities Ongoing/In-Progress

AWIPS development activities are constantly developing. Currently the following activities are in progress:

- The AWIPS team is expanding the Elasticsearch and Kibana capabilities to build a thorough reliable metrics and monitoring tool.
- The AWIPS team has been testing and implementing changes to improve and optimize our cloud EDEX server.
- The AWIPS team is exploring the possibility of adding additional data and increasing the archive time of some existing data.
- The AWIPS team is responding to all AWIPS support questions from the community and striving to provide realistic solutions in a timely manner.
- The AWIPS team is creating new branches for all our GitHub repos for our upcoming 18.2.1 release.
- The AWIPS team is in the process of officially certifying and notarizing our CAVE installation for the Mac through Apple's new required processes.
- The AWIPS team has worked through the build process for the Linux and Windows distributions of CAVE and is ready to make deployments as soon as the code is ready.
- The AWIPS team has started a new bi-weekly blog series called AWIPS Tips that began on April 7th, and is intended to highlight useful functionality and fundamentals for CAVE, EDEX and python-awips.

Future Activities

Future plans are currently in a state of flux. The AWIPS team is in the process of creating new branches for the OSX specific code base for CAVE for our upcoming 18.2.1 release. As soon as the code is in a finalized state for all three operating systems, we plan on making a new official AWIPS release. After that, the goal is to possibly look at developing a remote training workshop for either EDEX or CAVE, and then begin transitioning Unidata's AWIPS distribution away from version 18, onto version 20.

Metrics

Downloads August 1, 2020 - March 31, 2021

AWIPS downloads: 3,381
GEMPAK downloads: 954

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   The cloud-based EDEX data server continues to see widespread use and growing adoption. More and more datasets continue to be added to the server as Unidata deploys more decode/ingest nodes.

2. **Providing Useful Tools**
   All AWIPS tools (EDEX, CAVE, and python-awips) are freely available, and also incorporate LDM/IDD technology for accessing geoscience data.

3. **Supporting People**
   At this juncture, we are providing full technical support with regards to AWIPS for the community, and encouraging community members to assist each other through Unidata-managed forums for GEMPAK support.

Prepared *April 2021*

Status Report: Cloud Computing Activities

October 2020 - April 2021

Sean Arms, Shay Carter, Julien Chastang, Ward Fisher, Ryan May, Tiffany Meyer, Jen Oxelson, Mohan Ramamurthy, Jeff Weber, Tom Yoksas

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Has the COVID-19 pandemic and the shift to online learning increased your reliance on cloud computing technologies? How can Unidata better assist you in this area? What is missing in your toolkit to teach effectively during the pandemic?
2. Do you need a Unidata hosted JupyterHub for your classroom or workshop?
3. What new cloud technologies are our community members using and investigating on their own initiatives?
4. What cloud computing environments or platforms are our community members using? Commercial? E.g., Amazon. NSF? E.g., Jetstream.

Activities Since the Last Status Report

NSF Jetstream Cloud Grant Renewal for 2021-2022

Our NSF Jetstream Cloud allocation must be renewed annually with NSF XSEDE (Extreme Science and Engineering Discovery Environment) to ensure continuity of Unidata services (EDEX, THREDDS, JupyterHub servers, etc.) running on Jetstream. We submitted a “Research” application grant renewal for Jetstream resources for 2021-2022. We requested the same amount of support and compute resources as last year:

- ~4,000,000 Service Units (SUs) to run virtual machines on Jetstream.
- 40TB of disk storage.
- XSEDE, Extended Collaborative Support Service (ECSS) to continue JupyterHub cluster support with Andrea Zonca (San Diego Supercomputing Center).
- SGCI (Science Gateway Community Institute) Support.

JupyterHub Servers for Online Instruction During COVID-19 Crisis and 2021 Spring Semester

Unidata JupyterHub activities continue to advance since the last status report. These JupyterHubs are deployed In collaboration with XSEDE, ECSS and the Jetstream group at Indiana University (IU).

We have supported a number of semester-long classes, and workshops with JupyterHub

servers hosted on the Unidata Science Gateway. The JupyterHub servers are tailored to the instructor's objectives with pre-configured PyAOS (Python for the Atmospheric and Oceanic Sciences) environments, classroom material and data. Demand for Unidata JupyterHub servers has increased since the arrival of the COVID-19 pandemic and the transition to online learning. We are more than happy to assist instructors in this area, and would like to help in whatever way we can with these resources.

For the spring semester of 2021, Unidata assisted (or will help) seven universities and one institute with pre-configured PyAOS JupyterHub servers. Four of those servers are "carryovers" from the fall semester for instructors that wished to continue on with a previous server rather than creating a new server. We also help two new programs that we have not seen previously: University of Louisville (with professor Jason Naylor), and Pennsylvania State University (with professor Kevin Bowley). We have helped about 150 students this semester (see table below in the metrics section).

JupyterHub for Student Python Workshop at 2021 American Meteorological Society Virtual Meeting

Continuing from last year's successful Python workshop, Unidata hosted a Python workshop at the Annual Student Conference for the American Meteorological Society 2021 virtual meeting. The goal of this instructional event was to deliver an introduction to Python for the atmospheric sciences. Unidata presented the educational material in a synchronous and asynchronous format. During the conference, students accessed the dedicated, pre-configured JupyterHub to arrive at their instructional material. About 60 students participated in the workshop.

Disk Volume Problems Associated with JupyterHub Servers

Kubernetes JupyterHub servers operating on Jetstream occasionally experience disk volume problems where a student cannot login to their classroom JupyterHub server. Unidata worked with Kevin Goebbert to diagnose the problem, and Jeremy Fischer and Andrea Zonca to attempt to resolve it. Despite spending a fair amount of time attempting to solve the issue, we have never been able to arrive at the root cause. However, we have taken a number of mitigating steps such as improved monitoring and notification of volume problems to quickly implement workarounds. To date, we have never lost a student's data and we can typically deploy fixes within minutes or hours so that students can login once again.

Ongoing Activities

NOAA Big Data Program

- Unidata continues to manage the NEXRAD archive in Amazon S3, ensuring that realtime data are successfully delivered to the noaa-nexrad-level2 bucket. LDM, THREDDS Data Server (TDS), and THREDDS Docker software are employed to deliver these data.

- TDS on Jetstream for level II NEXRAD: http://thredds-aws.unidata.ucar.edu/thredds/catalog.html
- AWS Explorer (Public access): https://s3.amazonaws.com/noaa-nexrad-level2/index.html
- Public Bucket for level II NEXRAD: https://noaa-nexrad-level2.s3.amazonaws.com
- Continue to populate the NEXRAD level II archive with real time data.
- Continue to populate new GFS .25 degree output and NCEP HRRR output to an S3 bucket for access. We did not place a TDS on this collection as this output is available from our standard sources.

JupyterHub Demonstration Server

Unidata continues to enhance the Unidata JupyterHub demonstration server. This server needs to be regularly updated as the Jupyter, JupyterHub, and JupyterLab ecosystems rapidly evolve.

Docker Containerization of Unidata Technology

We continue to employ Docker container technology to streamline building, deploying, and running Unidata technology offerings in cloud-based environments. Specifically, we are refining and improving Docker images for the LDM, ADDE, RAMADDA, THREDDS, and AWIPS. In addition, we also maintain a security-hardened Unidata Tomcat container inherited by the RAMADDA and THREDDS containers. Independently, this Tomcat container has gained use in the geoscience community.

Progress has been made on the following

- Tomcat Docker container continues to be updated as new versions of Apache Tomcat become available. We try to do this quickly in the event of a Tomcat security update.
- We work closely with Jen Oxelson and the systems admin group to ensure that our containers are configured as securely as possible.
- Collaborated with Sean Arms for the deployment of TDS production and beta releases as new versions of the TDS are released.

Product Generation for IDD

For the past five years, Unidata generated products for the IDD, FNEXRAD and UNIWISC data streams have been created by a VM hosted in the Amazon cloud. This production generation has been proceeding very smoothly with almost no intervention from Unidata staff.

AWIPS EDEX in Jetstream Cloud

Unidata continues to provide an EDEX data server on the Jetstream cloud, serving real-time AWIPS data to CAVE clients and through the python-awips data access framework (DAF) API. The distributed architectural concepts of AWIPS allow us to scale EDEX in the cloud to account for the desired data feed (and size). We continue using Jetstream to develop

cloud-deployable AWIPS instances, both as imaged virtual machines (VMI) available to users of Atmosphere and OpenStack, and as Docker containers available on DockerHub and deployable with the science gateway toolset.

EDEX is designed with a distributed architecture, so different components can be run across separate virtual machines (VMs) if needed, to improve efficiency. Our current design makes use of two VMs: one large instance to process most of the data and run all of the EDEX services, and another smaller instance to handle radar data specifically.

For the past year, we have successfully created and maintained a duplicate set of VMs to mirror our production EDEX environment. These auxiliary VMs have served as a testing ground for implementing new changes, as well as a backup for when our production server is unavailable.

This coming year, we plan on potentially adding one more VM to our EDEX environment to help handle the large influx of CAVE connections we periodically encounter (for a total of two new VMs -- one for each of the production and backup systems). This scenario can happen in severe weather conditions, or when numerous students in a university class connect to our server simultaneously. Additionally, we have talked about implementing a totally independent case-study server for our community.

Beyond these EDEX servers, we also have one additional VM in Jetstream related to AWIPS. We have launched a machine to host and serve an Elasticsearch database, with a Kibana web interface. The purpose of this VM is to gather real-time metrics on all four of our EDEX machines and provide insight to potential problems.

Nexrad AWS THREDDS Server on Jetstream Cloud

As part of the NOAA Big Data Project, Unidata maintains a THREDDS data server on the Jetstream cloud serving Nexrad data from Amazon S3. This TDS server leverages Internet 2 high bandwidth capability for serving the radar data from Amazon S3 data holdings.

Jetstream Security

We work with the Unidata system administrator staff to ensure that our web-facing technologies and virtual machines on Jetstream adhere to the latest security standards. This effort involves such tasks as ensuring we are employing HTTPS , keeping cipher lists current, ensuring docker containers are up-to-date, limiting ssh access to systems, etc.

Unidata Science Gateway Website and GitHub Repository

Website

The Unidata Science Gateway web site is regularly updated to reflect the progress of what is available on the gateway. The news section is refreshed from time-to-time for announcements concerning the gateway. The conference section and bibliography is also

maintained with new information.

Repository

All technical information on deploying and running Unidata Science Gateway technologies is documented in the repository README. This document is constantly updated to reflect the current state of the gateway.

Presentations/Publications

- J. Chastang, D. J. Delene, W. J. Flynn, K. H. Goebbert, and J. T. Potemra. Assisting university earth science programs transition to online learning with computational notebooks. In *30th Conference on Education at the 101st AMS annual meeting*, New Orleans, Louisiana, USA, Jan. 12-16 2021.
- UCAR President's Council Presentation on Unidata Science Gateway
- S. Arms and J Chastang. Python workshop. In *101st AMS Annual Meeting*, New Orleans, Louisiana, USA, Jan. 12-16 2021
- Douglas Dirks. UCAR Mentorship and Remote Computing Resources Boost Remote Research for SOARS Student, Oct 2020.
- A. Zonca, R. P. Signell, J. Chastang, J. Fischer, J. M. Lowe, and R. S. Sinkovits. *Deploy kubernetes and jupyterhub on xsede jetstream*. In Gateways 2020 Proceedings, Oct. 19-23 2020. See also https://science-gateway.unidata.ucar.edu.

New Activities

Over the next three months, we plan to organize or take part in the following:

Forthcoming Virtual Conference Attendance

- 17th IEEE eScience Conference, September 20-23, 2021

AWS Public Dataset For NEXRAD Level 3

We were approached by AWS's public dataset program about NEXRAD Level 3 data. We have started uploading all NEXRAD level 3 data to the unidata-nexrad-level3 S3 bucket; this bucket is intended to store all data indefinitely, providing a full archive of data going forward.

Over the next twelve months, we plan to organize or take part in the following:

XSEDE ECSS JupyterHub Collaboration

We plan to continue our collaboration with Andrea Zonca (XSEDE ECSS, San Diego Supercomputing Center). We are currently testing out a new version of the Kubespray (version `2.15`) project for employing JupyterHub clusters.

Jetstream2

As described in the fall 2020 Cloud Computing Activities status report, the NSF Jetstream2 project, a follow-on to the current Jetstream facility, has been awarded funding. We plan on applying for resources to that system via the XSEDE allocation process possibly by an initial "Startup" grant followed by a much larger "Research" allocation. We plan to start exploring that system and its capabilities in the next year. Eventually, we plan on migrating Unidata operations currently running on Jetstream to Jetstream2. We would also like to experiment with object store technologies that may be available on Jetstream2.

Relevant Metrics

Spring 2021 JupyterHub Servers

	Nº of Users	Point of Contact	Notes
Valparaiso University	40	Prof. Kevin Goebbert, Department of Geography and Meteorology	
U of North Dakota	17	Dr. David Delene Prof Dept of Atmos Sciences U of North Dakota	Will happen later in Spring 2021
AMS 2021 Python Workshop	82	Sean Arms, Unidata	21 Admins
University of Louisville	17	Professor Jason Naylor	
Penn State University	7	Professor Kevin Bowley	They are using the main Unidata JupyterHub demonstration server.
University of Oklahoma	8	Shawn Riley, Ben Shenkel OU School Meteorology	JupyterHub continued from the fall 2020
U of North Dakota	5	Dr. Aaron Kennedy Assoc Prof Dept of Atmos Sciences U of North Dakota	JupyterHub continued from the fall 2020
Southern Arkansas University	1	Keith Maull (UCAR/NCAR Library)	JupyterHub continued from the fall 2020

Github Statistics

	Watches	Stars	Forks	Open Issues	Closed Issues	Open PRs	Closed PRs
sience-gateway	4	14	8	10	149	0	441
tomcat-docker	8	43	46	2	35	0	62
thredds-docker	10	18	18	3	104	0	142
ramadda-docker	2	0	1	1	10	0	19
ldm-docker	6	11	10	0	33	0	52
tdm-docker	3	2	6	1	9	0	13

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   Unidata supplies a good portion of the data available on the IDD network to the Jetstream cloud via the LDM and the high bandwidth Internet 2 network. Those data are distributed to the TDS, ADDE, RAMADDA and AWIPS EDEX installations running on Jetstream for the benefit of the Unidata community. Unidata also makes the AWS Nexrad archive data accessible through the TDS Nexrad server running on Jetstream at no cost to the community. These data can be accessed in a data-proximate manner with a JupyterHub running on Jetstream for analysis and visualization. Containerization technology complements and enhances Unidata data server offerings such as the TDS and ADDE. Unidata experts install, configure and in some cases, security harden Unidata software in containers defined by Dockerfiles. In turn, these containers can be easily deployed on cloud computing VMs by Unidata staff or community members that may have access to cloud-computing resources.

2. **Providing Useful Tools**

Jupyter notebooks excel at interactive, exploratory scientific programming for researchers and their students. With their mixture of prose, equations, diagrams and interactive code examples, Jupyter notebooks are particularly effective in educational settings and for expository objectives. Their use is prevalent in many scientific disciplines including atmospheric science. JupyterHub enables specialists to deploy pre-configured Jupyter notebook servers typically in cloud computing environments. With JupyterHub, users login to arrive at their own notebook workspace where they can experiment and explore preloaded scientific notebooks or create new notebooks. The advantages of deploying a JupyterHub for the Unidata community are numerous. Users can develop and run their analysis and visualization codes proximate to large data holdings which may be difficult and expensive to download. Moreover, JupyterHub prevents users from having to download and install complex software environments that can be onerous to configure properly. They can be pre-populated with notebook projects and the environments required to run them. These notebooks can be used for teaching or as templates for research and experimentation. In addition, a JupyterHub can be provisioned with computational resources not found in a desktop computing setting and leverage high speed networks for processing large datasets. JupyterHub servers can be accessed from any web browser-enabled device like laptops and tablets. In sum, they improve "time to science" by removing the complexity and tedium required to access and run a scientific programming environment.

3. **Supporting People**
   A Unidata science gateway running in a cloud computing setting aims to assist the Unidata community arrive at scientific and teaching objectives quickly by supplying users with pre-configured computing environments and helping users avoid the complexities and tedium of managing scientific software. Science gateway offerings such as web -based Jupyter notebooks connected with co-located large data collections are particularly effective in workshop and classroom settings where students have sophisticated scientific computing environments available for immediate use. In the containerization arena, Unidata staff can quickly deploy Unidata technologies such as the THREDDS data server to support specific research projects for community members.

Prepared *April 2021*

Status Report: Community Services

October 2020 - April 2021

Nicole Corbin, Doug Dirks, Jeff Weber, Joshua Young

Areas for Committee Feedback

We are requesting your feedback on the following topics:

Do your needs from the Unidata Program Center change during this unique time?

Activities Since the Last Status Report

News@Unidata blog

Posts to the News@Unidata blog appear regularly, but not on a specific schedule. Some highlights:

- NSF EarthCube funds Project Pythia
- UCAR Mentorship and Remote Computing Resources Boost Remote Research for SOARS Student
- AMS Student Conference Python Workshop
- Unidata Staff at AMS 2021 Meeting
- Offer: Unidata Science Gateway JupyterHub Resources Available for Spring 2021 Courses
- Call for Proposals: Unidata 2021 Community Equipment Awards
- What's New in MetPy 1.0?
- Wanted: Student Representative for Unidata Users Committee
- Mike Zuranski Receives 2020 DeSouza Award
- Unidata Program Center Welcomes Nicole Corbin
- Unidata Staff Teach MetPy at AMS *Python for Climate and Meteorology* Short Course
- A Proposal for Community Support of GEMPAK
- Welcome to AWIPS Tips!
- Unidata Welcomes New Committee Member
- Software release information
- Community job postings
- Community meetings and other announcements

Dependencies, challenges, problems, and risks include:

- Finding community members willing to contribute stories (or story ideas) for the blog is an ongoing challenge. We're starting to make progress working with committee members, but there is more to do.

Community Outreach and Services

The community services group continues to actively reach out to and engage with Unidata community members.

Progress has been made on the following:

- Continue to engage with underserved populations and institutions as part of Unidata's outreach efforts to groups such as Rising Voices, SACNAS and AISES
- Continue to serve on the CUAHSI HIS standing committee.
- Engage with the Arctic Research Consortium of the US on multidisciplinary projects
- We continue to update Unidata's social media channels (Facebook, Twitter, Google+).
- We continue to publish short videos/screencasts on the Unidata YouTube channel.
- Represent Unidata at the National Weather Service Partners events
- We continue to actively support the NCAR/SOARS program.
- Actively participate in Super Science Saturday.
- Engage and support the Undergraduate Leadership Workshop (ULW) at UCAR.
- Support the development and operation of the UCAR:NCAR Equity and InclusiON (UNEION) community of practice.

Dependencies, challenges, problems, and risks include:

- Facilitating community adoption of new technological services (cloud, etc)
- Engagement with Unidata social media streams among community members is not particularly high.

Learning and Outreach

The community services group has expanded efforts to promote learning Unidata products and workflows.

Progress has been made on the following:

- Delivering a virtual MetPy workshop with activities backed by adult learning science principles.
- Developing an appropriately scaffolded Python curriculum from initial programming/logical thinking principles to Unidata packages (e.g. MetPy, Siphon, Python-AWIPS)
- Performing a needs assessment for an upcoming AWIPS eLearning offering
- Sharing tips and best practices for using AWIPS through the AWIPS Tips blog series

Dependencies, challenges, problems, and risks include:

- Identifying community members with the availability to provide feedback throughout the development process for new learning materials. This feedback is critical for ensuring learning programs will directly meet the needs of the community.

Ongoing Activities

We plan to continue the following activities:

- Engagement with science or cyber communities at large
- NAWIPS migration to AWIPS, including the overall AWIPS project
- Ongoing development of news articles for publication through News@Unidata
- Continue to support and contribute to governing committees
- Seminars
- Outreach
- Inclusion and equity
- Engagement with professional societies
- Support for cloud-related projects
- Further development of the Data Management Resource Center
- Support the pursuit of funding
- Site visits as the budget and pandemic allow
- Engage other UCAR/NCAR divisions regarding Unidata software use i.e. CESM/IDV
- Active participation in CUAHSI HIS (Hydrologic Information System)
- Active participation in the Hydroshare Advisory Committee

New Activities

Over the next three months, we plan to organize or take part in the following:

- Expanded emphasis on engagement with MSIs
- Expanded effort organizing and supporting community seminars/working sessions

Over the next twelve months, we plan to organize or take part in the following:

- Make structural changes to broaden participation in Unidata community engagement
- Continue to engage the hydrologic community regarding WRF-Hydro/IDV interactions and the National Water Center's efforts
- Continue to engage the arctic community to find opportunities for collaboration
- Seek additional opportunities to engage and listen to the community
- Release a self-paced eLearning module for AWIPS
- Formalize a curriculum for learning Unidata-provided Python packages

Beyond a one-year timeframe, we plan to organize or take part in the following:

- Support the providing additional cloud-related training

Relevant Metrics

Statistics from the Community pages on the Unidata web site. Comparisons are made with statistics from the previous six-month period.

All community pages

Most recent six months:

- 46,703 unique pageviews (50,067 in previous period)
- 8.1% of total unique pageviews (7.6% in previous period)

Top community pages

1. All blog pages
   32397 unique pageviews (38443 in previous period)
   69% of total community pageviews (77% in previous period)
2. www.unidata.ucar.edu/community
   8592 unique pageviews (4682 in previous period)
   18% of total community pageviews (9% in previous period)
3. www.unidata.ucar.edu/events
   1927 unique pageviews (3438 in previous period)
   4% of total community pageviews (7% in previous period)
4. www.unidata.ucar.edu/about
   3010 unique pageviews (2992 in previous period)
   7% of total community pageviews (6% in previous period)

Social media statistics, April 16, 2021

1. # of Twitter followers: 1597 (up from 1386)
2. # of Facebook followers: 874 (up from 846)

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   We monitor and collaborate with data sources to stay apprised of impending changes and to advocate for the needs of our user community.

2. **Supporting People**
   We provide user workshops, tutorials, and community workshops to help build supportive relationships between community members.

We coordinate with our governing committees to find ways to expand Unidata's community participation. We use our web site, electronic newsletters, and social media to keep community members informed about enhanced data services, software tools, and cyberinfrastructure.

We participate in UCAR/NCAR and NSF projects for underrepresented populations and minority communities (SOARS, AIHEC, outreach to HBCUs). We provide services and tools to facilitate education and research in diverse communities. We work to broaden the Unidata community by participating in student and professional conferences.

Prepared *April 2021*

Status Report: Unidata Community Equipment Awards

Sponsored by the National Science Foundation

October 2020 - April 2021

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Possible theme(s) for the 2022 Unidata Community Equipment Awards;
2. Please consider volunteering to serve on the 2022 Review Panel;
3. Suggestions from previous panel members on how to improve the program

Community Equipment Awards

The NSF provides the Unidata Program Center up to $100k in equipment grant funds each year. In alignment with the Unidata 2024 proposal, the Equipment Awards Program is designed to broaden participation and promote the use of Unidata tools and systems (e.g., THREDDS, NetCDF, IDV, GIS connections) to support education and research on various aspects of climate studies (e.g., diagnostics, change and impacts), by providing grants to be used in the procurement of new computers and equipment including upgrades to existing classroom and laboratory equipment.

A Request for Proposals was sent out on December 1, 2020 with a March 26, 2021 submission deadline. The Review Panel will meet virtually on April 14 to finalize their reviews and to make funding recommendations.

Relevant Metrics

Since taking over the management and administration of the Equipment Awards program in 2003 on behalf of the NSF, Unidata has made 104 awards totaling over $1,300,000.

Status Report: GOES-R and NOAAPort Satellite Ingest

October 2020 - April 2021

Mike Schmidt, Tom Yoksas

Questions for Committee Members

- What image coverages, spatial and temporal resolutions and possibly projections should be considered for addition to the **UNIWISC** IDD feed?

- What kind(s) of data access methods are most desired/usable for the community?

  We currently provide access via the IDD (push), THREDDS Data Server (pull), McIDAS ADDE (pull) and AWIPS EDEX (pull).

- Other questions?

Activities Since the Last Status Report

- During the process of upgrading the OS from CentOS 6 to CentOS 8 on our GRB ingest machines, we encountered a problem where CSPP GEO processes couldn't keep up with GOES-17 ingest activities on one of our machines. The effect of this was the existence of gaps/blank spots in images. These gaps/dropouts were seen to affect processing by end users that were/are creating products from GOES-17 imagery.

  Prompted by a support request (NOAA/Raytheon) to enhance the NOAAPort ingest component of the LDM to be able to set its input buffer size, a test of significantly increasing (from 16 MB to 96 MB) the input UDP buffer size used by GRB processing software, CSPP GEO (UWisconsin/CIMMS), was initiated. Instances of ingest "Gaps" dropped from an average of ~3200/day to less than 15 per day with most days being more like a handful.

  Comment:

  By any measure, GOES-16/17 imagery continues to be **very** popular in the community!

Ongoing Activities

We plan to continue the following activities:

- Participate in UW/SSEC's "fanout server" sharing of GOES-R/S data (redistribution of

the GRB-200 UDP unicast stream over TCP) for GOES-17 GRB products.

We are feeding from SSEC's GOES-16/17 fanout servers, and they are feeding from the ingest machine that we operate. Sharing of the feed streams has allowed by SSEC and Unidata to minimize effects of solar and terrestrial interference.

- Ingest GOES ReBroadcast (GRB) streams from GOES-16 and GOES-17 in real-time

  Since repointing the 4.5m satellite dish at the NCAR Mesa Lab from GOES-16 (75 W to GOES-17 (137.2 W), the Terrestrial Interference (TI) that had been hampering GOES-16 data ingest activities has decreased to the point that our ingest quality rivals what UW/SSEC experiences on their 6.3m dishes.

  In the fall of 2017 we began experiencing significant TI in the GOES-16 signal being received by our 4.5m satellite dish at the NCAR Mesa Lab. An outcome of the discussions we had with Quorum Communications (the manufacturer of the electronics we use in our GOES-R/S ingest installations) was our moving of the GOES-16 ingest to a 3.8 m satellite dish located at the UCAR FL-2 location. The relocation of GOES-16 ingest required that an additional signal cable be pulled from the satellite dish that was repurposed from GOES GVAR ingest into the 2nd floor FL-2 NCAR/RAL computer room where our ingest electronics are located. The cost of this work was contributed by the UCAR/NCAR networking group.

  GOES-16 ingest on the Foothills Lab 2 (FL-2) satellite dish has been working well except during periods where maintenance is being done in the tree-lined plaza between FL-1 and FL-2. The worst interference is experienced when cleaning crews use gas-powered leaf blowers in the plaza. A secondary source of "interference" (signal degradation is a better description) is the trees that fill the plaza. As expected, signal levels and quality improve when the leaves fall off of the trees during the fall, and worsen when the leaves return in the spring. In the long term, the trees in the plaza will be removed (by UCAR) to facilitate construction to resolve drainage issues. Afterwards, some trees may be restored to the area, but we expect they will be sized and placed to avoid future problems. In the interim, we have been preparing to install a 3.8m mesh dish (from NCAR/EOL) on the western satellite pad at the Mesa Lab with the expectation that this location will be shielded from the TI problems that plagued our GOES-16 ingest.

  Very recently, we were given a 3.6m satellite dish that was being excessed by a private company that was relocating their operations. We are in the process of obtaining permission to install this dish at the UCAR Marshall site where, hopefully, there will be less of a chance of being affected by Terrestrial Interference created by 5G wireless technologies. The other possible use of this dish would be to install it on the western satellite pad at the NCAR Mesa Lab, and use it to ingest GOES-17. This use will only be possible if the TI that had been experienced bythe 4.5m dish has subsided.

- Continue to distribute GOES-16 and GOES-17 data via the LDM/IDD and serve the data via the TDS, ADDE and EDEX

The volume of data available in the SATELLITE datastream can be seen in:

http://rtstats.unidata.ucar.edu/cgi-bin/rtstats/iddstats_vol_nc?SATELLITE+oliver.unidata.ucar.edu

- Work with NOAA/GSL to get a feed of their **RRFS** (Rapid Refresh Forecast System) when it has matured enough for the products to be distributed to end users for evaluation.

Future Activities

CSPP GEO Gridded Geostationary Lightning Mapper (Gridded GLM)

On March 21, Graeme Martin (UWisconsin/CIMSS) announced the initial release of **Gridded Geostationary Lightning Mapper (Gridded GLM)** software package:

The software is capable of processing GOES-16 and GOES-17 GLM Level 2+ products in mission standard format, generating a new set of products which have been gridded to the Advanced Baseline Imager (ABI) 2-km resolution, and are aggregated at one-minute intervals. Spatial extent information that is not readily available in the GLM L2+ data is recovered and used to create the gridded products.

The following products can be produced:
- Minimum Flash Area
- Flash Extent Density
- Total Optical Energy

AWIPS-compatible tiles can optionally be generated, using functionality that was developed within the open source Python SatPy library.

Input GLM L2+ files can be obtained from the CSPP Geo GRB software running at a direct broadcast site, or from NOAA CLASS. Output is in NetCDF4 format.

We intend to implement this software, evaluate the products, and distribute them in the IDD when appropriate.

NOAAPort SBN

The NWS has informed COMET that they would no longer be supporting the NOAAPort downlink facility that is installed near the front entrance of Foothills Lab 2. At this point, it is unclear if the NWS would want to remove the dish or to transfer ownership of the dish to UCAR. Our position is that we want/need to keep the downlink capability working as it is one of the major sources of NOAAPort-derived content in the IDD.

We are trying to keep abreast of a possible expansion of the NOAAPort SBN that would, if implemented, increase available bandwidth twofold. As one might imagine, progress on this

and other fronts has been slowed by reorganization of priorities during the COVID-19 epidemic.

IDD NIMAGE and UNIWISC Datastreams

As noted earlier, both the **NIMAGE** and **UNIWISC** datastreams have been revamped to include GOES-16/17 imagery and products. We will add more products if asked to do so by the governing committees. The volume of data available in the **NIMAGE** and **UNIWISC** datastreams can be seen in:

http://rtstats.unidata.ucar.edu/cgi-bin/rtstats/iddstats_vol_nc?NIMAGE+oliver.unidata.ucar.edu

http://rtstats.unidata.ucar.edu/cgi-bin/rtstats/iddstats_vol_nc?UNIWISC+oliver.unidata.ucar.edu

VALUE-ADDED Products

Texas Tech University (Eric Bruning) is creating value-added Level 2 products created from Geostationary Lightning Mapper (GLM) images as a precursor for similar products potentially being added to NOAAPort. We added these Level 2 products to the NIMAGE IDD datastream at the same time that we added the NOAAPort-delivered Level 2 products. The GLM Level 2 products are directly usable by all of the analysis and display packages that we make available with the exception of GEMPAK.

More recently, the **NIMAGE** feed was enhanced by the addition of three RGB products being created at CSU/CIRA. The repurposing of **NIMAGE** from a feed that only contained imagery that derived from NOAAPort to one in which new, innovative Level 2 satellite products has seemingly been accepted by the community.

We welcome contributions of additional value-added products by TTU and other sites.

SSEC Collaboration

Continue working with SSEC on their *fanout* approach that insulates GRB ingestion from expected (e.g., NCAR twice per year power downs; twice per year solar interference periods; etc.) and unexpected (e.g., TI caused) service interruptions

L2 Product Creation Testbed

We still intend to establish a test bed for the creation of Level 2 (L2) products from GOES-16/17 imagery, model output and observational data.

The objective is to provide the capability of running user site submitted algorithms to create

L2 products and make them available for testing for a short period of time via the IDD, the TDS, McIDAS ADDE and AWIPS EDEX. This initiative has been slowed by the inability by most staff to work on-site.

Relevant Metrics

- Lots O Data!

  The volume of GOES-16 and GOES-17 GRB products (12 GB/hour ave and 16 GB/hour max; this is the most voluminous IDD feed!) can be seen in the real-time statistics plot from our GOES-R ingest machine:

  http://rtstats.unidata.ucar.edu/cgi-bin/rtstats/iddstats_vol_nc?SATELLITE+oliver.unidata.ucar.edu

- Feeding data to a slowly growing list of sites via the IDD:

  We are distributing all or part of the GOES-16 GRB products to:

  - Groups within UCAR/NCAR (3: all products Unidata, EOL, RAL))
  - U.S. Universities (25: variety of feeds; GLM very popular)
  - U.S. Government (3: all products to 2 NOAA sites and one Military site)
  - International (3: Full Disk imagery and GLM L2 products)

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   Providing TDS, ADDE and EDEX servers for GOES-16/17 imagery and products benefits the greater community by providing access to real-time observations from the U.S. operational satellite constellation.

2. **Supporting People**
   Providing access to data in real-time has been a fundamental Unidata activity since its inception. Continuing to provide data enables Unidata sites to focus on their educational and research activities.

Status Report: Internet Data Distribution

October 2020 - April 2021

Steve Emmerson, Mustapha Iles, Mike Schmidt, Jeff Weber, Tom Yoksas

Questions for Committee Members

- Do you have suggestions regarding content of data streams like CONDUIT, FNEXRAD, NIMAGE, UNIWISC, NLDN Lightning, etc.?

  We (UPC, the Unidata community and UAlbany for the NLDN component of LIGHTNING) have control of the content of these data streams, so their contents are open for suggestions.

Activities Since the Last Status Report

Internet Data Distribution (IDD)

IDD data volumes continue to increase. The following output is from a Linux-based motherlode clone that the UPC operates on behalf of the community, lead.unidata.ucar.edu:

```
20210415

Data Volume Summary for lead.unidata.ucar.edu

Maximum hourly volume 119058.717 M bytes/hour
Average hourly volume  66934.329 M bytes/hour

Average products per hour     502933 prods/hour

Feed                            Average             Maximum     Products
                       (M byte/hour)            (M byte/hour)   number/hour
SATELLITE               14504.943    [ 21.670%]    20110.400     6373.289
CONDUIT                 14241.170    [ 21.276%]    58807.539    90655.756
NGRID                   10877.426    [ 16.251%]    15887.426    74143.156
NEXRAD2                  8542.694    [ 12.763%]    10822.452   103162.978
NIMAGE                   7387.537    [ 11.037%]    11919.019     6004.178
HDS                      3740.647    [  5.589%]     9431.904    40612.378
NEXRAD3                  2288.151    [  3.419%]     2845.997   101206.733
FNMOC                    1800.878    [  2.691%]     9777.195     6913.911
UNIWISC                  1217.099    [  1.818%]     4872.587     2030.889
FSL2                      732.583    [  1.094%]     1793.870     1558.111
GEM                       690.749    [  1.032%]     3691.804    15961.289
IDS|DDPLUS                358.972    [  0.536%]     1249.387    53181.689
```

```
NOTHER                266.674    [  0.398%]      930.846       64.867
LIGHTNING             167.834    [  0.251%]     1089.446      935.533
FNEXRAD               113.174    [  0.169%]      175.129      103.511
GPS                     3.800    [  0.006%]       73.042       24.467
```

New Data Distribution:

IDD CONDUIT feed:

The GFS model was upgraded from v15.3 to 16.0. The increase in data volume and number of products in CONDUIT was significant enough to warrant Unidata upgrading its ingest machine to be able to handle the increased volume while continuing to create the content for the FNEXRAD (NEXRAD Level III national composites), NIMAGE (GOES-16 and -7 Level 2 images and products fully reconstituted images from NOAAPort tiles and with broadcast headers and footers stripped off to leave “bare” netCDF4 files), and UNIWISC (select GOES-16 and -17 images converted to McIDAS AREA format for ease of use in legacy systems like GEMPAK).

IDD NGRID feed:

MRMS imagery/products we are receiving from an NCEP LDM feed and that is in GRIB2 format will be added to the IDD NGRID feed as soon as internal testing in Unidata packages (e.g., TDS and AWI:S) is complete. These products are currently being relayed to lead.unidata.ucar.edu, so the NGRID volume in the above listing reflects the additional ~4 GB/hr volume that users will experience when the feed goes live.

IDD NIMAGE feed:

The NIMAGE feed was enhanced by the addition of three products being created by CSU/CIRA: GeoColor, DebraDust and CloudSnow. GOES-East CONUS and -West PACUS coverages are available for the GeoColor and CloudSnow products while the DebraDust product is available in a GOES-East CONUS coverage. All three of these are RGB products - displays are created using different wavelength channels to drive the Red, Green and Blue portions of a composite display. The GeoColor product is quite interesting and useful, so users are encouraged to take a look!

Experimental HRRR feed:

Unidata used to receive experimental High Resolution Rapid Refresh (**HRRR**) grids (both 2D and 3D fields) in an LDM/IDD feed from NOAA/GSL and feed these products to a small number of university sites on **hrrr.unidata.ucar.edu** (which is also known as **lead.unidata.ucar.edu**). Once the HRRR data went operational, NOAA/GSL stopped creating experimental **HRRR** output. The experimental **HRRR** is, however, being replaced by the **RRFS** (Rapid Refresh Forecast System) in NOAA/GSL. We have requested a feed of these data, but

we have been told that the **RRFS** is still a couple/few months away from being available.

Existing Data Distribution:

The primary top level IDD relay cluster, idd.unidata.ucar.edu, has been operating well since its move to the NCAR Wyoming supercomputer facility in Cheyenne, WY.

The data volume seen in the **SATELLITE** (which is known as **DIFAX** in LDM distributions prior to v6.13.6) listing above represents all products received in the GOES ReBroadcast (GRB) downlinks that we installed in UCAR (GOES-17 at the NCAR Mesa Lab and GOES-16 at UCAR Foothills Lab 2). The data volume seen in the **NIMAGE** entry represents GOES-16/17 ABI Level 2 imagery that has been reconstituted by stitching together tiles that are distributed in NOAAPort and all other Level 2 products. In both cases, binary headers and footers that are added to products before distribution in NOAAPort have been stripped off leaving "raw" netCDF4 files. The **UNIWISC** feed represents the volume of 3 select channels (0.64um VIS, 6.2um WV and 10.3um IR) for all coverages (CONUS, FullDisk, Mesoscale-1 and Mesoscale-2) of GOES-16/17 image products that are in PNG compressed McIDAS AREA format that is suitable for use in GEMPAK, the IDV, McIDAS-V, and McIDAS-X.

Challenges, problems, and risks:

More sites are installing intrusion detection/prevention systems (e.g., Palo Alto), which can adversely affect LDM throughput if not configured correctly.

Ongoing Activities

We plan to continue the following activities:

- Unidata took over the data distribution of GPS radio occultation solutions from COSMIC. COSMIC will still gather incoming GPS data and create the solutions, but due to hardware constraints COSMIC has requested Unidata to provide distribution from our top level IDD relay clusters (idd.unidata.ucar.edu and iddb.unidata.ucar.edu) to the community. The solutions (Precipitable Water Vapor and Total Electron Content-Ionosphere) are in netCDF format and are available in the GPS feedtype.

- Many, but not all, of the products in NCEP operational HRRR are being distributed in the NOAAPort SBN and relayed in the IDD NGRID feed. These products served with other model output in the TDS and in AWIPS EDEX:

- Other data sets we continue to explore with NOAA/GSD/ESRL are:
    - FIM
    - HIWPP
    - RRFS

- NCEP (operational) HRRR fields and forecasts times were added to the IDD CONDUIT

datastream.

NOAAPort Data Ingest

- Ingest of the DVBS-2 NOAAPort Satellite Broadcast Network (SBN) products and their relay to end-users via the IDD has been “operational” at the UPC since August 2014.

  Unidata has been assisting LSU/SRCC in the maintenance of their NOAAPort ingest capability for several years. Activities have included providing a spare LNB to bringing their NOVRA S300N receiver to Boulder for testing, configuration, power supply replacement and routine monitoring of their data and distribution. We have recently been informed that the effort to secure 5-year funding has failed, so all operations, including NOAAPort ingest, will be shut down. The most recent information that we have is some transition funding has been secured, so the NOAAPort ingest activities will continue for the “next fed months”.

  During this reporting period considerable effort has been expended to streamline our NOAAPort ingest systems and assist other sites (UW/SSEC, NOAA/GSL, NOAA/SPC) in troubleshooting problems being experienced in their systems. Most recently, the NOAAPort ingest capability in the LDM has been enhanced to allow the user to specify the size of the ingest buffers used. This capability, while requested by NOAA/Raytheon has increased our understanding of the need for additional input stream buffering, and the idea has been transferred to our GRB ingest efforts where significant improvements have been experienced.

- The NOAAPort-derived data streams (**HDS**, **IDS|DDPLUS**, **NGRID**, **NIMAGE**, **NEXRAD3** and **NOTHER**) are redundantly injected into the IDD at three geographically separate locations: UCAR/Unidata, UW/SSEC, and LSU/SRCC. The **NOTHER** data stream contains GOES-16 and GOES-17 tiles that need to be stitched together to make full image scenes usable to end-user applications. Unidata is using Ryan May’s ldm-alchemy package (available in the Unidata section of Github) to create full ABI L2 images that are then relayed in the **NIMAGE** datastream which was revamped earlier in the year since the content of the NOAAPort-received **NIMAGE** products dropped to essentially zero when GOES-15 was put into standby storage. Even though GOES-15 will periodically be taken out of storage for use in western Pacific hurricane monitoring, the GINI image products that were once created from its scans will no longer be produced by NOAA.

- We are now looking for a site that could replace LSU/SRCC as a third contributor to NOAAPort-derived content in the IDD. We are experimenting with using a feed froma commercial entity (Allisonhouse.com) on a quid pro quo basis.

- Unidata's NOAAPort ingest package is bundled with current versions of the LDM. The current LDM release is v6.13.13. A new LDM distribution is being readied for release.

Relevant Metrics

- Approximately **571** machines at **213** sites are running LDM-6 **and** reporting real-time statistics to the UPC.

  We routinely observe that the number of sites reporting real-time statistics fluctuates. We are not certain why this may be the case, but our best guess is that some sites do not keep their LDMs running all of the time; campus firewall adjustments block the sending of the statistics; and/or sites decide to stop sending statistics.

  We know that there are a number of sites that are participating in the IDD, but are not reporting real-time statistics back to us. Reporting of real-time statistics is not and never has been mandatory.

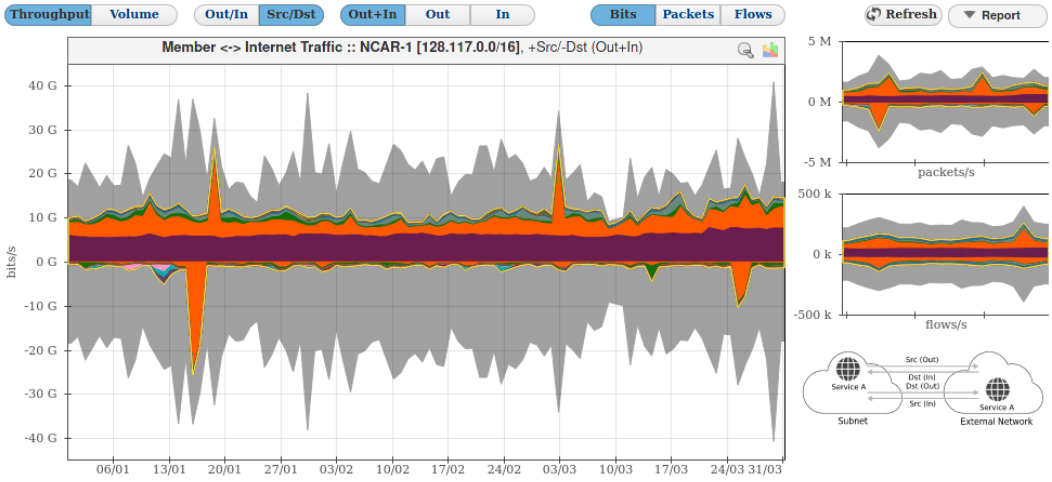
  Unidata staff routinely assist in the installation and tuning of LDM-6 at user sites as a community service. We have learned about sites not sending real-time statistics during these kinds of support activities.

- A number of organizations/projects continue to use the LDM to move substantial amounts of data that do not report statistics to Unidata: NOAA, NASA, USGS, USACE, Governments of Spain, South Korea, private companies, etc.).

- UCAR IDD toplevel relays, **idd.unidata.ucar.edu and iddb.unidata.ucar.edu**

  The IDD relay clusters, described in the June 2005 CommunitE-letter article Unidata's IDD Cluster, routinely relays data to more than 1205 downstream connections. The primary IDD relay cluster, **idd.unidata.ucar.edu**, was moved to the NCAR/Wyoming Super Computing facility in Cheyenne, WY in late August 2019.

  Over the period from January 1 through March 31, 2021 the average volume of LDM/IDD data flowing from the UCAR/NCAR network averaged around 5.5 Gbps (~59.4 TB/day), and peak rates reached 8 Gbps (which would be ~86.4 TB/day if the rate was sustained).



Services	Src (Out+In)		Dst (Out+In)		Total (Out+In)	
	Avg	Max	Avg	Max	Avg	Max
LDM [388]	5.5 Gbps (36.:	8.0 Gbps	42.3 Mbps (0.	59.9 Mbps	5.5 Gbps (18.:	8.1 Gbps
HTTPS [443]	2.7 Gbps (17.:	18.3 Gbps	364.6 Mbps (2	25.0 Gbps	3.1 Gbps (10.	28.2 Gbps
Port [22]	318.8 Mbps (2	1.9 Gbps	183.1 Mbps (1	3.2 Gbps	501.9 Mbps (1	3.4 Gbps
Port [80]	372.8 Mbps (2	2.5 Gbps	3.3 Mbps (<0.	21.5 Mbps	376.1 Mbps (1	2.5 Gbps
Port [5201]	1.3 Mbps (<0.	2.4 Mbps	178.7 Mbps (1	361.7 Mbps	179.9 Mbps ((	364.0 Mbps
Port [4501]	98.6 Mbps (0.	562.8 Mbps	32.1 Mbps (0.	298.2 Mbps	130.7 Mbps ((	631.1 Mbps
Port [112]	72.6 Mbps (0.	142.1 Mbps	681.7 kbps (<	1.3 Mbps	73.2 Mbps (0.	143.4 Mbps
Port [50040]	47.0 Mbps (0.	323.3 Mbps	2.9 Mbps (<0.	380.6 Mbps	49.9 Mbps (0.	420.3 Mbps

The following table shows how consistent the volume of data flowing to downstreams out of UCAR has been:

Date range	Src Ave Max	Dst Ave Max	Total Ave Max
20200508 - 20200630	5.4 7.5	42.1 52.9	5.5 7.5
20200701 - 20200930	5.4 7.9	41.9 60.3	5.4 7.9
20201001 - 20201231	5.2 6.9	39.9 55.9	5.3 7.0
20200101 - 20200331	5.5 8.0	42.3 59.9	5.5 8.1
20200401 - 20200415	6.1 15.5	46.4 112.7	6.1 15.7

NB: The units for Src and Total Ave and Max are Gbps (gigabits per second), and the units for Dst are Mbps (megabits per second).

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   The IDD project demonstrates how sites can employ the LDM to move and process data in their own environments.

2. **Providing Useful Tools**
   The freely available LDM software and the IDD project that is built on top of the LDM have served as a demonstration for distribution of real-time data for a variety of organizations including the U.S. National Weather service.

   The cluster approach for LDM/IDD data relay that Unidata pioneered has been been adopted by several Unidata university sites, and is currently being implemented at U.S. government sites.

   Unidata's NOAAPort ingest package, which is bundled with LDM-6, is being used by a variety of university, U.S. government, and private sector entities.

   Both the LDM and NOAAPort ingest packages are bundled with AWIPS.

3. **Supporting People**
   The IDD is the primary method that core Unidata sites use to get the meteorological data that they need. Providing access to data in near real-time is a fundamental Unidata activity. The IDD-Brasil, the South American peer of the North American IDD, and IDD-Caribe, the Central American peer of the North American IDD, are helping to extend real-time data delivery throughout the Americas

---

Prepared *April, 2021*

Status Report: IDV with RAMADDA

October 2020 - April 2021

Yuan Ho, Julien Chastang

Areas for Committee Feedback

We have no questions at this time.

Activities Since the Last Status Report

IDV Release

IDV 5.7 has been released on February 5 of 2020.

IDV System Changes

__Latest netCDF-Java Version__

The version of the netCDF-Java library currently distributed with the nightly release version (5.7u1) is the 5 .4.2-SNAPSHOT . The prior version of netCDF-Java to be distributed with the stable release IDV (5.7) was 4.6.15. There have been many new features and bug fixes in that range. The complete release notes for these versions can be found here. The nightly release version 5.7u1 has fully integrated with the netCDF-Java 5.4 distribution. There are many new developments related to the new grid coverage features.

__IDV Certificates__

Java Webstart, Windows app and MacOS certificates have been renewed and will be valid until at least May 30, 2020 (MacOS certificate is valid until 2024). Moreover, as properly signing the IDV under these different environments can be an involved process, this information has thoroughly documented here.

__Changes to nightly release that will eventually be incorporated into into stable version__

- IDV uses the latest Java 8 AdoptOpenJDK
- IDV employes latest Java3D (1.6.2)
- Updated the IDV code signing certificates on all platforms (i.e., MacOS, Windows, Webstart)

- IDV now “notartized” on MacOS
- Updated Unidata's Install4J license from version 5 to 8.
- Updated the IDV Install4J configuration.

IDV Display Changes

__New Volume Rendering__

The volume rendering capability native to IDV produces a semitransparent volume representation. The new volume rendering methodologies in IDV implement a on-the-fly resample to a Linear3DSet. By using the Stack2D algorithm, any field can be transformed into a binary representation. This feature is still under development and testing.

__Cross Seam Subsetting__

The cross seam subsetting in the grid coverage type works similar to the grid file subsetting where you can set the property for all fields in the Data Source or override the default for a particular field in the field selector subset panel. The only difference in the Coverage data spatial subset is that it will allow users to flip longitudes in a cyclic rect linear grid from 0/360 to -180/180 (or vice-versa) before performing the cross seam subsetting. This feature is still under development and testing.

__Text Html Display__

New enhanced Text/Html display avaiable under Displays > Special for displaying straight text or HTML, this display can also be added to the view window legend panel for descriptive text about the display.

__Grid Diagnostics Formulas__

The new GFS dataset collection has several parameters with multiple accumulation hours in the time axis for each time step and the previous logic in the IDV picks the first one in the list, unfortunately, always the longest one. So the new logic will try to pick the shortest accumulation hour for each time step, and what we got are 3 and 6 hours for GFS one degree or 0.25 degree, we can use the subtraction to get the 3-hour time step accumulation. This change has been added as a derived variable named: TimeStepAccumulatedPrecip. In general, this formula can be applied to any mixed interval parameters.

__Latest Version of VisAD__

The SSEC team at UW, Madison has made a number of improvements to support 3D trajectories.

IDV Community Support

In the transitioning to a remote-learning system as a result of the COVID-19 pandemic, help local MSI university to borrow several refurbished MacBook computers for the students to be able to run Unidata's Integrated Data Viewer (IDV) at home. I provided a remote IDV training class to a group of radar class students from the University of Millersville. Yuan also prepares several remote IDV training classes for the coming school year.

KIOSK IDV Project

In collaboration with UCAR Center for Science Education and Computational Information Systems Laboratory, the project developed an extended IDV package for a Real-Time Weather Museum Touchscreen. This new real-time weather museum touchscreen display will undergo further usability testing to eventually join other weather and climate exhibits at NCAR's Mesa Lab in Boulder, CO, and at the NCAR-Wyoming Supercomputing Center Visitor Center in Cheyenne, WY.

IDV Publication Highlights

Synoptic–Dynamic Meteorology in 3D: Introducing an IDV-Based Lab Manual by Gary Lackmann, B. Mapes and K. Tyle

A Google Scholar Search reveals a number of publications that cite use of the IDV (doi:10.5065/D6RN35XM).

IDV and RAMADDA Training, Conference Attendance and Presence

__2019 AGU Fall Meeting__

- Hydrology Model Water Resource Estimation with Advanced 3D Visualization and Analysis
- An update in Hydrology Model Development and Application

__2019/2020 IDV lectures at WRF tutorial workshop__

- Yuan delivered two IDV introduction lectures on two semi-annual WRF tutorial workshops.

Ongoing Activities

We plan to continue the following activities:

__IDV 6.0 __

The coming release of the IDV 6.0 will have the updated JRE (AdoptOpenJDK) and Java 3D library.

__Investigation of Java 3D Alternative__

Because of concerns about the long-term viability of the open-source Java 3D project, the IDV team has begun discussions with our University of Wisconsin, SSEC collaborators to replace Java 3D with a more viable alternative within the VisAD API. We have started investigating whether the Ardor 3D can meet that objective. Looking into alternatives to Java 3D was also a goal described in the Unidata 2018 Five-year plan.

New Activities

Over the past few months, we plan to organize or take part in the following:

We plan to finally migrate away from Oracle Java 8u51 and towards a more modern version of OPenJDK Java. This switch will necessitate altering the IDV building and distribution workflow to work with OpenJDK.

Relevant Metrics

__E-Support__

The IDV team continues to provide the geoscience community with high-quality support through e-support software and idv-users mail list. In the last half year the IDV team has closed ~40 e-support tickets. Each individual ticket may and often does involve many back-and-forth messages. There is an especially large number of support requests coming from international users.

Top ten universities running IDV are: Millersville, Oklahoma, University of Utah, St Cloud state, Plymouth, NC State, West Kentucky, Lyndon State, University of Illinois, and San Francisco State.

__GitHub Pull Requests__

In the area of greater collaborative development, since the migration of the IDV project to github, we have closed a total of 125 “pull requests” or code contributions from internal and external collaborators.

__Youtube IDV Instructional Videos__

In the area of online IDV training, the Youtube IDV instructional videos have been viewed thousands of times.

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   The IDV is a state of the art geoscience visualization application. It gives users the ability to view and analyze a rich set of geoscience data, including real time data, in a seamless and integrated fashion. This analysis is captured in IDV bundles. RAMADDA is a content management system and service specifically tailored towards the sharing and distribution of IDV bundles facilitating distribution of scientific data and analysis.

2. **Providing Useful Tools**
   The IDV has been an open-source project for several years. The IDV is available on the github version control platform for greater open-source collaboration. The IDV provides users the unparalleled ability to analyze, integrate, and visualize heterogeneous geoscience data in two, three, and four dimensions. The IDV coupled with RAMADDA enables geoscience specialists the capability to share and collaborate their IDV analysis via social scientific networks.

3. **Supporting People**
   Unidata offers yearly multi-day training and occasionally regional workshops for IDV and RAMADDA. The IDV coupled with RAMADDA enables our earth science community partners to distribute geoscience data and metadata through web-based technologies thereby fostering scientific collaborations. Moreover, the IDV's ability to share bundles through RAMADDA creates a scientific social and collaborative network for the geoscience community.

Status Report: Information Technology

October 2020 - April 2021

Mike Schmidt, Matt Perna, & Jennifer Oxelson

Major Activities

****Remote working**** -- with the entire Unidata staff working remote, it's been a successful test of the infrastructure including bandwidth through the gateway systems and remote supporting staff with software and hardware issues and well as security and update management. Fortunately, we now have all three people listed above allowed and certified to be onsite if the circumstances warrant.

****Network upgrades**** -- as UCAR upgrades their backbone infrastructure to 100Gb/s links, we will continue to upgrade our data movers (IDD cluster nodes, data aggregators, motherlode clones) from 2 x 1Gbp/s bonded ethernet to 10Gb/s as necessary.

****Security**** -- we continue efforts to keep services and systems secure which takes consistent attention and occasional herculean efforts (to patch everything all at once). Unidata staff have moved as a group to use Duo two factor authentication. Initial access to most Unidata and UCAR resources requires some form of two factor authentication.

****LDM 7 node**** -- we maintain a LDM7 test node at the Front Range GigaPOP (FRGP) just off downtown Denver in co-location with the major backbone networks supporting FRPG participants (UCAR, ..). We expect to support intensive data movement and LDM testing for the next few years on this effort.

Ongoing Activities

We plan to continue the following activities:

- Day-to-day system and network support to the community as needed
- Resolve daily staff help desk issues
- Maintain security profile and exceed UCAR security standards

Prepared *April, 2021*

Status Report: LDM

October 2020 - April 2021

Steve Emmerson, Tom Yoksas, Mike Schmidt, Mustapha Iles, Yuanlong Tan (University of Virginia)

Activities Since the Last Status Report

LDM

The LDM is the primary software package by which research and education institutions obtain near real-time meteorological and related data.

Progress has been made on the following:

- NOAAPortIngester:
    - GEMPAK tables:
        - Added many new WMO parameters
        - Added many NWS MRMS parameters
    - Added option to set the receive buffer size to accommodate a request from the NWS. The operating system must allow the new size. This had the added benefit of reducing the number of missed NOAAPort packets at the Unidata Program Center.
- pqact(1):
    - Squashed bug that caused abnormal termination under extremely rare conditions
    - Improved documentation and log messaging
- ldmadmin(1):
    - Made handling of netstat(1) output more general
    - Improved help for the “pqactcheck” command
- scour(1):
    - Replaced scour(1) script with multi-threaded C program for vastly improved performance
- LDM registry:
    - Added parameter for scour(1) configuration-file
    - Added SCOUR_EXCLUDE_PATH for pathname of file that lists directories to be excluded by new scour(1) program
- pqcreate(1):
    - Increased the mean number of bytes in a data-product from 51 k to 140 k for computation of the number of slots when the registry parameter is “default”
- Logging:
    - Added FATAL logging level for serious problems
- Miscellaneous:
    - Removed lint identified by clang(1) and Coverity Scan
- Support:
    - Answered many questions from Universities, NOAA, US Military, and

corporations
    - Troubleshot several sites that were having problems that were, overwhelmingly, network-related

Dependencies, challenges, problems, and risks include:

The LDM is sometimes held responsible for decisions made by the NWS when they don't follow their own policy on how to categorize and name data products (not a new challenge).

More sites are installing intrusion detection/prevention systems (e.g., Palo Alto), which can adversely affect LDM throughput if not configured correctly.

Mathew Lazarra (University of Wisconsin–Madison) is spearheading an effort to formally integrate the LDM into the Antarctic meteorological data-distribution network.

Multicast LDM (aka LDM-7)

The multicast LDM project is separately funded by CISE in NSF. The goal is to reduce the outgoing bandwidth requirement of the LDM -- yet retain the current level of reliability -- by converting it into a hybrid system that combines use of the new, semi-reliable multicast protocol developed jointly with the University of Virginia with the time-tested unicast capability of the current LDM.

This project ended April 1st, 2019, but a no-cost extension was approved because funding was delayed.

Progress has been made on the following:

- Improved handling of the OpenSSL library. (Because any site can send to an IP multicast group, security is a concern and was a point in the proposal.)
- Corrected bug in the logic for tracking of the last multicast product and made its handling more robust
- Created three levels for the message authentication code (MAC) for multicast packets depending on the value of environment variable FMTP_MAC_LEVEL:
    - Unset or 0: No MAC
    - 1: SHA256 Hash-based MAC
    - 2: Ed25519 Digital Signing Algorithm
- Created a "BlackHat" module for spoofing valid packets to verify the message authentication mechanism
- Conducted numerous test runs to explore the parameter space (i.e., performance as a function of 3 security-levels, 5 packet sizes, and 5 VLAN bandwidths). Each run takes an hour.
- Improved log messaging
- Improved documentation

Dependencies, challenges, problems, and risks include:

- The amount of manual intervention required to maintain the multipoint VLAN is

considerable

Ongoing Activities

We plan to continue the following activities:

- Support and maintain the LDM

New Activities

Over the next twelve months, we plan to organize or take part in the following:

- Work with NSF's Polar Programs regarding integration of the LDM into the Antarctic meteorological data-distribution network

Relevant Metrics

- Data on LDM downloads
- The LDM system at the Unidata Program Center powers the Unidata IDD (Internet Data Distribution) system. Metrics on that program can be found in the IDD status report.

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   By enabling researchers, teachers, and students to process a wide variety of meteorological and related data in near real time.

2. **Providing Useful Tools**
   By enabling researchers, teachers, and students to obtain a wide variety of meteorological and related data in near real time and at no cost via the Internet.

   By using the LDM to move data into the cloud and developing multicast technologies.

3. **Supporting People**
   By answering support questions, writing documentation, and conducting workshops.

Prepared *April 2021*

Status Report: McIDAS

October 2020 - April 2021

Tom Yoksas

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Are there any features that users would like to be added to Unidata McIDAS-X and/or ldm-mcidas?

Activities Since the Last Status Report

Aside from routine updates/bugfixes to existing code and tables, the main area of activity recently has been the incorporation of Unidata-developed code into the next UW/SSEC McIDAS-X/XCD release, v2021..

Current Activities

- Support use of McIDAS-X/XCD in the community
- Ensure that the Unidata instances of McIDAS ADDE continue to function efficiently (ADDE serves on the order of 1.7 TB/day)
- McIDAS-X is used to convert GOES-16 ABI imagery that is in netCDF4 format to McIDAS AREA format that is usable by all supported display and analysis packages except Python/MetPy for the Unidata-Wisconsin (**UNIWISC** aka **MCIDAS**) IDD feed.

The v2020 release features the following:

- SSEC added support for McIDAS-X on RHEL 8 systems, ended support of McIDAS-X on RHEL 6 and Windows 7 systems, and added preliminary support of McIDAS-X on Windows 10 systems.
- Updated GOES-R Series ABI servers to list calculated resolutions in **IMGLIST** FORM=BAND and FORM=ALL output, and to support GOES-17 Fusion imagery because of modifications to allow more flexible file naming formats.
- Updated **GEO** command with improved logic when using Himawari imagery, and when merging imagery from multiple satellites and the domain contains the dateline.
- Improved performance of **RGBDISP** command, especially when it is run multiple times in a script.
- Updated VIIR servers and calibration module to correctly set the RAW value of Band 18

(M13) pixels to zero in bowtie deletion and bad or missing line regions.

Ongoing Activities

We plan to continue the following activities:

- SSEC McIDAS Advisory Committee (MAC)

  The UPC (Yoksas, Ho) continues to participate as the Unidata representative to the McIDAS Advisory Committee (MAC) that is operated by SSEC.

  The MAC was assembled by UW/SSEC to advise SSEC on McIDAS-X users needs/concerns/desires for development in the next generation McIDAS, McIDAS-V. The MAC was modeled after the Unidata IDV Steering Committee.

- Interest in McIDAS by non-core users

  The UPC occasionally receives requests for McIDAS-X and help using McIDAS-X from international university users, U.S. government agencies and other non-traditional Unidata users (e.g., private businesses, etc.). Government agencies and non-traditional Unidata users are referred to UW/SSEC for access to McIDAS; international educational community user requests are granted on a case-by-case basis after they provide a clear statement of their acceptance of the terms of use provided by SSEC.

- Continued support of existing and new community members

New Activities

- Add support for new types of data when they become available, otherwise McIDAS-X support is in maintenance mode.

Relevant Metrics

- Data delivered by the Unidata McIDAS ADDE servers exceeds 1 TB/day. The great majority of the data being served is imagery from GOES-16 followed by imagery from GOES-17.

- McIDAS-X/-XCD Inquiry Metrics

ldm-mcidas Decoders Activities

Development

ldm-mcidas releases are made when needed to support changes in software development and operating system environments. **ldm-mcidas** v2012 was released at the end of September, 2012.

Planned for the next major addition to this package will be the development of a "decoder" for GRB delivered Geostationary Lightning Mapper (GLM) data. This development is aimed at greatly increasing the speed at which displays of the GLM data can be made in McIDAS-X, the IDV and McIDAS-V.

Geostationary Satellite Data Ingest and Data Serving

Unidata continues to ingest GOES-East and GOES-West imager data at the UCAR Foothills Lab and NCAR Mesa lab campuses in Boulder.

- Direct, programmatic access to real-time GOES-East (GOES-16) and GOES-West (GOES-17) data via McIDAS ADDE services on three publicly accessible servers (lead.unidata.ucar.edu, adde.ucar.edu (aka atm.ucar.edu) and adde.ssec.wisc.edu) has been averaging on the order of 48 TB/month for the past 6-9 months.

Planned Activities

Ongoing Activities

Continued ingest, distribution via the IDD and ADDE serving of GOES-East and GOES-West imagery from the GRB downlinks we installed in UCAR

Continued ingest and ADDE serving of GOES-15 and GOES-14 imagery when available. GOES-15 and GOES-14 were put into standby mode on March 2, 2020. GOES-14 remains in its standby location (104W) and will be turned on for periodic testing as needed. GOES-15 supplemental operations began on Sunday, August 9, 2020 at 0000 UTC and continues to provide surveillance during Pacific hurricane seasons.

These efforts require maintenance of the satellite ingest and data serving equipment.

New Activities

Establish a testbed for generating Level 2 products from GOES-16/17 imagery and select model output. The intention is to be able to test vetted algorithms submitted by community members for a long enough period for the algorithms to be fully tested.

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   Remote, programmatic access to data provided by the Abstract Data Distribution Environment (ADDE) environment of McIDAS has been a model for the development of remote access methodologies since 1994. Concepts articulated in ADDE inspired the development of THREDDS (to address the lack of rich metadata available in ADDE) and RAMADDA. ADDE remains one of the most used data services in the Unidata suite. ADDE servers operated by Unidata are currently serving in excess of 1.6 TB/day.

2. **Providing Useful Tools**
   McIDAS remains the application of choice for the satellite meteorology community. The Abstract Data Distribution Environment (ADDE) component of McIDAS was the first application offered by Unidata to provide remote, programmatic access to a wide variety of data that is important to the atmospheric science community.

   The fifth generation of McIDAS, McIDAS--V, unlike its predecessors, is a fully open source application that is in wide scale and growing use in the worldwide satellite meteorological community

   McIDAS ADDE continues to evolve and provide access to a rapidly increasing volume of imagery and non-image data.

3. **Supporting People**
   McIDAS is still in active use by those interested in satellite meteorology worldwide.

Prepared *April, 2021*

Status Report: netCDF

October 2020 - April 2021

Ward Fisher, Dennis Heimbigner

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Are there other cloud-based block storage formats/locations (TileDB, Azure, etc) that are actively in use? What is the next venue for investigation once we have our Zarr support in place?
2. How can we encourage more user testing of the release candidates we provide?

Activities Since the Last Status Report

We are using GitHub tools for C, Fortran and C++ interfaces to provide transparent feature development, handle performance issues, fix bugs, deploy new releases and to collaborate with other developers. Additionally, we are using docker technology to run netCDF-C, Fortran and C++ regression and continuous integration tests. We currently have **164** open issues for netCDF-C, **68** open issues for netCDF-Fortran, and **40** open issues for netCDF-C++. The netCDF Java interface is maintained by the Unidata CDM/TDS group and we collaborate with external developers to maintain the netCDF Python interface.

In the netCDF group, progress has been made in the following areas since the last status report:

- The release of ncZarr (netCDF with native Zarr support) has been released as of netCDF-C version 4.8.0.
- Migrated the NetCDF User's Guide to a new, separate repository. This repository will contain the concise, language-agnostic summary of the netCDF data model. Language-specific documentation (primarily used by developers) will remain associated with the individual code repositories.
- Further enhancements to the netCDF-C documentation, modernization of the netCDF-Fortran and netCDF-C++ documentation.
- We continue to see a high volume of contributions to the netCDF code base(s) from our community. While these contributions require careful review and consideration, it is encouraging to see this model of development (enabled by our move to GitHub) being more fully embraced by our community.
- We have prepared and presented multiple times on the netCDF-Zarr roadmap and plans to working groups at both NASA and NOAA.

Dependencies, challenges, problems and risks include:

- The small group of netcdf developers is under a lot of pressure to provide project management as well as implement new features, fix bugs, provide esupport, etc. With 1.5 FTE assigned to the project, the workload is significant.
- Rapid evolution of Zarr standard is very useful, but also provides a bit of a moving target.
- Increase in external contributions has greatly increased the project management overhead for netCDF-C/C++/Fortran.
- Advances in compilers (GCC 10.x) and newer architectures (such as Apple's ARM M1 architecture) are requiring additional overhead to ensure compatibility.

Ongoing Activities

We plan to continue the following activities:

- Continue work towards adoption of additional storage options, separating out the data model from the data storage format (as much as possible).
- Provide support to a large worldwide community of netCDF developers and users.
- Continue development, maintenance, and testing of source code for multiple language libraries and generic netCDF utility programs.
- Continue modernizing the documentation for netCDF-C, Fortran and C++ libraries.
- Extend collaboration as opportunities arise, for increasing the efficiency of parallel netcdf-3 and netcdf-4.

New Activities

NetCDF/Zarr Integration

The netCDF team has released the first public version of netCDF-C which provides Zarr I/O compatibility, dubbed 'ncZarr'. This work has been highly anticipated, and well received, by the broader netCDF community.

- Release of XArray compatibility.

Over the next three months, we plan to organize or take part in the following:

- Release the first version of netCDF with Zarr+Xarray support (ncZarr).
- Release subsequent versions of netCDF-C, netCDF-Fortran, netCDF-C++.
- Continue modernizing/editing the netCDF documentation to provide easy access to documentation for older versions of netCDF.

Over the next twelve months, we plan to organize or take part in the following:

- Release an official Windows port of the netCDF-Fortran and netCDF-C++ interfaces.
- Continue to encourage and support the use of netCDF-4's enhanced data model by third-party developers.
- Expand support for native object storage in the netCDF C library.

- Continue to represent the Unidata community in the HDF Technical Advisory Board process.
- Continue to represent the Unidata community in the Zarr/n5 collaboration conference calls.

Beyond a one-year timeframe, we plan to organize or take part in the following:

- Improve scalability to handle huge datasets and collections.
- Improve the efficiency of parallel netcdf3 and parallel netcdf4.
- Continue to add support for both file-storage and object-storage options.

Relevant Metrics

Static Analysis Metrics

There are currently about 226,892 lines of code (up from 202,428 lines of code) in the netCDF C library source. The Coverity estimate for defect density (the number of defects per thousand lines of code) in the netCDF C library source has slightly decreased to **0.68,** where it was **0.68** six months ago. According to Coverity static analysis of over 250 million lines of open source projects that use their analysis tools, the average defect density with 100,000 to 500,000 lines of code is **0.50**.

Google Metrics

Google hits reported when searching for a term such as netCDF-4 don't seem very useful over the long term, as the algorithms for quickly estimating the number of web pages containing a specified term or phrase are proprietary and seem to change frequently. However, this metric may be useful at any particular time for comparing popularity among a set of related terms.

Currently, Google hits, for comparison, are:

- **951,000** for netCDF-3
- **884,000** for netCDF-4
- **1,130,000** for HDF5
- **116,000** for GRIB2

Google Scholar hits, which supposedly count appearances in peer-reviewed scholarly publications, are:

- **440** for netCDF-3
- **972** for netCDF-4
- **17,500** for HDF5
- **1,380** for GRIB2

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   by supporting the use of netCDF and related technologies for analyzing, integrating, and visualizing multidimensional geoscience data; enabling effective use of very large data sets; and accessing, managing, and sharing collections of heterogeneous data from diverse sources.

2. **Providing Useful Tools**
   by developing netCDF and related software, and creating regular software releases of the C, C++ and Fortran interfaces; providing long-term support for these tools through the various avenues available to the Unidata staff (Github, eSupport, Stackoverflow, etc).

3. **Supporting People**
   by providing expertise in implementing effective data management, conducting training workshops, responding to support questions, maintaining comprehensive documentation, maintaining example programs and files, and keeping online FAQs, best practices, and web site up to date; fostering interactions between community members; and advocating community perspectives at scientific meetings, conferences, and other venues.

Prepared *April, 2021*

Status Report: Outreach to Underserved Communities

October 2020 - April 2021

Doug Dirks, Jeff Weber, Joshua Young

Areas for Committee Feedback

We are requesting your feedback on the following topics:

Are you currently collaborating with an MSI?

Are there MSI's geographically close to you that you have not engaged with?

Activities Since the Last Status Report

Attended AISES-NSF Proposal Workshop

- Successfuly submitted a proposal to NSF solicitation CISE 21-533, Titled:
  A Sovereign Network System for Environmental Monitoring, Data and Information Exchange, and Collaboration among Tribal Colleges and Universities
- Institutions involved with Unidata are: Southwestern Indian Polytechnic Institute, Navajo Technical University, and Tohono O'odham Community College

Engaged with Rising Voices

- Involved with Indigenous Peoples Climate Change Working Group **(IPCCWG)**
- Involved with Indigenous-**FEWSS**. Indigenous Food, Energy & Water Security and Sovereignty

Internships

- Active engagement in the SOARS program
  - Selection Committee - completed for 2021

Progress has been made on the following:

- Designing structural changes (e.g. modifications to how equipment awards, internships, workshops, and committee placements are announced and selected)

Dependencies, challenges, problems, and risks include:

- The only known dependency is regarding funding and time both of which have been dedicated to this effort. Since this is a new project, other dependencies or risks have not been identified at this time.

Ongoing Activities

We plan to continue the following activities:

- SACNAS and Rising Voices engagement

New Activities

Over the next three months, we plan to organize or take part in the following:

- Continue to develop outreach stakeholder list for broadening distribution of opportunities (equipment awards, internships, etc) and engage remotely

Over the next twelve months, we plan to organize or take part in the following:

- Implement changes to the process of how Unidata opportunities are announced and awarded
- Plan exhibition or other activities at subsequent appropriate conferences
- Identify relevant metrics (contacts, partners identified, meetings attended)
- Identify sustaining partnerships for the next five years

Beyond a one-year timeframe, we plan to organize or take part in the following:

- This effort is an ongoing commitment for the next award period (5 years); however, during this first year we are piloting efforts and then will apply lessons learned for the next 4 years.

Relevant Metrics

Relevant metrics should be discussed and decided for reports going forwarded

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Providing Useful Tools**
   Better understanding these communities and characterizing their needs will allow us to develop more fit for purpose tools that can and will be adopted

2. **Supporting People**
   Unidata has always served the broad geoscience community; however we are making a concerted effort to expand our reach to underrepresented individuals and organizations as an emphasis of our new award

---

Prepared *April, 2021*

Status Report: Python

October 2020 - April 2021

Ryan May, Drew Camron, Sean Arms, Julien Chastang

Areas for Committee Feedback

We are requesting your feedback on the following topics:

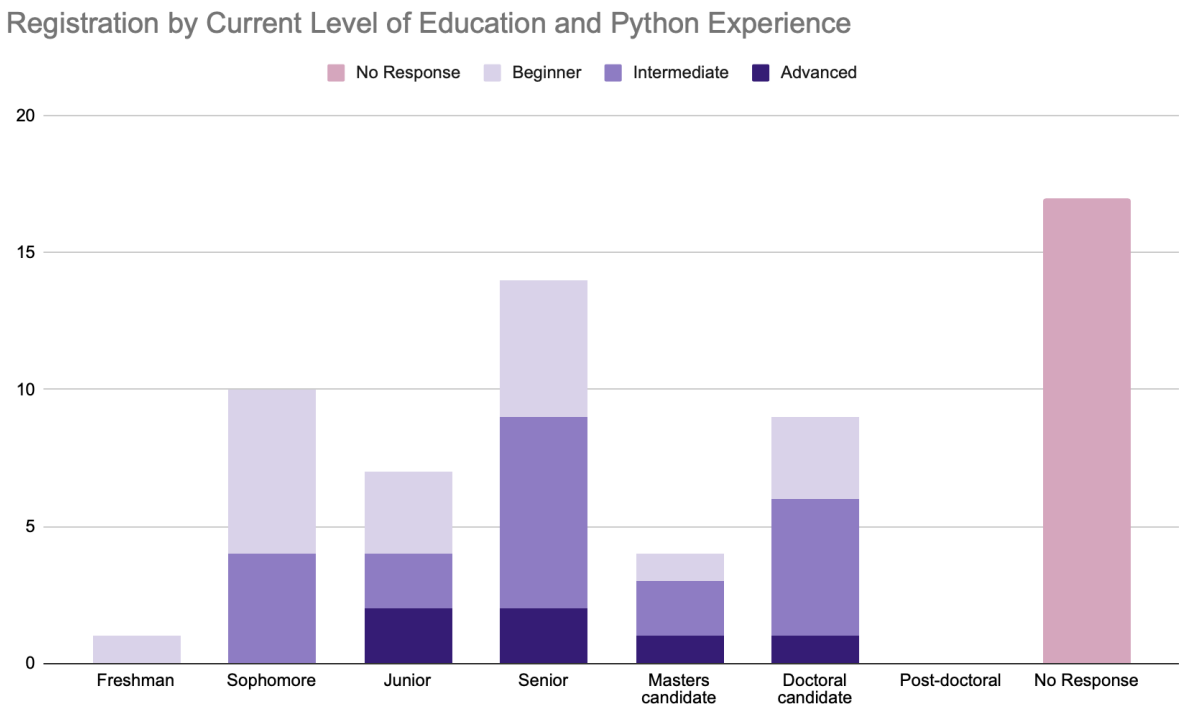
1. What are the pain points you encounter when trying to use Python for your work? When using MetPy and/or Siphon?

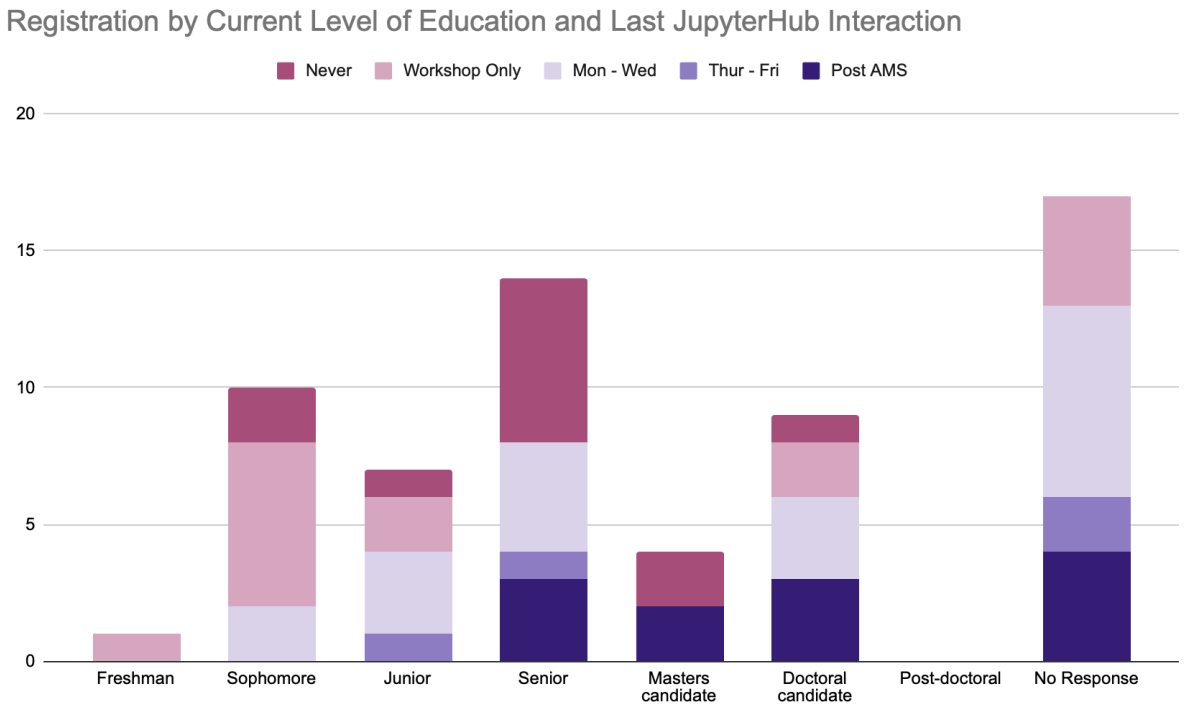
Activities Since the Last Status Report

Python Training Efforts

Python training efforts continue to be an important part of the Python portfolio. We continue to be successful in identifying opportunities to offer training within our resource constraints. Not only do these generate significant goodwill and grow our audience, but they are a significant source of information to inform our library development. One challenge is to balance time dedicated to creation of training materials, workshop preparation, and logistics against time devoted to support and Python software development. Motivated by COVID-19, remote work, and pushes for equity and inclusion in our offerings, we've enlisted the help of Unidata's instructional design staff to create new materials and tailor prior curricula for a primarily-virtual audience. Some of this material was first practiced at our contribution to the 2021 AMS Beginner Python Short Course and will continue to be a focus moving forward.

Alongside this, we are completing the unification of Unidata's Python training materials using sustainable technologies in-line with our other Python projects, including an automatically tested and deployed Sphinx-based website with Unidata-consistent theming. The materials are being restructured and reorganized with an eye to potential new scheduled training endeavors. This infrastructure and reorganization was sampled in constructing the AMS 2021 student conference workshop materials. That AMS 2021 student conference workshop had wide array of experiences and background register and participate:





Progress has been made on the following:

- Taught a virtual Python workshop at the 2021 AMS Student conference

- Taught a virtual Python short course through AMS in March 2021, in collaboration with other instructors including Kevin Tyle (SUNY Albany), Scott Collis (Argonne National Lab), and organized by Damien Irving (University of South Wales)
- A regular collaborative workflow has been established for contributing to Project Pythia, a collaborative Earthcube project with CISL, CGD, and the University of Albany, focused on building out a larger Python training portal with a cloud-deployable platform to execute the training materials.
- The AMS Intermediate Short Course on MetPy, collaborating with Kevin Goebbert, has been delayed until the 2022 AMS Annual Meeting.
- John Leeman continues to lead the "MetPy Mondays" effort.

MetPy

Development continues to be driven by requirements for our dedicated awards (in addition to bug reports and pull requests from community members). The latest developments have revolved around the MetPy 1.0 release, 1.0.1 release, and establishment of its path forward from there. Developments include

- "Finalizing" the API for user reliability across the entirety of 1.x.
- Constructing the foundation for the entire unit-aware xarray-based data model to enhance the functionality and consistency of our calculations.
- Full release of the declarative interface for rapid visualization

The focus for the 1.1 release is trying to merge in some features that have been languishing in "almost ready" state for some time, including plotting fronts and several community-contributed Pull Requests. We will also be dropping support for Python 3.6, following the trend in the broader scientific Python community. We will also be working on the early version of the "automated solver", which will allow MetPy to automatically calculate some parameters through a chain of calculations, based on the input data. (This is the last technical deliverable on the MetPy SI2 grant.) Beyond 1.1, we'll be working on porting some remote access functionality from Siphon into MetPy, as well as adding support for new sources like Bufkit.

Progress has been made on the following:

- MetPy 1.0rc2, **1.0**, and 1.0.1 were released
- MetPy 1.1 will be released late Spring 2021
- A proposal to NSF CSSI for enhancing MetPy's performance and scalability was submitted
- Work towards requirements of MetPy-related NSF awards
- Community awareness continues to grow, with the volume of engagement and mentions on social media growing; the MetPy twitter account has reached 1858 followers (35% growth in 6 months). There are now job postings that include MetPy in their list of skills.
- We have submitted a proposal to AMS for an article on MetPy in the Bulletin of the AMS.

Siphon and Data Processing

Siphon continues to grow and develop, though at a slower pace than MetPy; its development

tends to be driven by obstacles to access of remote data. The most pressing developments we anticipate for Siphon are improvements to working with Siphon in interactive sessions, like the Jupyter notebook environment: improved catalog crawling interface, better string representations, and tab completion. Siphon has seen infrastructure and usability improvements over these months, and the decision has been made to separate **non-TDS functionality** (e.g. Wyoming Upper Air archive access) out into a new remote-access toolset contained within MetPy.

We also continue to maintain the LDM Alchemy repository as a collection of LDM processing scripts in Python. Currently this includes the code powering the AWS NEXRAD archive as well as the program that reconstitutes NOAAPORT GOES-16/17 imagery. As we transition more of our internal data processing to Python, this repository will hold those scripts. We have seen several community questions regarding both the GOES and NEXRAD processing software.

Progress has been made on the following:

- Siphon 0.9 was released in March 2021, prior to the AMS Beginner Python Short Course

External Participation

The Python team attends conferences as well as participates in other projects within the scientific Python ecosystem. This allows us to stay informed and to be able to advocate for our community, as well as keep our community updated on developments. As participants in a broader Open Source software ecosystem, the Python team regularly encounters issues in other projects relevant to our community's needs. As such, we routinely engage these projects to address challenges and submit fixes. We also continue to host Jeff Whittaker's netCDF4-python project repository; Jeff continues to be the active maintainer of the project. The overall involvement helps ensure that important portions of our community's Python stack remain well-supported. Ryan May continues to serve as a core developer for CartoPy as well as a member of Matplotlib's Steering Council, while joining conda-forge's core team.

Progress has been made on the following:

- We continue to engage with the Pangeo project, a grass-roots effort to develop a community stack of tools serving the atmospheric, oceanic, land, and climate science. This engagement is enhanced by work on the Pangeo EarthCube award, which will likely drive some contributions to the XArray project.
- Similar EarthCube efforts see us acting as scientific reviewers for Jupyter notebooks for its annual meeting, as well as joining the Pythia Project.
- Ryan May continues to work as a developer on the matplotlib and CartoPy projects, and has joined the conda-forge core team.
- We also continue to actively engage with the xarray, conda-forge, and pint projects.

Ongoing Activities

We plan to continue the following activities:

- Supporting Unidata's collection of online Python learning materials
- Engaging in synchronous Python teaching opportunities, virtual or otherwise
- Growing Siphon as a tool for remote data access across a variety of services
- Growing and developing MetPy as a community resource for Python in meteorology
- Continued participation in the scientific Python community as advocates for the atmospheric science community
- Working with JupyterHub as a way to facilitate data-proximate analysis
- MetPy Mondays for engaging the community

New Activities

Over the next three months, we plan to organize or take part in the following:

- Deploy a new version of the Python Training website using sustainable technologies
- Offer a redesigned virtual Python training workshop
- Release MetPy 1.1
- Engage in the rollout of Project Pythia and adjacent UCAR Python education efforts

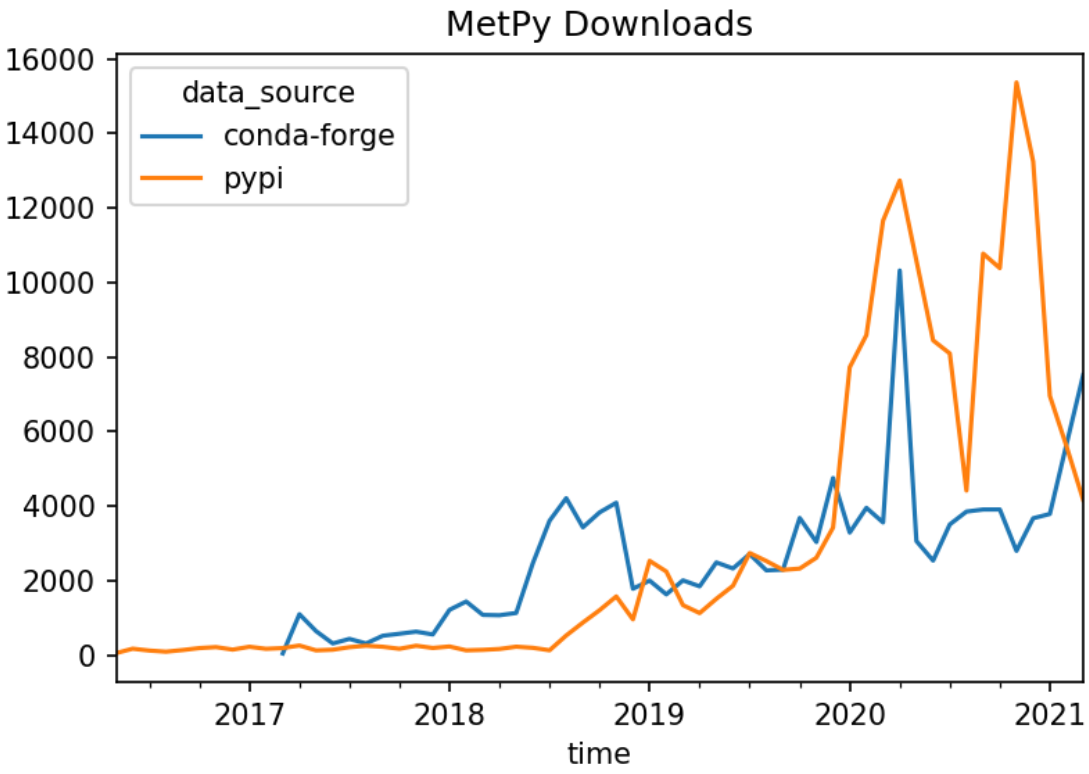
Over the next twelve months, we plan to organize or take part in the following:

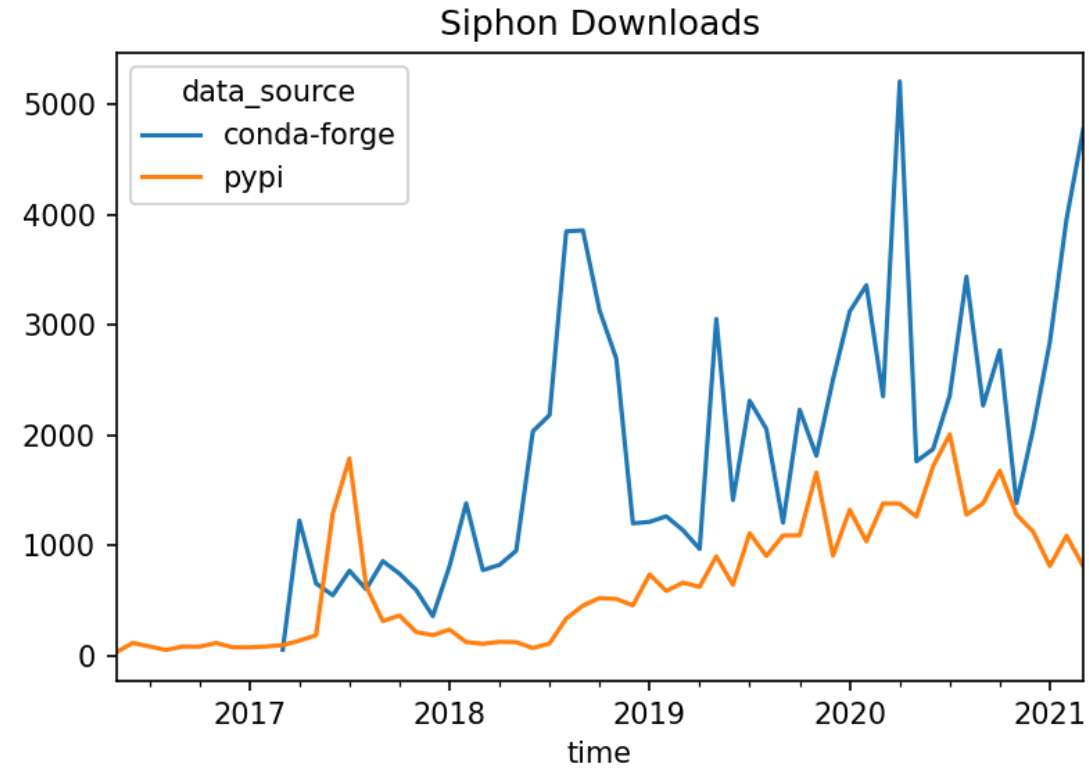
- Teach intermediate MetPy short course at AMS 2022
- Engage in AMS 2022 student conference
- Present annual update on Python libraries at AMS 2022
- Separate non-TDS siphon capability into new MetPy remote functionality

Beyond a one-year timeframe, we plan to organize or take part in the following:

- Evaluate the possibility of extending siphon functionality to interface with the AWIPS-II EDEX server
- Evaluate ways to improve MetPy performance and scalability using tools like Dask and Numba

Relevant Metrics


MetPy Downloads
data_source
conda-forge
pypi
16000
14000
12000
10000
8000
6000
4000
2000
0
2017
2018
2019
2020
2021
time


Siphon Downloads
data_source
conda-forge
pypi
5000
4000
3000
2000
1000
0
2017
2018
2019
2020
2021
time

MetPy

- 96% test coverage
- Watchers: 60
- According to GitHub, 160 repositories and 18 packages depend on MetPy
- Downloads for the releases made in the last year (Conda + PyPI):
    - 0.12.2: 55211
    - 1.0.0rc2: 285
    - 1.0: 27305
- Since 1 October 2020
    - Active Issues: 114 (68 created, 46 closed)
    - Active PRs: 212 (180 created, 169 closed)
    - External Issue Activity: 42 opened, 115 comments
    - External PR Activity: 26 opened, 62 comments
    - Unique external contributors: 48
    - Stars: 97 (722 total)
    - Forks: 2 (274 total)
    - Commits: 409
- Since 1 May 2020
    - Active Issues: 193 (123 created, 83 closed)
    - Active PRs: 328 (279 created, 268 closed)
    - External Issue Activity: 73 opened, 198 comments
    - External PR Activity: 49 opened, 153 comments
    - Unique external contributors: 84
    - Stars: 169 (724 total)
    - Forks: 13 (274 total)
    - Commits: 676

Siphon

- 98% test coverage
- Watchers: 16
- According to GitHub, 75 repositories and 8 packages depend on MetPy
- Downloads for the last year (Conda + PyPI):
    - 0.8.0: 37922
    - 0.9.0: 2915
- Since 1 October 2020
    - Active Issues: 21 (9 created, 6 closed)
    - Active PRs: 50 (47 created, 42 closed)
    - External Issue Activity: 6 opened, 11 comments
    - External PR Activity: 2 opened, 0 comments
    - Unique external contributors: 9
    - Stars: 10 (136 total)
    - Forks: 0 (53 total)
    - Commits: 96
- Since 1 May 2020

- Active Issues: 24 (13 created, 6 closed)
- Active PRs: 61 (50 created, 43 closed)
- External Issue Activity: 9 opened, 18 comments
- External PR Activity: 5 opened, 2 comments
- Unique external contributors: 14
- Stars: 18 (136 total)
- Forks: 1 (53 total)
- Commits: 101

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Providing Useful Tools**
   Python has become a key tool in the atmospheric sciences, and the geosciences in general. MetPy leverages the rest of the scientific Python ecosystem to provide a suite of documented and tested domain-specific functionality, supporting greater use of Python by the community. Siphon serves to provide access to the growing collection of remote data sets. Together, MetPy and Siphon give the community a platform for scripted analysis of real-time and archived weather data. These tools are also readily used in the Jupyter Lab/Notebook environment, for ease of use in cloud and HPC computing environments, facilitating data-proximate analysis. We also participate in a variety of projects in the broader scientific Python ecosystem, to help ensure the ecosystem's viability and that it continues to meet our community's needs.

2. **Supporting People**
   We provide a variety of online training resources to facilitate our community's education and use of Python. We also regularly conduct training workshops to teach attendees how to use tools and apply them to their problems and challenges in research and education.

Prepared *April, 2021*

Status Report: Support

October 2020 - April 2021

Jennifer Oxelson, Tom Yoksas, UPC Staff

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Is the support that we provide sufficient for the community's needs?

   If not, what else should we be doing?

Activities Since the Last Status Report

Training

- Since 2018, the UPC has been focusing its in-person training efforts on regional workshops and short courses.
- Additional resources will be directed towards developing online training materials.
- Unidata recently hired a specialist in online training to help with modernization of our training materials

New Activities

In order to fulfill our objectives articulated in the Unidata 2018 Proposal, focused efforts are needed in two major areas:

- Enhance electronic support offerings
- Create instructional materials for online virtual training

Relevant Metrics

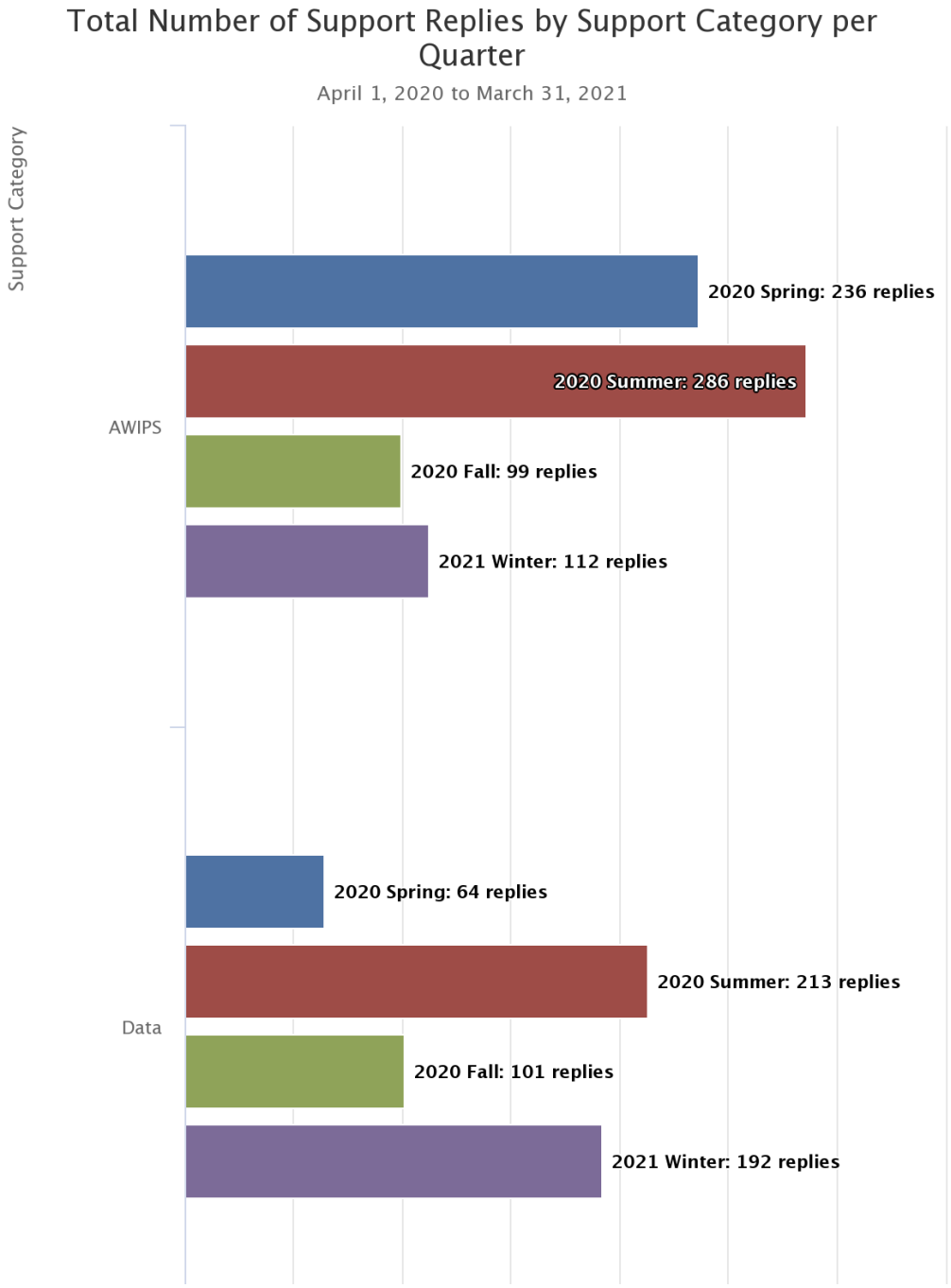
Since January 26, 2006 over 64200 user support "transactions" (new inquiries and follow-ups) have been processed through the Unidata inquiry tracking system. Other methods of providing answers to questions posed (e.g., Github, Stack Overflow, Jira, mailing list replies, etc.) add substantially to the support load.

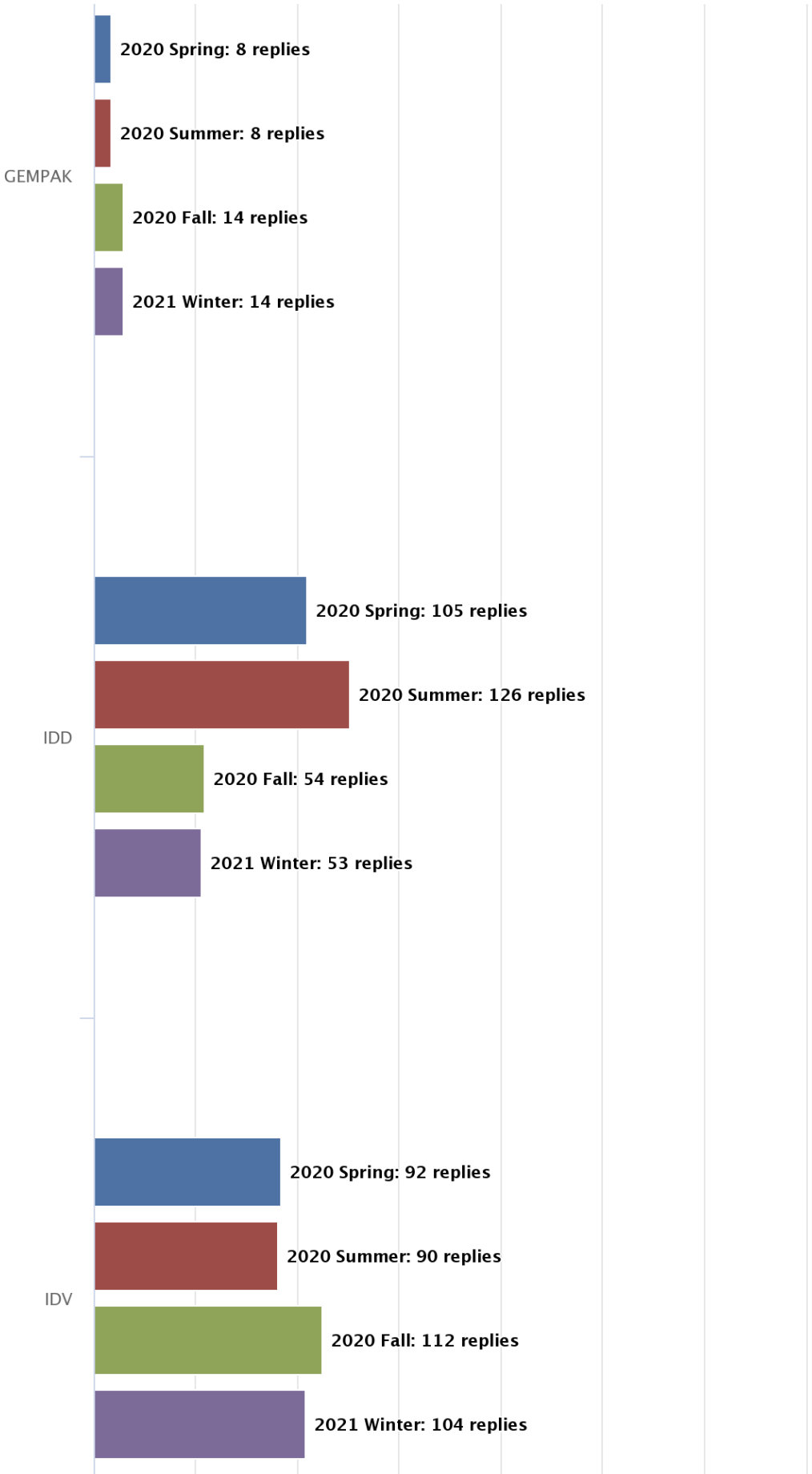
Additional metrics may be found in the Comprehensive Metrics Data portion of this meeting's agenda.

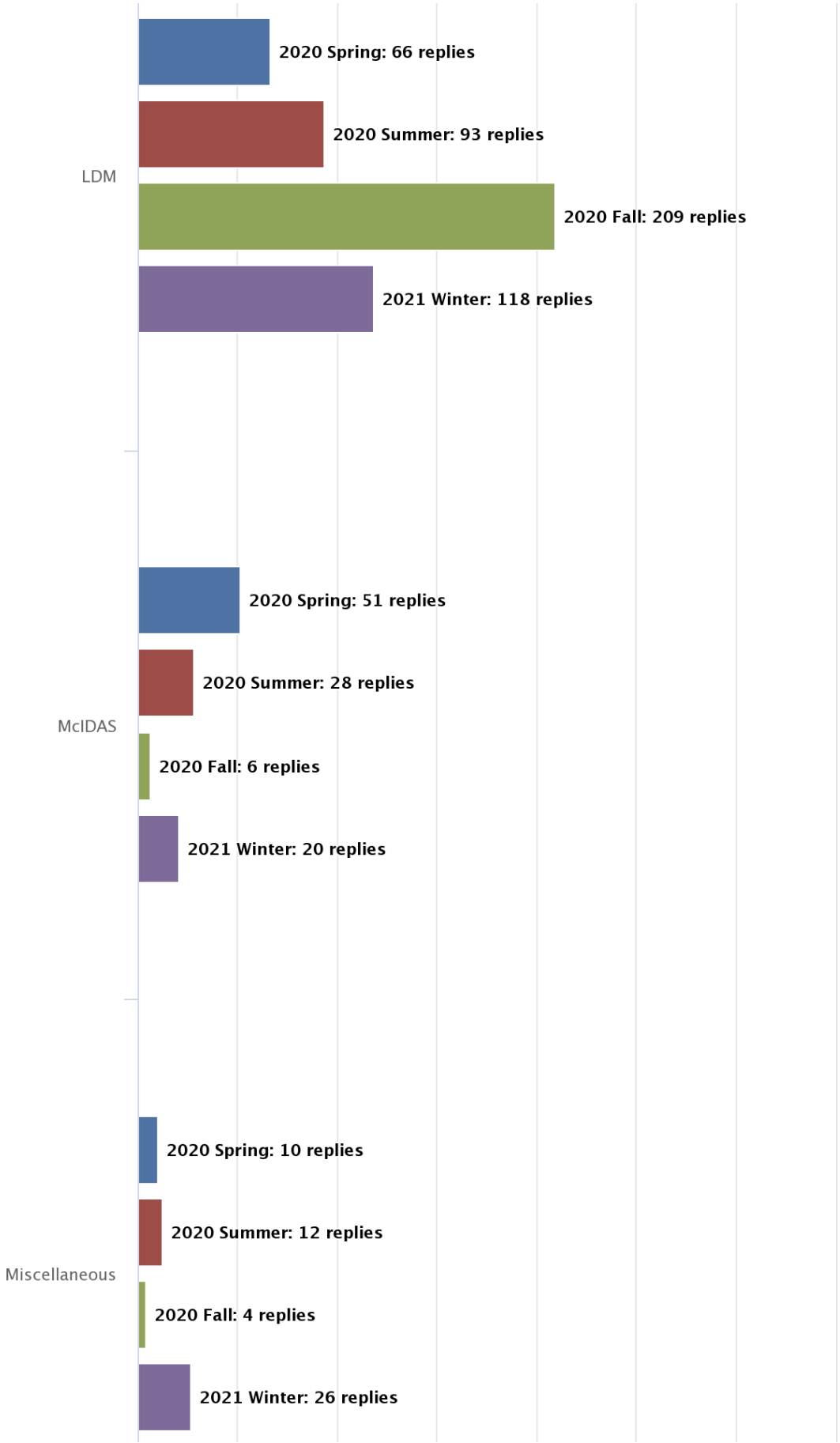
Fig. 1: Below are histograms that portray the number of Unidata email responses for categories of support logged in the Unidata Inquiry Tracking System for the 12 month period from August 1, 2019 until July 31, 2020.

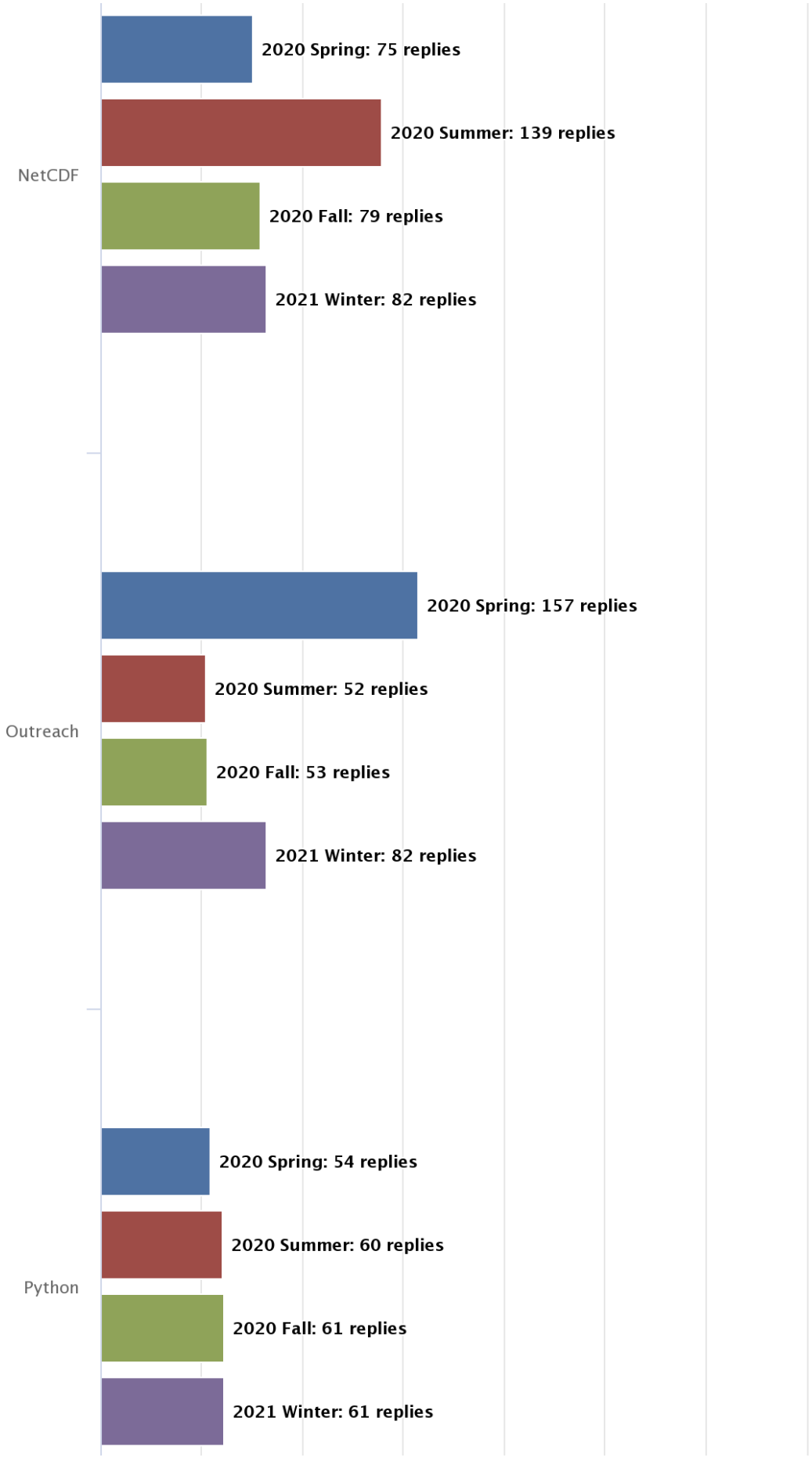
The quarters shown are defined as:

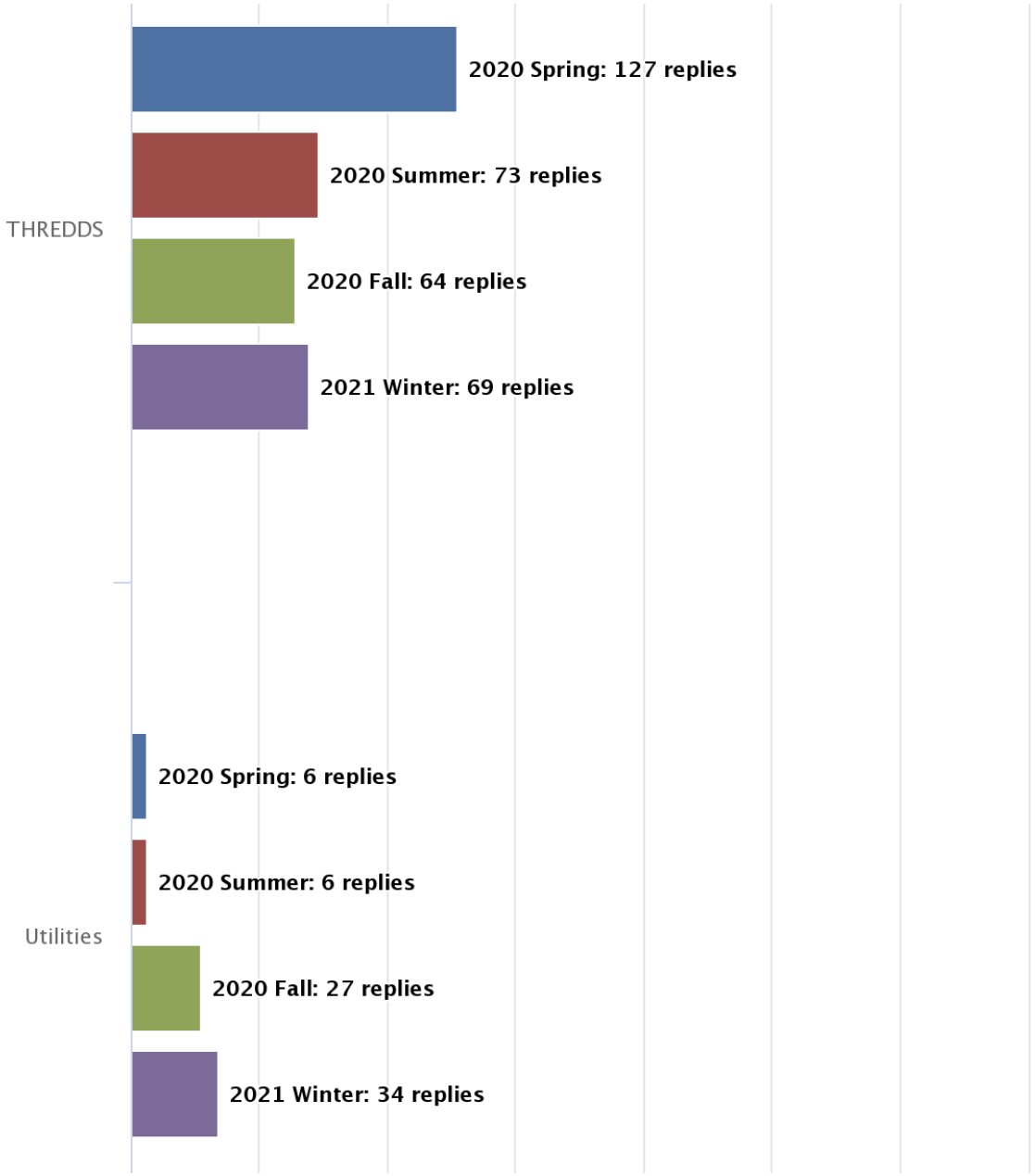
Winter:	**Spring:**	**Summer:**	**Fall:**
January, February, March	April, May, June	July, August, September	October, November, December




GEMPAK
2020 Spring: 8 replies
2020 Summer: 8 replies
2020 Fall: 14 replies
2021 Winter: 14 replies
IDD
2020 Spring: 105 replies
2020 Summer: 126 replies
2020 Fall: 54 replies
2021 Winter: 53 replies
IDV
2020 Spring: 92 replies
2020 Summer: 90 replies
2020 Fall: 112 replies
2021 Winter: 104 replies


LDM
2020 Spring: 66 replies
2020 Summer: 93 replies
2020 Fall: 209 replies
2021 Winter: 118 replies
McIDAS
2020 Spring: 51 replies
2020 Summer: 28 replies
2020 Fall: 6 replies
2021 Winter: 20 replies
Miscellaneous
2020 Spring: 10 replies
2020 Summer: 12 replies
2020 Fall: 4 replies
2021 Winter: 26 replies


NetCDF
2020 Spring: 75 replies
2020 Summer: 139 replies
2020 Fall: 79 replies
2021 Winter: 82 replies
Outreach
2020 Spring: 157 replies
2020 Summer: 52 replies
2020 Fall: 53 replies
2021 Winter: 82 replies
Python
2020 Spring: 54 replies
2020 Summer: 60 replies
2020 Fall: 61 replies
2021 Winter: 61 replies



Individual support activities included in the categories shown above are listed in the following table.

Category	eSupport Departments
AWIPS	Support AWIPS
Data	Support CaseStudy, Support CONDUIT,

	Support Datastream, Support LEAD, Support Level II, Support NOAAPORT, Support SUOMINET
GEMPAK	Support GEMPAK
IDD	Support IDD, Support IDD Antarctica, Support IDD Brasil, Support IDD Cluster, Support IDD SCOOP, Support IDD TIGGE
IDV	Support IDV, Support IDV Storm, Support McV, Support VisAD
LDM	Support LDM
McIDAS	Support McDevelop, Support McIDAS
Miscellaneous	Administration, Development, Plaza, Staging Folder, Support, Support eSupport, Support Miscellaneous, Support Platforms, Support Plaza, Student Interns, Systems
NetCDF	Support LibCF, Support netCDF
Outreach	Outreach, Polcomm, Science Gateway, Support Egrants, Support News, Support Outreach, Support Workshop, Usercomm, Student Interns
Python	Support Python
RAMADDA	Support RAMADDA
THREDDS	Support netCDF Java, Support THREDDS
Utilities	Support LDM-McIDAS, Support netCDF Decoders, Support netCDF Perl, Support OPeNDAP, Support Rosetta, Support UDUNITS

Comments

- The total support provided by the UPC continues to be substantial: yearly totals have shown a slight decline over the past several years, but this is most likely attributable by the increased ways support is being provided. Overall support activities vary by somewhat by quarter. Spikes in support for individual packages is largely correlated with the releases of new distributions of the packages.
- Support for netCDF continues to be substantial, and is understandable given the very

large number of users of the package worldwide.
- Support for the legacy visualization packages GEMPAK and McIDAS has decreased over the past several years, most likely due to GEMPAK users investigations of use of AWIPS and Python/MetPy.
- Support for AWIPS has steadily increased and has exceeded that for GEMPAK over the past couple of years.
- Support for Python scripting using MetPy is growing markedly.
- Support for LDM, IDD, and Data continue at a high levels and show some variability throughout the year.

Notes

These numbers and conclusions should not be taken too literally, for several reasons:

- For some packages, multiple responses in the same thread may be bundled into a single archived email. Other packages have each response in a thread counted separately.
- After a new release of software, there may be a flurry of the same or similar questions, which can be answered in separate emails or in a single mailing list posting.
- The graph primarily represents support of end users and site administrators, not developers. Support for non-Unidata developers in projects such as THREDDS, IDV, GEMPAK, and McIDAS requires significant resources, but is difficult to assess.
- Not all support records were indexable for this report. Given this, the above numbers are an **underestimate** of the actual support being provided by the UPC.

Additional User Support Metrics

Strategic Focus Areas

We support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   Unidata User Support enables access to geoscience data by supporting the use of tools created and/or supported by the UPC.

2. **Providing Useful Tools**
   A significant part of providing useful tools is providing support for those tools. Unidata has always provided world class support for all of the tools that it makes freely available to the greater geoscience community.

3. **Supporting People**
   The user support provided by the UUPC is recognized throughout the atmospheric science community. Unidata's outreach efforts are routinely noted as being exceptional in surveys of the NCAR/UCAR community.

Prepared *April, 2021*

Status Report: THREDDS

October 2020 - April 2021

Sean Arms, Hailey Johnson, Jennifer Oxelson, Ryan May, Ethan Davis, Dennis Heimbigner

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. gCDM (see below)...thoughts?
2. How can we help you and your students? We can do much more than Java programming - we love Python too! Our team comes from a variety of academic backgrounds as well, including Meteorology (boundary-, surface-, and canopy-layer, in-situ observations, radar), Computer Science, Oceanography, Chemistry, and Physics!
3. Several of us across Unidata have gained quite a bit of experience with GitHub Actions over the past year. If you are interested in (free) automated testing of your github-hosted code, but don't quite know how to get start, please let us know!
4. Have you ever used one of the Web Start viewers from the TDS? If so, be aware that as of Java Web Start has been deprecated as of Java 9, and has been removed in Java 11 onward. As such, at some point in 2021 (post TDS v5 release) we are planning to remove all netCDF-Java and TDS related Web Start files from Unidata servers, and the Web Start functionality provided by TDS v4.6 and prior will no longer work. Will this impact your education and research efforts?

Activities Since the Last Status Report

The THREDDS Project

The THREDDS Project encompases four projects: **netCDF-Java, the THREDDS Data Server (TDS), Rosetta, and Siphon** (the Unidata Python client to interact with remote data services, such as those provided by the TDS). For specific information on Siphon, please see the Python Status Report. An update regarding cloud efforts related to the TDS, including the popular Docker container effort, can be found in the Cloud Computing Activities Status Report.

Released netCDF-Java 5.4.1

- NetCDF-Java version 5.0.0 was released on 29 July 2019. NetCDF-Java version 5.4.1 was released on 17 December 2020.
- AWS S3 support was introduced in 5.3.1, so files readable by netCDF-Java can be read directly off of an S3 object store. Since then, general support has been added for accessing data from object storage (those which offer S3 compatible APIs), NcML aggregations, and all of the authentication methods supported by AWS.
- Version 5.4.1 includes the first fruits of efforts to create a well defined public API for the library. A summary of common client changes can be found on the netCDF-java github wiki.

- A roadmap for netCDF-Java has been published on the netCDF-Java github wiki.
- Version 6 of the library is anticipated to be released in beta within the next month. Beta testers welcome!

Released TDS version 4.6.16.1 (Stable)

- TDS version 4.6.16.1 was released on 23 February 2021. This release contains a variety of bug fixes, as well as updates to third-party libraries, including security updates. Version 4.6.16.1 of the TDS is the current stable release, and as such we recommend anyone running a TDS to upgrade to this version

TDS version 5.0.0

- We anticipate releasing a stable version of TDS 5.0.0 this summer.
- General object store support (AWS S3 compatible API) has been extended to the TDS. New capabilities have been extended to the TDS since our last report, such as datasetScan and NcML aggregation support for object stores.
- In partnership with NASA JPL, initial TDS benchmarks show that the move from local disk based data holdings to Object Store based data holdings is mostly a horizontal move in terms of performance from an end users perspective.
- We also ran benchmarks against the TDS with data stored both on disk and in an object store as part of a study by CLS (Collecte Localisation Satellites - a French earth observing group). We worked with NCAR/CISL to use their on-prem ActiveScale object store.

Documentation for netCDF-Java / TDS (Beta) v5

- The Java-based THREDDS projects have changed significantly over the past few years, and have not been migrated to use a Jekyll-based documentation theme specifically designed for Unidata documentation (generic preview). We hope the new documentation system provides for a consistent set of docs, and facilitates documentation contributions from users. The new netCDF-Java documentation can be found at https://docs.unidata.ucar.edu/netcdf-java/current/userguide/index.html, and the new THREDDS Data Server documentation can be found at https://docs.unidata.ucar.edu/tds/5.0/userguide/index.html.

Rosetta

Rosetta continues to progress following a very successful NASA ACCESS grant (the Oceanographic In-situ data Interoperability Project, or **OIIP**), in which Unidata partnered with the PO.DAAC at JPL and UMASS-Boston.

Progress has been made on the following:

- Server technology requires continual bedrock level work, most of which for Rosetta has been focused on upgrading 3rd party libraries to keep up with security patches.

While this type of work does not result in new features, it's the kind of ongoing work any web-based service needs to remain secure and usable into the future.
- We recently became aware that a customized version of Rosetta was in use at the Nansen Environmental and Remote Sensing Center, which is one of four partners comprising the Bjerknes Center for Climate Research in Bergen, Norway. Founded in 1986, NERSC is a non-profit research foundation under which interdisciplinary research within earth science systems focused on the High North and the Arctic is pursued to investigate and map marine, cryospheric and atmospheric conditions.

gCDM (gRPC for the Common Data Model)

- gCDM is a new effort currently being developed in the _develop_ branch of netCDF-Java, and is a new way to allow netCDF-Java to communicate remotely (although it goes well beyond this).
- gCDM stands for "gRPC Cdm", where gRPC is a recursive acronym that stands for "gRPC Remote Procedure Calls". For more information on gRPC, checkout the gRPC FAQ.
- RPC is much different than REST or OpenAPI when it comes to API design. As a user of RPC, you never ask the question "what does the URL structure look like?" or "do I need to use an HTTP PUT request or POST request? What about GET?"
- gCDM is based around a set of services and messages which are defined in a language independent set of Protocol Buffers definition files.
- The proto compiler tool can generate both client and server code referred to as "stubs" in multiple languages. This generated code handles all of the HTTP communication aspects of the client and server interaction, including serialization and deserialization of data.
- gRPC supports bi-directional streaming with http/2 based transport. It uses a highly efficient, binary-based on the wire transport format (no json, no xml, etc.).
- The gCDM project in the netCDF-Java repository includes a java-based client and server. The gCDM server functionality essentially replaces the functionality of all of the data services of the TDS (array level access, coordinate access, subsetting), but in 300 lines of non-generated code.
- In collaboration with NASA JPL (shoutout to Vardis Tsontos and Joe Roberts), we've created a simple python and javascript example demonstrating what it is like to work with the generated code (functionally equivalent, mostly print statements):
    - python (output)
    - javascript (output)

  Note that ideally a library like Siphon would provide "the final foot" for the generated stub code, and not individual users.
- You can spin up a local gCDM server by double-clicking a jar file. No configuration. At that point, you could read NEXRAD level 2 files off your local disk in your browser
- Because gRPC is bi-directional, this will allow not only non-Java based code to take advantage of the CDM, but will also allow the CDM to take advantage of non-Java code.
- Many gCDM possibilities: highly efficient bi-directional communications between geoscience focused microservices, chains of serverless functions, and "traditional" stand-alone servers; a cross kernel communications layer for geoscience data access in

Jupyter; new browser-based and mobile app functionality (without JSON/XML/etc overhead); configurless local server for accessing data via CDM in non-java based language.

General dependencies, challenges, problems, and risks include:

- Calling all beta testers! The goal of beta testing TDS 5 is to ensure that the current capabilities of 4.6.x are working in the new version (and if some bugs get fixed in the process, even better!). Beta testing by our users is critical, and so far we have had several community members offer their help (special thanks to Rich Signell, Peter Pokrandt, Victor Gensini, the NCAR RDA, etc.!). Beta 9 will be the final beta before the TDS 5 release candidate.

Ongoing Activities

We plan to continue the following activities:

- Maintain thredds.ucar.edu and keep up with the addition of new datasets to the IDD.
- Closely monitor the security status of our project dependencies, and provide updated versions of our libraries and server technologies to address as needed.
- Clearly define the public API of netCDF-Java.
    - Wrap-up initial work on version 6 of netCDF-Java has started. Much of the previously deprecated code has been removed, and only the public API is used.
    - Work will soon begin on version 7 of netCDF-Java. We will begin using Java 11 features, which means Java 11 or higher will be required.
    - Many thanks to John Caron, who has been making the vast majority of this work happen. You likely remember John (or have heard his name) from his many dedicated years on staff at Unidata as the lead of the THREDDS projects since their inception in the late 90s. John left Unidata in 2015, and has remained a critical and active part of the community since then, even while holding a full time position at Google. John's talent, insights, and mentorship have been greatly appreciated and cannot be overstated.

The following active proposals directly involve THREDDS work:

- USGS has awarded Unidata a project with the goal of improving the TDS's ability to leverage cloud computing technologies (like object stores) to serve very large datasets. The award will support just over 0.5 FTE over the next year. The three focus areas are 1) Zarr support in netCDF-Java and TDS; 2) add a TDS Service API to simplify adding new data services to the TDS; and 3) add support for the OGC EDR (Environmental Data Retrieval) API using the new TDS Service API.
- We are in our final six months of the NOAA IOOS award titled "A Unified Framework for IOOS Model Data Access", in which we are partnered with Rich Signell and Axiom Data Science. The goal is to enable support of the UGRID specification within the THREDDS stack, as well as create a GRID featureType to allow for serving large collections of gridded datasets (including UGRID). This work **stragiticialy aligns with

the Unidata 2024 focus area “Managing Geoscience Data, Making Geoscience Data Accessible** by improving the reliability and scalability of the TDS to handle very large collections of gridded datasets, as well as **“Managing Geoscience Data, Enhancing Community Access to Data”** through the addition of UGRID support (example: MPAS output is on a mesh, a.k.a. “unstructured”, grid).

- Unidata has been included as a subcontractor on the most recent phase of the CEOS Ocean Variables Enabling Research and Applications for GEO (COVERAGE) initiative. COVERAGE is a NASA-led research and development project and cross-cutting, collaborative effort within the Committee on Earth Observation Satellites (CEOS). The goal for the THREDDS project work is to extend the NCSS service of the TDS to facilitate matchup services between gridded and point datasets for data stored in traditional and cloud based (AWS S3, WEkEO) environments, and enabling a coverageJSON-based response from NCSS. This work **stragiticialy aligns with the Unidata 2024 focus areas “Managing Geoscience Data, Making Geoscience Data Accessible and Making Geoscience Data Usable** and **“Managing Geoscience Data, Enhancing Community Access to Data”**

New Activities

Over the next three months, we plan to organize or take part in the following:

- Community
    - Developer Hailey Johnson will be acting as a mentor for one of the three Unidata Summer Interns!
- netCDF-Java
    - Release Initial support for reading Zarr and ncZarr.
    - Begin beta testing of netCDF-Java 6.
    - Begin to modularize (Java Platform Module System) while maintaining Java 8 compatibility.
    - Commit to semantic versioning.
    - Better curate existing documentation into four documentation sets: tutorial, developer/reference (nitty-gritty details, for those interested in learning more or hacking on the netCDF-java codebase), THREDDS Client Catalogs (language agnostic), and NetCDF Markup Language (NcML) (language agnostic).
    - Continue experimenting with gCDM.

- TDS
    - Getting TDS v5.0 to a stable release (release candidate targeted for spring 2021).
    - Better curate existing documentation into four documentation sets: server administrator (with quick start guide), end user (browser), developer (web access via api), reference (nitty-gritty details, for those interested in learning more or hacking on the TDS codebase).

Over the next twelve months, we plan to organize or take part in the following:

- netCDF-Java

    - Obtain 90%+ test coverage of the public API in version 6+.
    - Initial support for writing Zarr and ncZarr.
    - Complete command line tool creating a WRF intermediate file from a subsetted GRIB dataset.
    - Break out language agnostic schema and documentation (THREDDS Client Catalogs, NcML) into new, standalone repositories.
- TDS
    - Release TDS version 5 (Stable)
    - Create a TDS Registry
    - Implement option to create WRF intermediate files from GRIB datasets via TDS user interface. Support storage of pre-defined dataset variables for ease of WRF file recreation.

Beyond a one-year timeframe, we plan to organize or take part in the following:

- netCDF-Java
    - Fully support Java 11 and the Java Platform Module System (end of Java 8 support)
- TDS
    - Re-evaluate new development of the TDS in light of gCDM.

Relevant Metrics

NetCDF-Java

End of 12 month period	Number of Downloads (cdm/cdm-core.jar, netcdfall.jar, toolsui.jar)
2020-02-29	36,897
2020-07-31	49,181
2021-03-31	104,516

THREDDS Data Server

Downloads

End of 12 month period	Number of Downloads (thredds.war, thredds-classes.jar)
2020-03-31	2,218

Startup Metrics

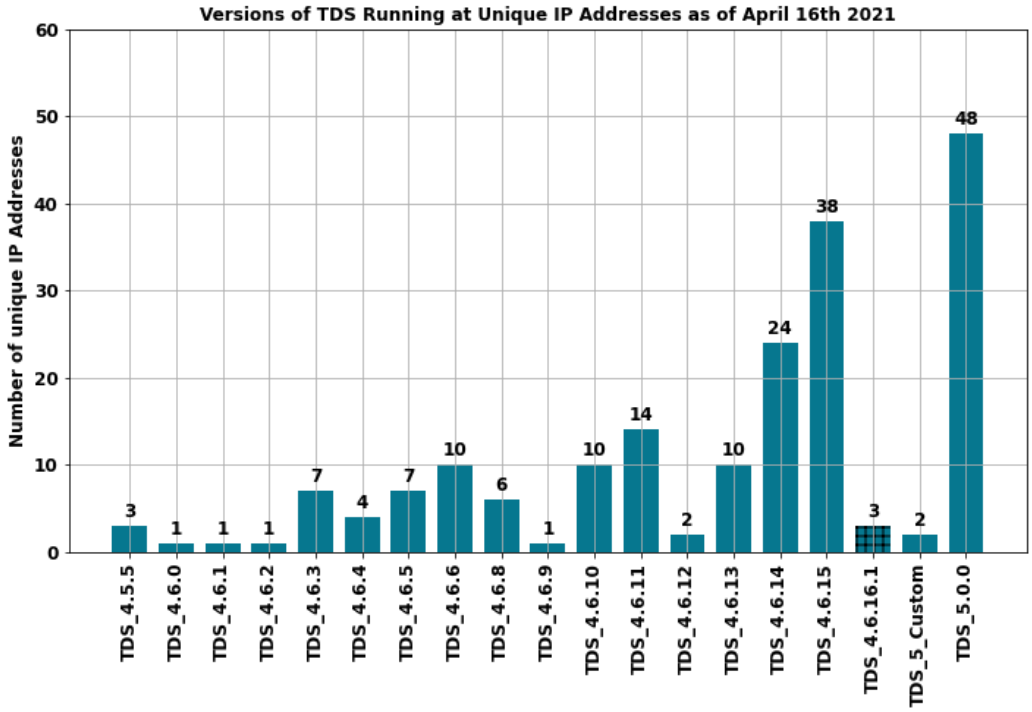
	2020-08 — 2021-04	2014-08 — 2021-04
TDS Startup (unique IP address count)	4,054	34,164

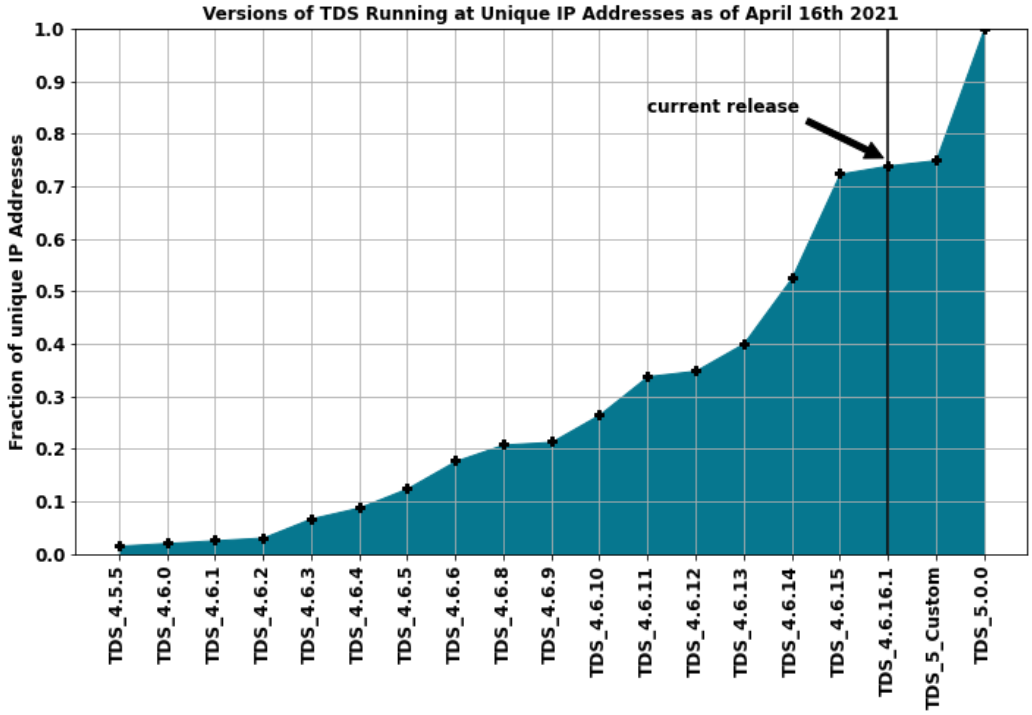
	Total Servers	Information page updated
Publicly Accessible[1] TDS count	192	99

Over the past six months, **4,054** unique IPs started up the TDS (August 2020 through March 2021). Since we've started tracking these metrics (v4.5.3, August 26th, 2014), we've seen the TDS startup from **34,164** unique IP addresses. There are currently **192** publically accessible TDSs running "in the wild" (the same as our last report) . Furthermore, of the **192** publically accessible servers, **99** have updated the name of their server in their server configuration file (taken as a sign that they are maybe, possibly, intended to be used by others...maybe…).

The figures below show the distribution of TDS versions (top), and the fractional share of servers running version X or older (bottom). Each labeled version includes betas and snapshots, not just the official release of that version, for presentation simplicity. The majority of the publicly accessible servers are running v4.6.13 or above (v4.6.16.1 was the most current release during this period, and was released on 23 February 2021). TDS v5 is the dominant specific version running in the wild.

[1] "Publicly accessible" means we could find a top-level THREDDS Client Catalog. We checked <server>/thredds/catalog.xml (version 4), <server>/thredds/catalog/catalog.xml (version 5), including the most common ports of 80, 8080, 443, and 8443.





In the next year, we will be working towards enabling TDSs, on an opt-in basis, to officially advertise their availability to the community through a centralized resource.

Strategic Focus Areas

The THREDDS projects covered in this report support the following goals described in Unidata Strategic Plan:

1. **Managing Geoscience Data**
   The component software projects of the THREDDS project work to facilitate the management of geoscience data from four points of view: __Making Geoscience Data Accessible, Making Geoscience Data Discoverable, Making Geoscience Data Usable, and Enhancing Community Access to Data__ . As a client-side library, **netCDF-Java** enables end users to read a variety of data formats both locally and across numerous remote technologies. Less user-friendly formats, such as GRIB, are augmented with metadata from community driven metadata standards (e.g. Climate and Forecast metadata standards), and viewed through the more user friendly Common Data Model (very similar to the netCDF Data Model), providing a single set of Java APIs for interacting with a multitude of formats and standards. The **THREDDS Data Server** exposes the power of the netCDF-java library outside of the Java ecosystem with the addition of remote data services, such as __OPeNDAP__ , __cdmremote__ , __OGC WCS__ and __WMS__ , __HTTP direct download__ , and other remote data access and subsetting protocols. The TDS also exposes metadata in standard ways (e.g. ISO 19115 metadata records, json-ld metadata following schema.org), which are used to drive search technologies. **Rosetta** facilitates the process of translating ascii based observational data into standards compliant, archive ready files. These files are easily read into netCDF-Java and can be served to a broader community using the TDS.
2. **Providing Useful Tools**
   Through Rosetta, the THREDDS project seeks to intercede in the in-situ based observational data management lifecycle as soon as possible. This is done by enabling those who produce the data to create archive ready datasets as soon as data are collected from a sensor or platform without the need to write code or intimately understand metadata standards. NetCDF-java and the TDS continue to support legacy workflows by maintaining support for legacy data formats and decades old data access services, while promoting 21st century scientific workflows through the creation of new capabilities (such as adding Zarr support) and services.
3. **Supporting People**
   Outside of writing code, the THREDDS project seeks to support the community by __providing technical support, working to build capacity through Open Source Software development, and by building community cyber-literacy__ . The team provides expert assistance on software, data, and technical issues through numerous avenues, including participation in community mailing lists, providing developer guidance on our GitHub repositories, and leading and participating in workshops across the community. The team also actively participates in “upstream” open source projects in an effort to help sustain the efforts of which we rely and build upon. We have mentored students as part of the Unidata Summer Internship Program, and worked across organizations and disciplines in support of their internship efforts.