Users Committee/Strategic Advisory Committee

Joint Meeting Agenda

Tuesday, October 1, 2019 (Users Committee)

9:00 - 9:45 New Members Orientation (Josh Young/Staff)

10:00 - 11:00 Break (and Video Recording at Center Green 1 for some)

Returning Members Convene by 11AM

10:45 - 11:00 Break/Full Committee Convenes

11:00 - 11:15 Welcome and Introduction (Mohan Ramamurthy/Chair)

11:15 - 11: 30 Administrative Items (Actions, Next Meeting Dates) (Chair/Josh)

11:30 - 11:45 Equipment Awards

11:45 - 12:00 RT Stats (Daryl Herzmann)

12:00 - 1:00 Lunch

1:00 - 1:30 Users Workshop Regional Follow Up Events (Casey Davenport, John Allen, Shawn Riley, and Josh Young)

1:30 - 1:45 DeSouza Award Text (John Allen)

1:45 - 2:30 Around the Table Reports (Committee)

2:30 - 3:15 Status Reports (Committee/Staff)

3:15 - 3:45 All Other Business/Users Committee Session Adjourn (Josh)

3:45 - 5:00 Developers Available

Wednesday October 2, 2019 (Joint Committee)

9:00 - 10:15 Director's Report (Mohan Ramamurthy)

10:15 - 11:00 Users Workshop Regional Follow Up Events

11:00 - 11:10 Break

11:10 - 11:45 MSI Engagement Strategy (Josh Young and Mohan Ramamurthy)

11:45 - 1:00 Lunch

1:00 - 1:30 NOAA Update - Scott Jacobs

1:30 - 2:00 NASA Update - Chris Lynnes

2:00 - 2:35 Science Gateway Capacity Test (Mohan Ramamurthy and Julien Chastang)

2:35 - 2:45 Group Photograph (location TBA)

2:45 - 2:55 Break

2:55 - 3:10 New UCP Structure (Mohan Ramamurthy)

3:10 - 3:45 NOAA Data Updates - Carissa Klemmer

3:45 - 4:00 Transition to Auditorium

4:00 - 5:00 DeSouza Award Presentation entitled "You can't spell UNIDATA without Data" by 2019 Award Recipient Pete Pokrandt

6:30 Collaborative discussion on the day's proceedings over dinner at the Med, 1002 Walnut St, Boulder, CO 80302 (map)

Thursday, October 3, 2019 (Strategic Advisory Committee)

9:00 - 9:15 Welcome and Administrative Items (Chair/Josh)

- Spring meeting date and location

9:15 - 10:00 Budget Report (Terry Mitchell-Sur)

10:00 - 10:30 NSF Report (Bernard Grant)

10:30 - 10:50 Business Development Update (Josh Young)

10:50 - 11:00 Break

11:00 - Noon Review of preceding day's activities/Discussion (Mohan Ramamurthy)

Noon - 1:00 All Other Business/Lunch

1:00 Committee Member Video Recording at Center Green 1

Status Report: Strategic Advisory Committee Actions

April 2019 - October 2019

Unidata Program Center Staff

Actions from the Previous Meeting

Action 1

(Josh) Update SAC following the Saudi Arabia and Met Tech World Expo meetings

Result

An update was sent following the visit to Saudi Arabia with the intention to do the same after the Met Tech World Expo; however, that activity was overcome by other events

Action 2

(Doug) Coordinate youtube committee videos during the Joint Session

Result

Scheduled as part of this Joint Session

Action 3

(Josh) Follow up with committee after Mind the Gap about the idea of an industry panel at either joint session or workshop follow up

Result

Conversation with Chairs held following Mind the Gap prior to planning this conversation; not explicitly discussed though more appropriate for regional follow ups pending their final structure

Action 4

(Josh & Committee Chairs) Invite an education researcher as speaker at either the Joint Session or workshop follow ups

Result

Conversation with Chairs held following Mind the Gap prior to planning this conversation; not explicitly discussed though more appropriate for regional follow ups pending their final

structure

Action 5

(Ethan) Discuss the possibility of Data Product training with Chris Lynnes at ESIP Summer Meeting

Result

Ongoing

Prepared *September 2019*

Status Report: Users Committee Actions

April 2019 - October 2019

Unidata Program Center Staff

Actions from the Previous Meeting

Action 1

New DeSouza award plaque in conference room (Josh/Doug)

Result

Complete

Action 2

Revise DeSouza awards nomination description based on the ideals of the Unidata community instead of the current text. (John Allen)

Result

Underway update scheduled as part of Joint Session

Action 3

Create a joint-subcommittee to generate outreach to nominee (Alex, Kevin)

Result

Complete

Action 4

Identify committee volunteers to partner on planning follow up workshop (John, Casey, Shaun)

Result

Committee volunteers and staff had a call between meetings to identify a path forward for further discussion during Joint Session

Action 5

15 minutes in the fall agenda to figure out the future for RTSTATS and ID3V (searchable database of what comes over IDD) (Daryl)

Result

Scheduled as part of Joint Session

Action 6

Blog post from Rich S on running WRF in the cloud (Rich/DougD).

Result

Postponed, still possible

Action 7

Try and work to get more oceanographic data into IDD (Alex Davies)

Result

Ongoing

Action 8

Notify Pete of award (Kevin)

Result

Complete

Action 9

30 minutes sessions on NASA data approach (Chris L/Josh)

Result

Requested and scheduled as part of Joint Session

Status Report: AWIPS and GEMPAK

April 2019 - October 2019

As you are aware, UPC developer Michael James passed away suddenly in late June 2019. The UPC had already hired Shay Carter as a second AWIPS developer, with the expectation that Michael would serve as her mentor as she gained experience with the package. Shay joined the UPC at the beginning of September, and is in the process of coming up to speed with the AWIPS code base.

In an effort to continue to provide the university AWIPS community with seamless support, we have sought the cooperation of several experienced AWIPS developers from NOAA ESRL, NCEP, and elsewhere. It is our hope that we can call on these developers to get more immediate feedback and answers for users who are experiencing difficulties, while Shay learns the ins and outs of AWIPS. The group is also 0 intended to be a bit of a backup mechanism for her if she runs into problems or questions while she is teaching herself about AWIPS and beginning to develop the software.

Activities Since the Last Status Report

AWIPS

Unidata's Jetstream EDEX server continues to serve real-time AWIPS 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.

Software Releases

18.1.1-6 March 2019

AWIPS 18.1.1-6 is a minor update.

- Updates to Surface, NCEP/Hydro, Upper Air menus and bundles, including return of SIGMET/AIRMET menu.
- Fix for the reprojection of MetarSurfaceTemp data.
- Fix coloring of MetarSurfaceTemp data to avoid rendering of point data as

black-on-black.
- Added RTOFS to LDM/EDEX, as well as CAVE menus and data bundles.
- Linux-specific fix for CAVE to allow for RPM updates.
- Linux-specific EDEX dependency updates to include net-tools.
- LDM updated to version 6.13.10.

GEMPAK/NAWIPS

GEMPAK 7.5.1, released in February 2019, is the most recent Unidata release.

https://github.com/Unidata/gempak/releases/tag/7.5.1

Activities Ongoing/In-Progress

AWIPS and GEMPAK development activities are currently in a state of flux. Unidata Program Center staff are currently:

- Helping Shay familiarize herself with the AWIPS code base.
- Seeking to hire an AWIPS technical lead.
- Striving to answer AWIPS support questions from the community.

Future Activities

Future plans are currently in a state of flux. We expect the situation to begin to resolve as Shay comes up to speed and we hire a second developer to enhance the AWIPS team.

Metrics

Downloads March - September 2019

AWIPS downloads: 2975
GEMPAK downloads: 599

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**
   Both AWIPS and GEMPAK are freely available, and both incorporate LDM/IDD technology for accessing geoscience data.

3. **Supporting People**
   At this juncture, we are providing a limited amount of technical support for the community, and encouraging community members to assist each other through Unidata-managed forums.

Prepared *September 2019*

Status Report: Cloud Computing Activities

April 2019 - September 2019

Julien Chastang, Ward Fisher, Ryan May, Jen Oxelson, Mohan Ramamurthy, Jeff Weber, Tom Yoksas

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. We are planning on forming a Unidata Science Gateway advisory panel. Are you interested in participating?
2. Have you had a chance to experiment with the Unidata JupyterHub server?
3. Do you need a Unidata hosted JupyterHub for your classroom or workshop use?
4. What new cloud technologies are our community members using and investigating on their own initiatives?
5. 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

Unidata JupyterHub Demonstration Server

In collaboration with the eXtreme Science and Engineering Discovery Environment (XSEDE) Extended Collaborative Support Services (ECSS) team and Indiana University (IU), Unidata continues to enhance the Unidata JupyterHub demonstration server.

After a lengthy summer outage the Unidata JupyterHub is running and available once again. This JupyterHub server was previously deployed at the Jetstream TACC (Texas Advanced Computing Center at the University of Texas) data center. We encountered numerous problems at TACC that led to the JupyterHub not working properly and was therefore unavailable for a few weeks. We moved the JupyterHub to the IU data center where it appears to be functioning more reliably.

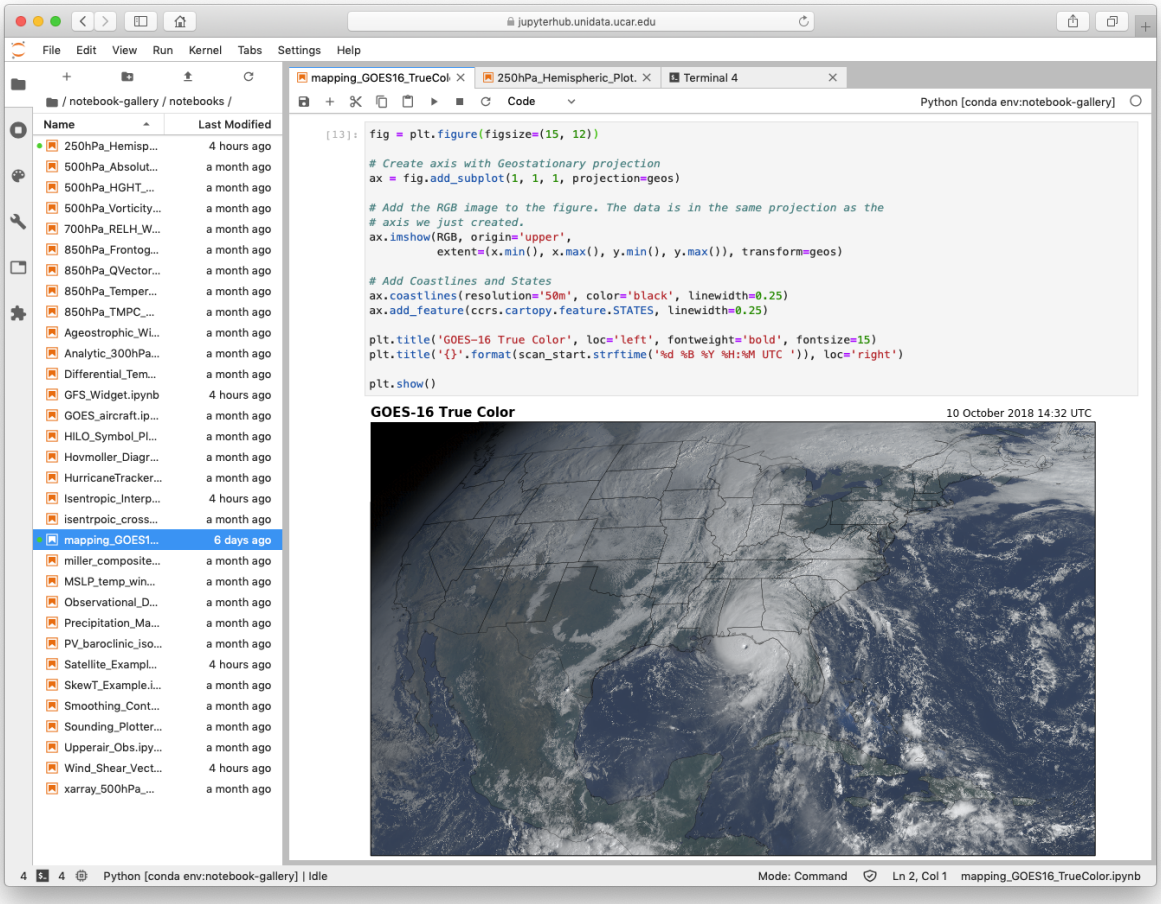
In addition, this JupyterHub server has been updated with the latest stable versions of JupyterHub and JupyterLab. We also augmented the disk quota available to each user. It was previously at 1 GB, now it is 10GB. We have added cluster monitoring software with Grafana and Prometheus.

We also implemented an important usability improvement where users are now dropped into the correct kernel when accessing one of the pre-loaded notebooks from the Gallery, Online Training or Workshop projects.

We are seeing quite a bit of word-of-mouth growth of this server. Before the problems at TACC, we were up to 60 users who had tried the JupyterHub server at one time, with a smaller

subset using the server regularly. Since the move to IU, we are back up to 29 users.

We have also been working with Ben Schenkel (Research Scientist, University of Oklahoma, Cooperative Institute for Mesoscale Meteorological Studies) who has been providing us feedback for this JupyterHub server. He is directing his NSF REU students to use this solution because it requires no installation of local software.



Unidata Python Gallery GOES-16 true color notebook by Brian Blaylock

JupyterHub for Southern Arkansas University (SAU)

In collaboration with Keith Maull (UCP), we have deployed a JupyterHub for a data science class at SAU. Keith has provided invaluable feedback for previous JupyterHub deployments both at SAU and for the UCAR SOARS Summer Internship program.

JupyterHub for UCAR SOARS Summer Internship Program

In collaboration with Keith Maull (UCP), Unidata deployed a JupyterHub for the UCAR SOARS Summer Internship program. Unfortunately, this JupyterHub suffered from some of the same

problems described previously at the TACC data center. We expect to have a more successful result next summer at the IU data center.

JupyterHub for AMS 2020

In preparation for a Unidata PyAOS (Python in the Atmospheric and Oceanic Sciences) demonstration at the Winter 2020 AMS involving an audience of approximately 200 participants, we have begun experimenting with a JupyterHub that can accommodate a large number of workshop attendees.

ESIP 2019 Summer Meeting

Presented Deploying a Unidata JupyterHub on the NSF Jetstream Cloud, Lessons Learned and Challenges Going Forward.

Successfully Awarded New Research Allocation on NSF Jetstream Cloud

Unidata successfully requested and obtained a renewal to our Jetstream allocation worth $530,000:

- Awarded Resources: IU/TACC (Jetstream): 3,560,940.0 SUs
- U/TACC Storage (Jetstream Storage): 40,000.0 GB
- TACC Data Analytics System (Wrangler): 8,640.0 Node Hours
- TACC Long-term Storage (Wrangler Storage): 40,000.0 GB
- The Science Gateways Community Institute (SGCI) support
- XSEDE Extended Collaborative Support Services (ECSS)

Science Gateway News Feed

There is now a Unidata Science Gateway news feed to inform users of gateway developments.

Ongoing Activities

NOAA Big Data Project

- In collaboration with Unidata, NOAA is delivering 20+ years of NEXRAD Level II data via Amazon S3 and the NSF Jetstream cloud. LDM and THREDDS Data Server (TDS) THREDDS Docker software are being 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.
- Unidata continues to get requests from other UCAR/NCAR groups, to partner and lend assistance in cloud computing, especially in the AWS cloud.

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 IDV, 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 (which the THREDDS and RAMADDA containers depend on) has been enhanced to accept "relaxed characters" for compatibility with OpenDAP.
- Revamped THREDDS Docker container README.
- Released THREDDS Docker containers corresponding to 4.6.14 of the TDS and 5.0 beta 7.
- Released LDM Docker container corresponding to the 6.13.11 version of the LDM.

Product Generation for IDD

For the past four 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 (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 work using Jetstream to develop cloud-deployable AWIPS instances, both as imaged virtual machines (VMIs) available to users of Atmosphere and OpenStack, and as docker containers available on Docker Hub and deployable with the xsede-jetstream toolset.

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.

New Activities

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

Forthcoming Conference Attendance

- Science Gateways and eScience 2019 Conferences | September 23-27, San Diego, CA
- 2020 AMS Annual Meeting | January 12-16, Boston, MA

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

XSEDE ECSS Jetstream JupyterHub Collaboration

We plan to continue our collaboration with Andrea Zonca (XSEDE ECSS, San Diego Supercomputing Center) to implement an autoscaling solution for the Unidata JupyterHub. Currently, the Unidata JupyterHubs suffer from the problem that only "manual" scaling is available to accommodate additional users. We are collaborating with Andrea to develop an autoscaling solution that can automatically add more cluster nodes as more users come online and remove nodes when they are no longer needed.

Relevant Metrics

Github Statistics

	Watches	Stars	Forks	Open Issues	Closed Issues	Open PRs	Closed PRs
xsede-jetstream	5	8	7	4	146	2	309
tomcat-docker	8	29	32	2	28	0	39
thredds-docker	10	11	15	1	103	0	124
ramadda-docker	3	0	1	1	10	0	15

ldm-docker	7	7	8	0	30	0	43
tdm-docker	4	2	5	1	9	0	11

JupyterHub User Statistics

- Unidata JupyterHub demonstration server has **29 users**.
- Unidata JupyterHub for Southern Arkansas University has **10 users**.

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 with 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.

Prepared *September 2019*

Status Report: Community Services

April 2019 - October 2019

Doug Dirks, Jeff Weber, Joshua Young

Areas for Committee Feedback

We are requesting your feedback on the following topics:

All questions are featured as agenda topics

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:

- High Plains Regional Climate Center Inaugurates new THREDDS Data Server
- What's New in IDV 5.6
- Unidata Funding Proposal Approved by National Science Foundation
- NetCDF and Native Cloud Storage Access via Zarr
- Meet Unidata's 2019 Summer Interns
- LDM data feeds, THREDDS server now available from UW-Madison AOS Department
- Unidata Program Center Welcomes Zach Bruick
- Fresh IDV Videos on YouTube
- Summer 2019 Unidata Interns Wrap Up Their Projects
- Unidata Program Center Welcomes Shay Carter
- 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:

- Engagement with CUAHSI to support the NFIE and WRF-Hydro at the NWC.
- Assisting the EarthCube Science Support Office staff with back-office tasks.
- Continue to serve on the CUAHSI HIS standing committee.
- We continue to update Unidata's social media channels (Facebook, Twitter, Google+).
- We continue to publish short videos/screencasts on the Unidata YouTube channel.
- 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.

Ongoing Activities

We plan to continue the following activities:

- Event planning activities for the Users Workshop follow-up (date and topic TBD)
- 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 allows
- Engage other UCAR/NCAR divisions regarding Unidata software use i.e. CESM/IDV
- Active participation in CUAHSI HIS (Hydrologic Information System)
- Artive 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
- Development of a Users Committee workshop follow-on event

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
- Seek additional opportunities to engage and listen to the community

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

- Provide additional data management and 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:
- 51,066 unique pageviews (54,596 in previous period)
- 10.5% of total unique pageviews (10.3% in previous period)

Top community pages

1. All blog pages
   42256 unique pageviews (42529 in previous period)
   83% of total community pageviews (78% in previous period)

2. www.unidata.ucar.edu/community
   3112 unique pageviews (4321 in previous period)
   6% of total community pageviews (8% in previous period)

3. www.unidata.ucar.edu/events
   2856 unique pageviews (4884 in previous period)
   6% of total community pageviews (9% in previous period)

4. www.unidata.ucar.edu/about
   2395 unique pageviews (2406 in previous period)
   5% of total community pageviews (4% in previous period)

Social media statistics, March 4, 2019

1. # of Twitter followers: 1132 (up from 1016)
2. # of Facebook followers: 786 (down from 788)

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.

Status Report: Community Equipment Awards

Sponsored by the National Science Foundation

April 2019 - October 2019

Admin Group

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Possible theme(s) for the 2020 Unidata Community Equipment Awards;
2. Please consider volunteering to serve on the 2020 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.

This year, in conjunction with last year's Triennial Users Workshop, special consideration was given to proposals that included one or more of the following:

- Projects that advance data-proximate analysis of large remote datasets (e.g. GOES-16 satellite data).
- Projects that facilitate the use of machine learning techniques and data analytics.
- Installation of equipment that provides student access to and use of GOES-R series satellite data.
- Installation of equipment for the operational use of ensemble models and the development of those models' predictions to share with the broader geoscience community.

A Request for Proposals was sent out on November 16, 2018 with a March 08, 2019 submission deadline. The Review Panel met on March 22 after the Users Committee Meeting at the Unidata Program Center and recommended that the following proposal be funded:

- Northern Illinois University, Dr. Victor Gensini, "Bringing Back Weather.NIU.EDU: A Multifaceted Server at Northern Illinois University" proposal

Congratulations to Dr. Gensini and a special thank you to the Review Panel and the NSF for making the Unidata Equipment Awards program possible.

Relevant Metrics

Since taking over the management and administration of the Equipment Awards program in 2003 on behalf of the NSF, Unidata has made 99 awards totaling over $1,250,000.

Prepared September 2019

Status Report: GOES-R Class Satellites

April 2019 - October 2019

Mike Schmidt, Tom Yoksas

Questions for Committee Members

- A reorganization of the content of the LDM/IDD NIMAGE was finalized and went live in June. A reorganization of the UNIWISC datastream is in progress; details are included below

  Given the reorganization, what image coverages, spatial and temporal resolutions and possibly projections should be considered for the UNIWISC IDD feed? For instance, should coverages formerly provided by the NWS via NOAAPort be added to the IDD?

- 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

- Upgraded the IDD NIMAGE datastream by including the full content of NOAAPort-delivered GOES-16/17 ABI and other Level 2 imagery and products.

- Set up remote data serving of GOES-16/17 NOAAPort ABI imagery community data servers that we operate in UCAR.

  By any measure, the GOES-16/17 imagery is **very** popular in the community.

- Implemented UW/SSEC's "fanout server" (redistribution of the GRB-200 UDP unicast stream over TCP) for GOES-17 GRB products.

  We will be feeding from one of the SSEC GOES-17 fanout servers, and they are already feeding from the ingest machine that we operate.

  This setup mirrors the one that we have operated for GOES-16 GRB products for the past year.

- 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.

Background:

In the fall of 2017 we began experiencing significant TI in the GOES-16 signal being received by the 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.

Ongoing Activities

We plan to continue the following activities:

- Ingest GOES ReBroadcast (GRB) streams from GOES-16 and GOES-17 in real-time
- Continue to distribute GOES-16 and GOES-17 data via the LDM/IDD and serve the data via the TDS, ADDE and EDEX
- Canvas the community to learn more about GOES-16/17 data needs
- Investigate additions to the IDD CONDUIT data stream that would be useful for creation of new GOES-16/17 based Level 2 products

Future Activities

Reorganize the contents of the LDM/IDD UNIWISC datastream

GOES ReBroadcast (GRB)

The UPC has been making all products being distributed in the GRB since installation of downlinks for both GOES-16 (Winter 2017) and GOES-17 (Fall 2018), first as test point-to-point IDD SPARE datastream before the satellites were declared operational, and then in the IDD SATELLITE datastream (known as DIFAX in LDM versions prior to and including v6.13.6) and via our data server technologies, the THREDDS Data Server (TDS), McIDAS ADDE and AWIPS 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

NOAAPort SBN

As part of the transition of GOES-East from GVAR GOES-13 to GOES-R GOES-16, the NWS stopped creating/distributing a wide variety of image sectors, in GINI format, in the NOAAPort SBN. These were replaced by Sectorized Cloud and Moisture Imagery (SCMI) tiles that need to be stitched together to make Full Disk, CONUS and one or two Mesoscale "floater" sectors. As far as we can tell, the NWS' plan is to discontinue the GINI image sectors created from GOES-15 imagery when GOES-15 is turned off on December 31. GOES-17 assumed the role as GOES-West on February 11.

The GOES-East/West SCMI tiles that are being distributed in NOAAPort as tiles are available (along with some other products) in the IDD NOTHER datastream.

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

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

Unidata provides a Python script that is used to stitch together the tiles into full scenes in the ldm-alchemy section of the Unidata GitHub presence, and this utility is being used by several sites.

IDD NIMAGE Datastream

Since the process of stitching together NOAAPort-delivered SCMI tiles into usable full scenes requires local processing resources that we believe would be better used for other activities at user sites, we took on the job of reconstituting full images created from tiles and then distributing the full imagery. This was a relatively simple/straightforward effort for the UPC since we are already stitching together the NOAAPort-delivered tiles and making the reconstituted full scenes available by our TDS, ADDE and EDEX servers. The reconstituted

images were added to the rIDD NIMAGE datastream (NIMAGE was named to represent NOAAPort IMAGE) in June.

Also available in the NOAAPort SBN are Level 2 products created from GOES-East imagery. Conversion of the products into forms readily usable by Unidata-supported display and analysis packages was straightforward in that all that needed to be done was to strip off a WMO style header and trailer from the product and then adding the netCDF4 products to the NIMAGE datastream.

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

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

VALUE-ADDE 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 should be directly usable by all of the analysis and display packages that we make available with the possible exception of GEMPAK.

IDD UNIWISC Datastream

After we begin making GOES-East/West images reconstituted from NOAAPort tiles and Level 2 products that originate from NOAAPort and value-added providers (we anticipate that TTU will be the first, but not the only site creating value added products), we plan to review the image products that we create for the IDD Unidata-Wisconsin (UNIWISC) datastream and remove those products that are believed to be redundant. Exactly what the removed images will be replaced with will be determined after consultation with the Unidata Governing Committees and with the Unidata community.

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

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.

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.

Prepared *September 2019*

Status Report: Internet Data Distribution

April 2019 - October 2019

Steve Emmerson, Mike Schmidt, Jeff Weber, Tom Yoksas

Questions for Committee Members

- Suggestions regarding content of data streams like CONDUIT, FNEXRAD, UNIWISC and NLDN LIGHTNING?

  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:

```
20190913

Data Volume Summary for lead.unidata.ucar.edu

Maximum hourly volume  99401.429 M bytes/hour
Average hourly volume  63480.742 M bytes/hour

Average products per hour     492715 prods/hour

Feed                            Average             Maximum     Products
                       (M byte/hour)            (M byte/hour)   number/hour
SATELLITE               14261.190    [ 22.465%]    19464.428     6298.884
CONDUIT                 11616.690    [ 18.300%]    33164.870   107700.140
NGRID                   10322.023    [ 16.260%]    14477.420    66708.605
NEXRAD2                  9739.079    [ 15.342%]    12810.922    91559.698
NIMAGE                   6263.972    [  9.868%]     9218.347     5977.953
FSL2                     6053.238    [  9.536%]    18060.528     8622.930
NEXRAD3                  2915.835    [  4.593%]     3645.234   113627.395
HDS                      1372.762    [  2.162%]     1774.155    42401.581
GEM                       390.110    [  0.615%]     1584.383     2299.256
NOTHER                    218.287    [  0.344%]      787.182       40.814
FNEXRAD                   137.937    [  0.217%]      168.403      104.233
UNIWISC                    96.626    [  0.152%]      142.687       50.023
```

```
IDS|DDPLUS              74.236    [  0.117%]       93.000    46812.674
EXP                      9.808    [  0.015%]       14.135      113.558
LIGHTNING                8.860    [  0.014%]       15.688      396.651
GPS                      0.089    [  0.000%]        1.062        0.767
```

The 14-19 GB/hr value in IDD data volume shown as SATELLITE represents the ingestion of GOES-16 and GOES-17 data in the GOES ReBroadcast (GRB). The 6-9 GB/hr in the NIMAGE entry represents the full volume of reconstituted GOES-16/17 Level 2 products that are distributed in NOAAPort.

New Data Distribution:

The top level IDD relay clusters we maintain were upgraded to support relay of the full volume of GOES-16/17 imagery, products being received via redundant GOES-16/17 downlinks and pre-processed Level 2 products that derive from NOAAPort. The primary top level IDD relay cluster, idd.unidata.ucar.edu, was moved to the NCAR Wyoming super computer facility in Cheyenne, WY.

The data volume seen in the **SATELLITE** (which is known as **DIFAX** in LDM distributions prior to v6.13.6) feed 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.

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.

- Unidata receives the NOAA/GSD experimental High Resolution Rapid Refresh (**HRRR**) grids (both 2D and 3D fields) in an LDM/IDD feed from NOAA/GSD and feed these products to a small number of university sites on **hrrr.unidata.ucar.edu**.

- The NCEP operational HRRR is being served with other model output at:

  http://thredds-test.unidata.ucar.edu/thredds/catalog/idd/forecastModels.html (.xml for machine access)

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

- 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.

  The UCAR NOAAPort dish pointing was moved from SES-1 (approx. 101 W) to Galaxy 23 (approx. 89.5 W) in December 2017. Since the move, the indicated Carrier to Noises/EsNos reported by Unidata Novra S300N receivers improved from the mid-15s to mid-17s with peaks occasionally higher than 18. EsNos in this range indicates exceptionally good reception quality.

  Unidata works with LSU/SRCC to maintain their NOAAPort ingest capability. Activities have included providing a spare LNB to bringing their NOVRA S300N receiver to Boulder for testing, configuration and power supply replacement.

- The NOAAPort-derived data streams (**HDS**, **IDS|DDPLUS**, **NGRID**, **NIMAGE**, **NEXRAD3** and **NOTHER**) are being 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 a revamped **NIMAGE** datastream.

  We continue to look for a fourth ingest site to increase robustness of the IDD distribution of NOAAPort derived data.

- Unidata's NOAAPort ingest package is bundled with current versions of the LDM. The current LDM release is v6.13.11.

- After discovering that the LDM/IDD Product IDs, which are essentially WMO IDs, for the GOES-16 Level 2 (L2) products that are being distributed in NOAAPort are not sufficiently descriptive to easily allow for saving of all products to disk via LDM pattern-action file actions, we implemented enhanced processing to insure that all L2

products are processed and then made available via our data services. These L2 products are also being relayed in the revamped **NIMAGE** datastream.

Relevant Metrics

- Approximately **585** machines at **229** sites are running LDM-6 **and** reporting real-time statistics to the UPC.

  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.

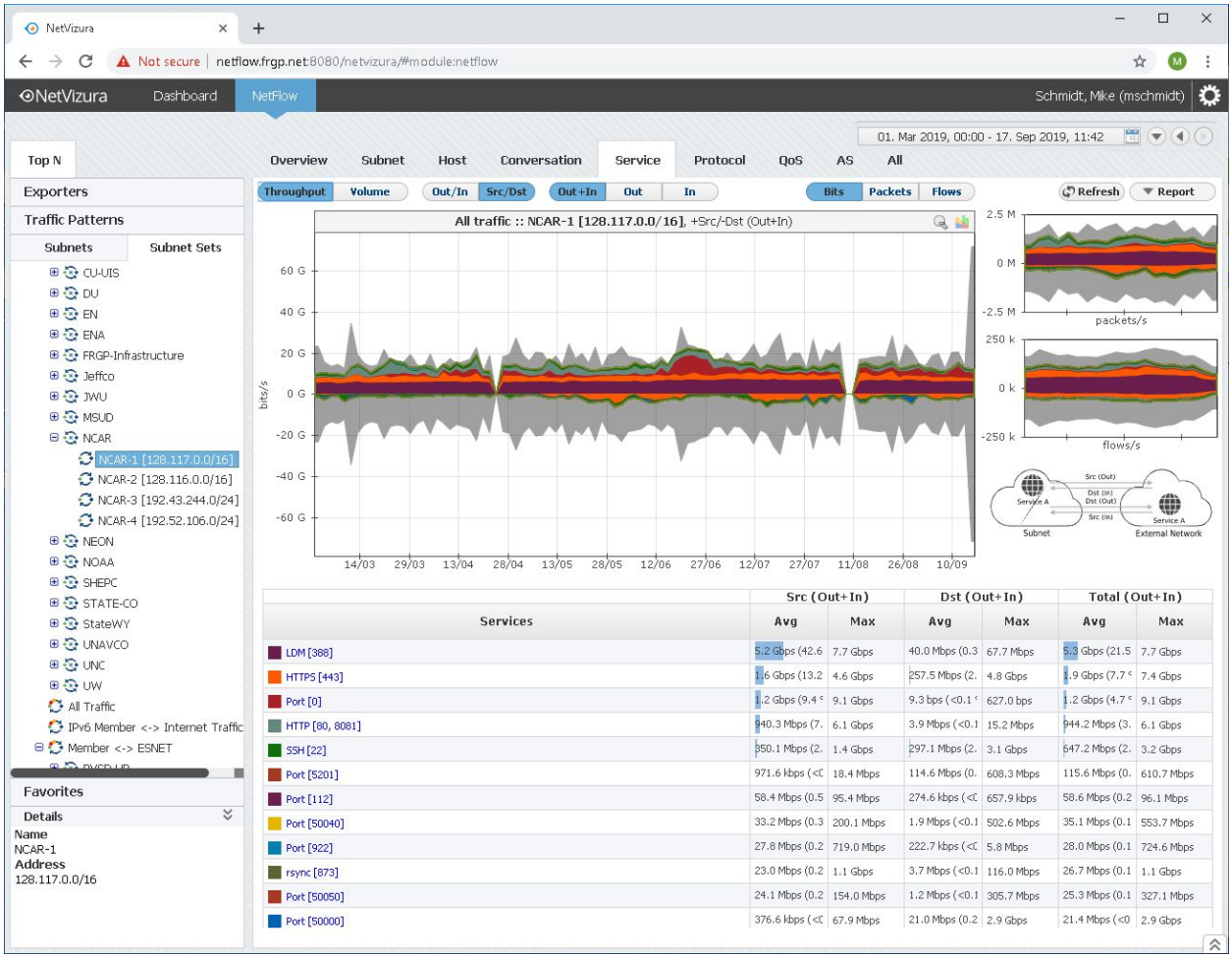
  Unidata staff routinely assist in the installation and tuning of LDM-6 at user sites as a community service.

- A number of organizations/projects continue 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 cluster, described in the June 2005 CommunitE-letter article Unidata's IDD Cluster, routinely relays data to more than 1076 downstream connections. It was moved to the NCAR/Wyoming Super Computing facility in Cheyenne, WY in late August.

  Over the period from March 11, 2019 through September 17, 2019 the average volume of LDM/IDD data flowing from the UCAR/NCAR network averaged around 5.2 Gbps (~54 TB/day), and peak rates reached 7.7 Gbps (which would be ~126 TB/day if the rate was sustained). The volume of data being distributed via the IDD from UCAR increased significantly with the augmentation of the **NIMAGE** feed in June.



Our new cluster machines have 10 Gbps Ethernet capability.

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 *September 2019*

Status Report: IDV with RAMADDA

April 2019 - October 2019

Yuan Ho, Cece Hedrick, Julien Chastang

Areas for Committee Feedback

We have no questions at this time.

Activities Since the Last Status Report

IDV Release

IDV 5.6 has been released in May 24 of 2019.

__Youtube IDV Instructional Videos__

Our summer intern Jessica Wiedemeier has created more than 15 IDV videos and plan to continue producing more instructional videos on the IDV. We would appreciate input and suggestions on specific video topics.

IDV System Changes

__Latest netCDF-Java Version__

The version of the netCDF-Java library currently distributed with the latest stable IDV version (5.6) is the 4.6.13-SNAPSHOT . The prior version of netCDF-Java to be distributed with the IDV was 4.6.12. There have been many improvements and bug fixes in that range. The complete release notes for these versions can be found here.

__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.

IDV Display Changes

__Integrated DRILSDOWN plugins__

Integrated DRILSDOWN Mapes collection plugin into the IDV release. This includes new display types of probing in pressure coordinates and cross section in pressure coordinates.

__GOES 16/17 GLM ADDE Chooser__

The GLM L2 product consists of geo-located and time-stamped events, groups, and flashes, with associated calibrated optical amplitudes (in units of Joules). The IDV has a new data chooser to access three GLM source types: Geostationary Lightning Mapper EVENT, GROUP, and FLASH Data through the UNIDATA ADDE data server. The GLM data queries can be performed on the relative times or absolute times.

__3D Streamline Display__

Streamlines are an instantaneous tangent to the velocity vector of the flow and a snapshot of the flow field at some particular time. In the IDV 3D space, streamline display provides the direction in which a massless fluid element will travel at any point in time. 3D streamline display type is available for both derived parameters: Grid 3D Trajectory and 3D Flow Vectors, and the initial area of the 3D streamlines can be a surface or a 3d volume.

__3D Trajectory Display Update__

The grid 3D trajectory display include new display types: Tracer and Tracer Point. The initial area of the 3D Grid Trajectory can be set at XZ or YZ plane, and controlled by the data index in the X or Y direction.

__3D Isentropic Advection__

The new grid 3D isentropic advection parameters interpolate advection from isobaric space into the isentropic space and creates the 3D Theta surface advection displays.

__Grid 3D Vertical Profile in Pressure Axis__

Traditional Skew-T Log-p plots are very poor at understanding tropical convection - moisture interactions, at lower atmosphere. New vertical profile display using pressure as its axis to emphasize information in the lower levels.

__Latest Version of VisAD__

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

IDV EarthCube Drilsdown Project

In collaboration with University of Miami Professor Brian Mapes, the project facilitates the use of data visualizations within curated, executable notebook documents, in the service of improving statistical algorithms for the datarich geosciences. The project integrates Jupyter Notebooks as the medium for literate computing and documentation; Unidata's IDV for its powerful data integration, visualization, and algorithm design capabilities; and RAMADDA as a metadata-rich repository for the packages that are the objects of this workflows.

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 American Meteorological Conference (AMS) Annual Meeting__

- Visualization of Wind Parcel Trajectories Using McIDAS-V and IDV
- 3D Space Isentropic Analysis in the Unidata's IDV
- VISUALIZATION FOR IMPROVING ALGORITHMIC MEASURES OF EVENTS

__2019 IDV lectures at WRF tutorial workshop__

- Yuan delivered two IDV introduction lectures on two semi-annual WRF tutorial workshops.
- Julien delivered one IDV presentation on the WRF user workshops.

Ongoing Activities

We plan to continue the following activities:

__NetCDF Java 5.0.0 integration__

The new version of netCDF-Java will be integrated with the IDV. There have been many improvements and bug fixes in this new netCDF java library.

__Investigation of Java 3D Alternative__

Because of concerns about the long-term viability 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:

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 ~20 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.

__Usage Metrics__

https://www.unidata.ucar.edu/software/idv/logging/left.html.

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

April 2019 - October 2019

Mike Schmidt, Matt Perna

Major Activities

Cybersecurity Analysis -- UCAR is in the midst of a cybersecurity evaluation and assessment commissioned by the UCAR Board of Trustees. The entire process is expected to take as much as six months, and be completed in stages with reporting, problem resolution and meetings throughout. As of yet, we are not aware of any serious findings with regard to Unidata's operations.

IDD cluster -- We've switched the operational IDD cluster from UCAR's Foothills Lab campus to the NCAR Wyoming SuperComputing Center (NWSC) in late August. Notable changes with this cluster include 10Gb/s network service and the ability for LDM queue sizes to hold several hours worth of data .

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.

UCAR FISMA -- UCAR currently has contractual FISMA requirements for a small parts of the organization (none for Unidata currently), but the plan is to start down the path of making the entire organization FISMA compliant at some level yet to be determined (probably low, possibly medium). We continue to attend meetings to stay connected to the process

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 just 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 *September 2019*

Status Report: LDM

April 2019 - October 2019

Steve Emmerson, Tom Yoksas, Mike Schmidt, 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:

- ldmd(8):
    - Top-level Server:
        - Eliminated having the operating-system assign a transitory port number for the LDM server if the server couldn't get its official port number
        - Configuration-file:
            - Improved signal-handling
            - Eliminated memory-leak found by Coverity Scan
        - Logging: Improved support for standard error stream being redirected to `/dev/null`
    - Downstream/Receiving LDM: Eliminated `SIGSEGV` bug
- pqact(1):
    - Shielded `EXEC` entries from ldm process-group signals
    - Improved signal handling
- ldmping(1): Made the main loop interruptible. Reconciling signal-handling amongst the various programs apparently caused this one to loop forever.
- Logging:
    - Unconditionally included the ulog(3) logging module in the LDM library for use by third-party software (e.g., McIDAS)
    - Improved concurrency locking for access by multiple threads
    - Made functions async-signal-safe
    - Improved format of log messages
    - Ensured that all programs check the return-value of `log_init()` and `log_set_destination()`
- NOAAPort:noaaportIngester(1):
    - Corrected initialization of last received frame when the run number changes. This lint was found by Coverity Scan.
    - Redirected `stderr` to `/dev/null` if it's closed because some GRIB/GEMPAK functions mistakenly write to it and it was open on the LDM's product-queue -- causing an assertion failure
- Misc:
    - Compiled under "`gcc -Wall`" and fixed almost all warnings

    - Improved many log messages.
- Released versions 6.13.9, 6.13.10, and 6.13.11

Dependencies, challenges, problems, and risks include:

The LDM is sometimes held responsible for decisions made by the NWS 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.

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 at the University of Virginia with the time-tested unicast capability of the current LDM.

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

Progress has been made on the following:

- LDM configuration-file:
    - RECEIVE entry:
        - Corrected description of <host>
        - Improved description of <al2sVlanId>
    - MULTICAST entry: Eliminated parameters
        - Multicast interface
        - Time-to-live
- Downstream/Receiving LDM7:
    - Corrected AL2S VLAN endpoint <-> FMTP interface consistency-test
    - Increased RPC timeout for subscribe_7() because adding a node to an AL2S multipoint VLAN can take longer than the default RPC client timeout
    - Modified to pass CIDR address for virtual interface to vlanUtil(1)
    - Made down7_halt() async-signal-safe to obviate deadlock if interrupted while logging
    - Differentiated "backlog" from "backstop" product-insertion messages
    - Removed lint found by Coverity Scan
    - Added number of FMTP retransmissions to "Received" log message
- OESS Scripts:
    - Replaced calls to ulogger(1) with sys.stderr.write() because sudo() handles the child process writing to stderr
    - provision.py:
        - Replaced writing of circuit-ID to stderr with writing to stdout
        - Added case where multipoint VLAN already exists
    - Scripts now expect the name of the OESS workgroup to be in the OESS

account file
- UpMcastMgr(3):
  - Added installation of account.py(1) and edit.py(1)
  - Deleted invocation of Python interpreter because scripts are executable
  - Removed acquisition of root privileges when calling "provision.py"
  - Added printing of "provision.py" command-string on failure
  - Deleted logging of child process that error-exits
  - Added to mldm_ensureExec() removal of terminated sender from mldm_sender_map(3)
- FMTP Module:
  - `fmtpRecvv3::Start()`: Modified to iterate its attempt to connect to the sending FMTP server because it can take time for the AL2S VLAN to be provisioned. It will attempt every 5 seconds for 2 minutes.
  - fmtpSendv3(3):
    - `::RunRetxThread()`: Don't terminate process on `TcpSend::parserHeader()` error
    - Adapted `::logMsg()` to use LDM logging
    - Eliminated this module terminating the process if an exception occurs
    - Eliminated slicing of the exception that caused an instance to stop working
  - fmtpRecvv3(3): Added per-product tracking of number of FMTP data-block retransmissions
- Deployed to several sites, ran tests, and collected performance metrics. Now in the process of analysis and paper submission.

Dependencies, challenges, problems, and risks include:

The trial deployment of the multicast LDM has revealed a problem with Internet2's Advanced Layer 2 System (AL2S): reconfiguring the multipoint VLAN to add a new site causes the entire VLAN to be destroyed and then re-created with the new site. This process can take several minutes and can cause severe loss of data and/or thrashing of the multipoint VLAN. We are working with Internet2 on a solution.

Ongoing Activities

We plan to continue the following activities:

- Support and maintenance of the LDM
- Continue adapting the LDM to the Docker container technology to support cloud activities and (perhaps) make life easier for LDM users

New Activities

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

- Continued deployment and testing of LDM-7

- Working with Internet2

Relevant Metrics

- Data on the LDM package can be found here
- 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 *September 2019*

Status Report: McIDAS

April 2019 - October 2019

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 was the creation of ADDE servers for NOAAPort-delivered GOES-16/17 satellite imagery and merging the resulting code into the McIDAS core at UW/SSEC.

Current Activities

- Unidata McIDAS version 2019 is currently being prepared for release

  v2019 will include all SSEC versions up to and including the current McIDAS-X and -XCD releases, both of which are v2019.1.

- 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 latest releases feature the following:

- Updated GOES-R Series ABI servers to store GOES-16 and -17 Temperature Data Quality Flag (TDQF) information and report the values in **IMGLIST** command's FORM=AUX output, support ABI L2 image products that derive from NOAAPORT distributed image tiles, remove restrictions on names of directories containing the ABI files, and to skip past corrupted files so that IMGLIST and other commands will proceed to subsequent "good" files.
- Updated VIIR servers with improvements and enhancements. The changes include increasing the number of files allowed in a dataset from 5,000 to 20,000 and removing previous dataset descriptor naming requirements. The servers were also updated to use the SATANNOT file's new sensor source numbers 320 (SNPP-SDR) and 321 (NOAA20-SDR).
- Added DAY= and TIME= keywords to **IMGPROBE** command to allow use with archive or other servers that require the DAY= keyword to be specified. The **D** command (Alt-D) was also updated to include the day value from the frame directory in its

request so that it also works with those servers.

- Updated **FRMLABEL** and **ZA** commands to include new *(DAYJ)* placeholder that plots the date in format *dd mon ccyyddd*, e.g., 12 OCT 2019285. Also updated FRMLABEL to identify RGB frames (displayed using the **RGBDISP** command) and thus apply an appropriate image label with no overlapping characters.

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.

The next major addition to this package is 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 has been averaging over 40 TB/month since the since the late spring of this year

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 (when available) imagery. GOES-15 is scheduled to be decommissioned on December 31 of this year. GOES-14 will remain in its standby location (104W) and will be turned on for periodic testing as needed.

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 1.4 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 sot for use by those interested in satellite meteorology worldwide.

Prepared *September 2019*

Status Report: netCDF

April 2019 - October 2019

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 (zarr, Azure, etc) that are actively in use? That we should consider storing.
2. Are there any emergent avenues (stack overflow, etc) for user support which the netCDF team should investigate?
3. 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 **90** open issues for netCDF-C, **43** open issues for netCDF-Fortran, and **15** 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:

- Multiple releases of the core C library, as well as the Fortran and C++ interfaces.
- Refinement of user-defined compression filters.
- Work towards enhanced parallel I/O.
- Further extension of the netCDF build-and-test platforms using Docker technology.
- Further enhancements to the netCDF-C documentation, modernization of the netCDF-Fortran documentation.
- Extended continuous integration platforms have been adopted.
- An architecture roadmap is available describing how the netcdf-c library will support thread-safe operation in *nix* and Windows environments. The draft proposal is available as netcdf-c github issue #382.
- We have seen 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.

Dependencies, challenges, problems and risks include:

- Small group of developers for supporting large project.
- Dependency on HDF5, controlled by external group.
- Rapid evolution of Zarr standard is both 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.

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 organization of Doxygen-generated 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 begun the technical work of adopting Zarr functionality in the core C library. This will allow for object-based storage (Amazon S3, etc).

Status

- Meta data is being properly written and read
- The content data of variables is not yet supported.
- ncgen and ncdump -h work with some limitations (as shown)
- Much testing:
    - move and modify tests from the ncdump/ and the nc_test4/ directories to zarr_test/.

Support for pure zarr

- Requires producing simulated data when _ncz... is missing:
    - simulated _nczarr: get the zarr version from the root group, use library built-in value for nczarr_version.

- simulated _nczcontent: Assume we have a group whose key is e.g. /y/z/g.
- dims: All variables whose shape is, say, (m,n), create dimensions in the root group of form dim_m=m and dim_n=n.
- vars: collect values of X for all keys of form "/y/z/g/X/.zarray".
- grps: collect values of X for all keys of form "/y/z/g/X/.zgroup".
- simulated _nczvar - contains netcdf-4 specific information for a
- Variable
    - dimrefs: Using the shape of the variable to figure out the dim names
    - contiguous is always false
- simulated _nczattr: infer attribute type from the values of the
- Attribute
    - (process is somewhat complex and is similar to algorithm in ncgen)

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

- Seek out, and prepare material for upcoming, conferences and other outreach opportunities.
- Work on reducing the defects reported by static analysis.
- Release the next versions of netCDF-C, netCDF-Fortran, netCDF-C++.
- Continue modernizing the netCDF documentation to provide easy access to documentation for older versions of netCDF.
- Publish formal documentation/roadmap for implementing cloud-native storage (object storage).

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.
- Implement support for native object storage in the netCDF C library.
- Enhance thread-safety for the netCDF C library.
- Participate in the HDF Technical Advisory Board process.
- Participate 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

There are currently about 202,428 lines of code (up from 192,041 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 **0.69** 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**. The jump in defect density is a result of the addition of the **DAP4** code. As this is new code, the initial defects are still being worked out.

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:

- **925,000** for netCDF-3
- **951,000** for netCDF-4
- **1,110,000** for HDF5
- **172,000** for GRIB2

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

- **351** for netCDF-3
- **744** for netCDF-4
- **13,000** for HDF5
- **1,070** 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 *September 2019*

Status Report: Outreach to Underserved Communities

April 2019 - October 2019

Doug Dirks, Jeff Weber, Joshua Young

Areas for Committee Feedback

We are requesting your feedback on the following topics:

All questions are featured as agenda topics

Activities Since the Last Status Report

Activity 1

- Development of a draft outreach engagement strategy
- Identification of priority community events (SACNAS, AISES, etc)

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:

- This is a new activity as an intentional project

New Activities

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

- Community services will participate in the AISES and SACNAS meetings this fall

- Develop outreach stakeholder list for broadening distribution of opportunities (equipment awards, internships, etc)

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 *September 2019*

Status Report: Python

April 2019 - October 2019

Ryan May, Zach Bruick, Sean Arms, Julien Chastang

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. What can we do to improve the utility of our training resources for you and your students? How can Unidata do more as a core Python training resource for the community? How do we grow our reach and serve the community better in this area?
2. Are there additional opportunities (e.g. conferences) we should explore as convenient venues for teaching workshops/courses?
3. Are we missing something? Anything you notice as lacking in MetPy or Siphon?

Activities Since the Last Status Report

Staffing Changes

In May 2019, Zach Bruick was hired to fill the position vacated by John Leeman. Zach has been working hard coming up to speed on the various projects and responsibilities of the Python team, including making numerous contributions to MetPy and helping teach the workshop at Valparaiso University. John Leeman has been retained on a contract basis to continue the production of MetPy Mondays.

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. We also continue to explore other options for conferences for teaching Python workshops. One thought is to look at traditionally science-focused conferences as a potential venue to teach more short courses.

We have begun working to unify our various training efforts (gallery, workshop, and online python training) into a proper “Python Training” site. The goal of this effort is to consolidate so that discoverability of the resource is improved, as well as producing a more cohesive effort around Python training. This effort should also make it easier for members of the community to contribute to the collection.

Progress has been made on the following:

- Ryan May and Zach Bruick, together with Kevin Goebbert, will be teaching a short course using MetPy at the 2020 AMS Annual Meeting, focused on MetPy's simplified plotting interface.
- Ryan May and Max Grover, one of Unidata's summer interns, taught a workshop to 31 attendees at SUNY Albany.
- Ryan May and Zach Bruick taught a regional workshop at Valparaiso University on 12-14 August. There were 17 attendees from a variety of institutions, including Ball State, Central Michigan, Northern Illinois, Indiana, Purdue, Notre Dame, and Valparaiso.
- Zach Bruick and Ryan May have taught MetPy introductory tutorials as part of two different Python tutorials jointly hosted by NCAR's CGD and CISL groups.
- John Leeman continues to lead the "MetPy Mondays" effort.
- We assisted Brian Blaylock (University of Utah, PhD student) with a GOES-16 True Color Jupyter Notebook. We helped Brian ensure the notebook conforms to modern coding standards as well as assisting him with conversion from Basemap to Cartopy and remote data access. This is now available as part of the Python Gallery.

MetPy

MetPy continues to grow, especially in terms of use. During the last year, MetPy's web documentation averaged 16347 page views per month; this is approximately double the rate from the previous year. MetPy was also downloaded 50812 times last year. Upon digging into the statistics, many of these downloads are from older versions, which is difficult to interpret. The 3 releases from the last year were downloaded a total of 29882 times. This represents a large increase over the previous year. There are also now 19 scholarly publications (journal articles, theses, technical reports) that cite or mention MetPy; this count stood at 3 at this time last year.

Development continues to be driven by requirements for our dedicated awards (in addition to bug reports and pull requests from community members). The primary efforts will be focused around the GEMPAK-like interface, improved units support, and integration with xarray. MetPy 1.0 is planned for release later this year, shortly before AMS 2020. Also, to try to foster more community discussion of MetPy's goals and plans, we have published a general MetPy roadmap that tries to capture our plans from GitHub in a more friendly format. The 0.11 release ended up delayed from early summer to late September; this was a consequence of the staffing changes and a focus on training efforts.

Progress has been made on the following:

- Community awareness continues to grow, with the volume of engagement and mentions on social media growing; the MetPy twitter account has reached 932 followers (24% growth in 6 months).
- MetPy 0.10.2 was released with a few fixes for compatibility issues with upstream dependencies
- MetPy 0.11 was recently released, including additional enhancements to the GEMPAK-like plotting functionality (including docs), Skew-T enhancements, and many fixes, including addressing some issues with profile calculations.
- Work towards requirements of MetPy-related NSF awards

- Max Grover, as part of his summer internship, finished work on a parser for METAR observations

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 continues to see community contributions trickle in. One of our summer interns, Aodhan Sweeney, extended Siphon to interface some of the publicly available, text-based data products from the National Hurricane Center (NHC) and the Storm Prediction Center (SPC). Thanks to Aodhan's work, Siphon can now read storm reports directly from the SPC and, building off of a previous interns work (Florita Rodriguez, 2014), can now access track and intensity data from the National Hurricane Center Database (NHCD). More information can be found on Aodhan's Developers@Unidata blog post.

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.

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.

Progress has been made on the following:

- Ryan May and Zach Bruick attended SciPy 2019
- Ryan May will attend a NumFocus Project Summit for the matplotlib project
- 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
- Ryan May continues to work as a developer on the CartoPy project and will be managing another release of CartoPy in the Fall of 2019.
- We also continue to actively engage with the xarray, numpy, and pint projects

Ongoing Activities

We plan to continue the following activities:

- Unidata Python training workshop
- 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:

- Teach additional Python regional workshops
- Restructure our annual Python training materials into a more unified Python training site
- Present at AGU session on open source geoscience tools
- Teach an advanced, application-focused Python workshop at Texas Tech University

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

- Teach another short course on MetPy at AMS 2020
- Teach a short python overview course at AMS 2020 student conference (Sunday, January 12, 2020 from 2:00pm to 3:45pm)
    - Julien is setting up a JupyterHub server for use at this workshop (needs to support up to 200 students)
- Present annual update on Python libraries at AMS 2020
- Teach an additional Python workshop at a Minority-serving institution

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

Relevant Metrics

MetPy

- 98% test coverage
- Watchers: 50
- Downloads for the releases made in the last year (Conda + PyPI):
    - 0.9.2: 12882

    - 0.10.0: 11801
    - 0.10.1: Not Released
    - 0.10.2: 5301
- Since 1 March 2019
    - Active Issues: 87 (50 created, 49 closed)
    - Active PRs: 90 (80 created, 72 closed)
    - External Issue Activity: 31 opened, 102 comments
    - External PR Activity: 30 opened, 77 comments
    - Unique external contributors: 36
    - Stars: 90 (470 total)
    - Forks: 6 (187 total)
    - Commits: 189
- Since 1 September 2018
    - Active Issues: 129 (93 created, 70 closed)
    - Active PRs: 119 (113 created, 102 closed)
    - External Issue Activity: 57 opened, 158 comments
    - External PR Activity: 47 opened, 87 comments
    - Unique external contributors: 55
    - Stars: 165 (470 total)
    - Forks: 7 (187 total)
    - Commits: 282

Siphon

- 97% test coverage
- Watchers: 16
- Downloads for the last year (Conda + PyPI):
    - 0.8.0: 22132
- Since 1 March 2019
    - Active Issues: 6 (6 created, 2 closed)
    - Active PRs: 11 (10 created, 8 closed)
    - External Issue Activity: 2 opened, 0 comments
    - External PR Activity: 3 opened, 9 comments
    - Unique external contributors: 4
    - Stars: 8 (99 total)
    - Forks: 1 (41 total)
    - Commits: 31
- Since 1 September 2018
    - Active Issues: 23 (20 created, 3 closed)
    - Active PRs: 17 (15 created, 13 closed)
    - External Issue Activity: 9 opened, 15 comments
    - External PR Activity: 5 opened, 9 comments
    - Unique external contributors: 13
    - Stars: 19 (99 total)
    - Forks: 2 (41 total)
    - Commits: 43

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 *September 2019*

Status Report: Support

April 2019 - October 2019

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?

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.

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

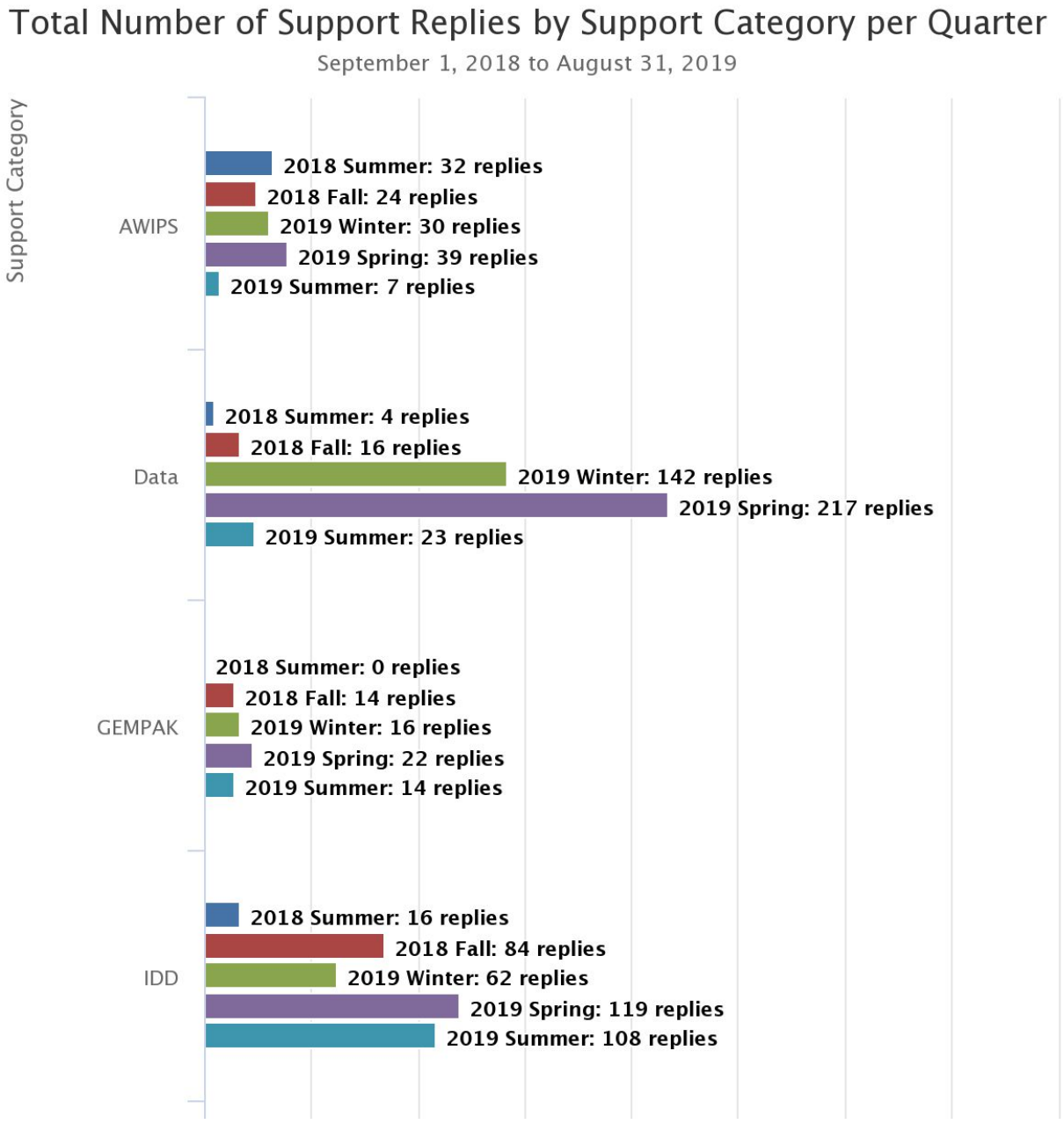
Since January 26, 2006 over 59600 user support "transactions" (new inquiries and follow-ups) have been processed through the Unidata inquiry tracking system. Other methods of providing answers to questions posed (e.g., Github, Stack Overflow, Jira, mailing list replies, etc.) add substantially to the support load.

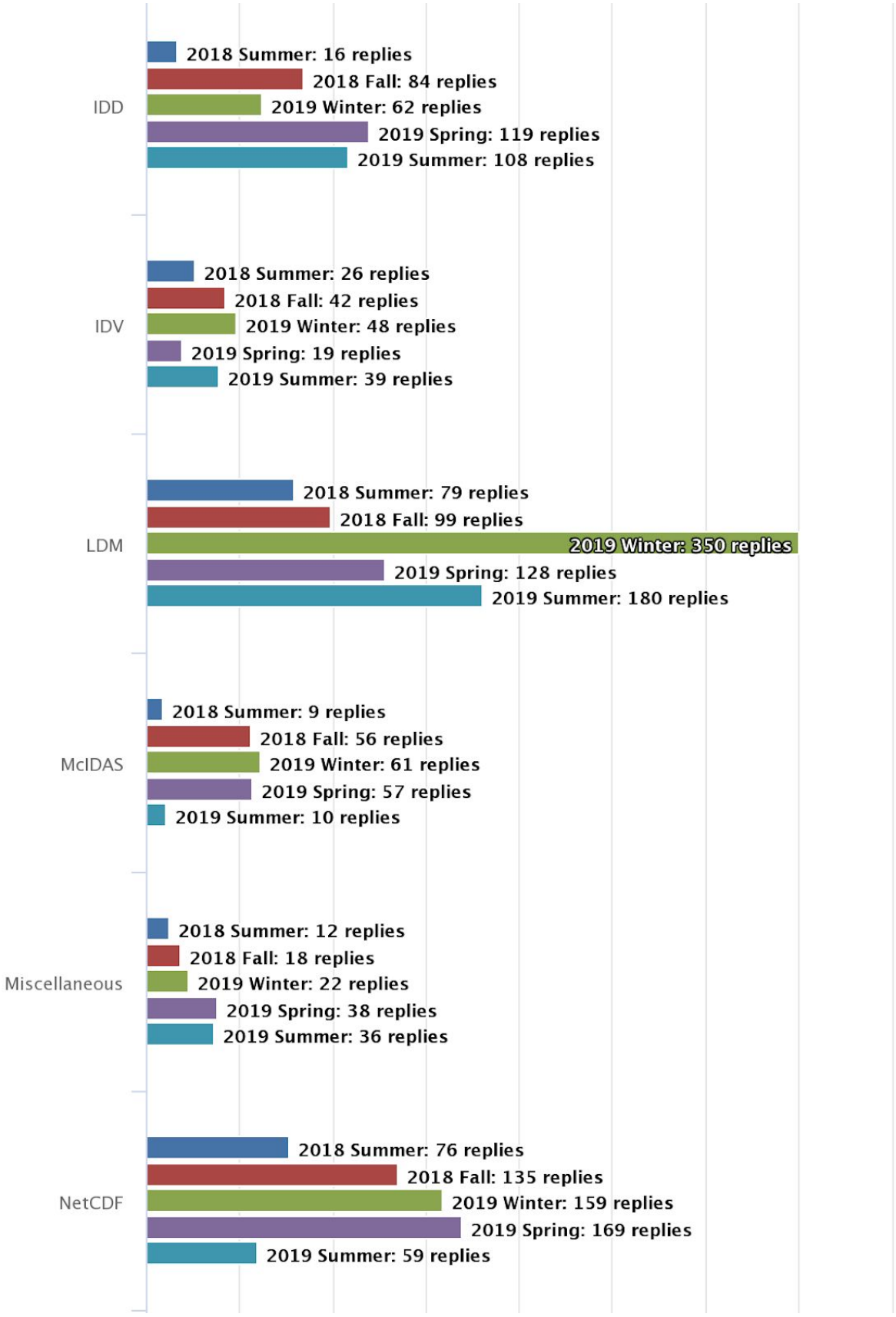
Additional metrics may be found in the Comprehensive Metrics Data portion of this meeting's agenda.

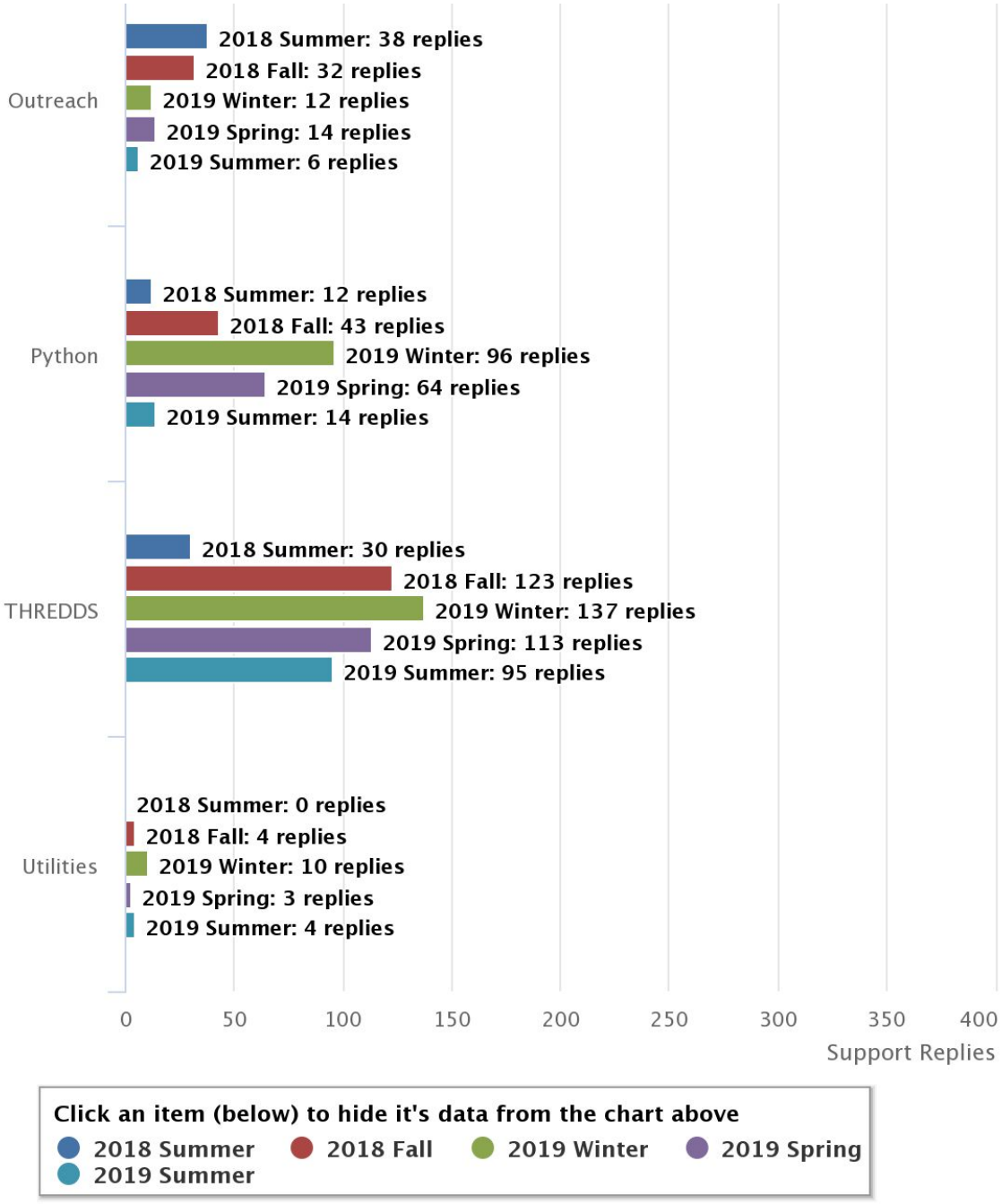
Fig. 1: Below are histograms that portray the number of Unidata email responses for categories of support logged in the Unidata Inquiry Tracking System for the 12 month period from September 1, 2018 until August 31, 2019.

The quarters shown are defined as:

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




IDD
2018 Summer: 16 replies
2018 Fall: 84 replies
2019 Winter: 62 replies
2019 Spring: 119 replies
2019 Summer: 108 replies
IDV
2018 Summer: 26 replies
2018 Fall: 42 replies
2019 Winter: 48 replies
2019 Spring: 19 replies
2019 Summer: 39 replies
LDM
2018 Summer: 79 replies
2018 Fall: 99 replies
2019 Winter: 350 replies
2019 Spring: 128 replies
2019 Summer: 180 replies
McIDAS
2018 Summer: 9 replies
2018 Fall: 56 replies
2019 Winter: 61 replies
2019 Spring: 57 replies
2019 Summer: 10 replies
Miscellaneous
2018 Summer: 12 replies
2018 Fall: 18 replies
2019 Winter: 22 replies
2019 Spring: 38 replies
2019 Summer: 36 replies
NetCDF
2018 Summer: 76 replies
2018 Fall: 135 replies
2019 Winter: 159 replies
2019 Spring: 169 replies
2019 Summer: 59 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, 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 *September 2019*

Status Report: THREDDS

April 2019 - October 2019

Sean Arms, Ethan Davis, Dennis Heimbigner, Cece Hedrick, Ryan May, Jennifer Oxelson, Howard Van Dam II

Areas for Committee Feedback

We are requesting your feedback on the following topics:

1. Do you know of anyone using netCDF-Java to read Vis5D grid files?
2. Do you know anyone who just **loves** BUFR and understand it inside and out? Unidata could use your help (knowledge of Java **not** required)!
3. With an eye towards guiding our outreach efforts, what THREDDS Data Servers do you use, and are there any that you consider critical to your needs?

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.

The various THREDDS projects were featured in a paper regarding Data Interoperability presented at OceanObs'19, a decadal conference series which seeks to improve response to scientific and societal needs of a fit-for-purpose integrated ocean observing system, for better understanding the environment of the Earth, monitoring climate, and informing adaptation strategies as well as the sustainable use of ocean resources.

Data Interoperability Between Elements of the Global Ocean Observing System. Derrick P Snowden, Vardis Tsontos, Nils Olav Handegard, Marcos Zarate, Kevin M. O'Brien, Kenneth S Casey, Neville Smith, Helge Sagen, Kathleen Bailey, Mirtha Lewis, Sean Arms. Review, Front. Mar. Sci. - Ocean Observation, Submitted on: 15 Nov 2018. DOI: 10.3389/fmars.2019.00442

Released netCDF-Java 5 (Stable)

- NetCDF-Java version 5.0.0 was released on 29 July 2019. NetCDF-Java version 5.1.0 was released on 12 September 2019.
- Prior to version 5, the netCDF-Java/CDM library and the THREDDS Data Server (TDS) have been built and released together. Starting with version 5, these two packages have been decoupled and now live in separate git repositories, allowing new features or bug fixes to be implemented in each package separately, and released

independently.
- The codebase of netCDF-Java can be found at https://github.com/unidata/netcdf-java

Released TDS version 4.6.14 (Stable)

- TDS version 4.6.14 was released on 29 July 2019.
- As part of this release, netCDF-Java version 4.6.14 was also released. However, the 4.6.x line of development for netCDF-Java is now in maintenance mode, and will only include security related fixes. All users of netCDF-Java are strongly encouraged to move to the latest netCDF-Java (as of this report, version 5.1.0).

Released TDS version 5.0.0-beta7

- TDS version 5.0.0-beta7 was released on 29 July 2019.
- We anticipate releasing a stable version of TDS 5.0.0 before the next committee meeting.
- Starting with TDS v5.0.0-beta7, the TDS codebase can be found at https://github.com/unidata/tds

Documentation for netCDF-Java / TDS (Beta) v5

- Both netCDF-Java and the THREDDS Data Server documentation sets have changed significantly, and now use Jekyll and a markdown driven documentation theme for documentation generation. Previously, the documentation was a loose collection of Adobe Dreamweaver generated HTML. 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/5.1/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. A poster was presented at the Fall AGU 2018 meeting with respect to the advances in Rosetta related to OIIP. We continue to work with JPL as part of their user acceptance process, with the ultimate goal being the operational use of Rosetta at the PO.DAAC. We, along with the PI of the OIIP project, participated in an ESIP Marine Data Cluster presentation.

Progress has been made on the following:

- Supporting user selection of an appropriate standard name from the CF Conventions has been particularly challenging. Work has been done to create a mind-map of the various standard names as a design prototype for building a standard name selection

widget. While it is easy to stuff all of the available standard names into a dropdown list (>4000 names), it is not particularly user friendly, does not promote discoverability, and is certainly not performant.

General dependencies, challenges, problems, and risks include:

- While all java based components in the THREDDS project run under Java 11, all except Rosetta currently build with Java 8. Portions of our build infrastructure need to be reworked to use Java 11. The end of 2020 is generally marked as end-of-life for Java 8, and thus moving to Java 11 is a priority.
- 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 Pokrant, Victor Gensini, the NCAR RDA, etc.!).
- While the list of names on this report seems large, the current staffing levels on all three components covered in this report is less than 2.5 FTE (including externally funded efforts). A similar resource landscape can be seen for nearly every other project run by the Unidata Program Center. External funds help (currently seeking one opportunity), but rarely provide the ability to bring on and sustain new staff members, which results in taking resources away from other projects and efforts within the Unidata Program Center. Given your position as members of our governing committees, you play a critical role in helping us set priorities. Your feedback is greatly desired and very much appreciated.

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.

The following active proposals directly involve THREDDS work:

- Thanks to Rich Signell, we, along with Axiom Data Science, submitted and were **awarded** a NOAA IOOS grant. The proposal was entitled “A Unified Framework for IOOS Model Data Access”, and the goal 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 will fund a Undiata staff member at 0.5 FTE for two years, as well as two co-PIs at Axiom Data Science at a slightly lower level. 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).

- We partnered with JPL / PO.DAAC on a NASA ROSES Advanced Information Systems Technology (AIST) solicitation. The proposal was titled “MIKITA – Multi-sensor Data Integration Microservices for Knowledge InTensive Applications.” The focus of the work is on extending the Metadata Profile Service (MPS) and creating Semantic Technology based microservices which leverage the MPS. While the bulk of the work is on the JPL side, we are proposing to extend Rosetta to interact directly with the MPS through its web API, and extending the TDS to provide metadata records in a more semantic friendly format, such as JSON-LD based on schema.org. While we at Unidata are not experts on the semantic web, we can certainly empower those who are. The semantic technology work **strategically aligns with the Unidata 2024 focus area “Managing Geoscience Data, Making Geoscience Data Discoverable”** by fostering community adoption of standard data discovery services to help users locate and acquire data appropriate for their projects, working to ensure that data available from Unidata-managed services are discoverable through standard data discovery mechanisms, and using Schema.org markup of datasets in support of search engine discovery (i.e. Google DatasetSearch), as well as **“Managing Geoscience Data, Making Geoscience Data Usable”** by creating ways that netCDF-Java and the TDS can take advantage of “Linked Data” best practices for exposing, sharing, and connecting datasets across networks. Our work on Rosetta **strategically aligns with the Unidata 2024 focus area “Providing Useful Tools, Creating Modern Data Management Workflows”** by further enabling long-tail data providers to produce files that conform to widely adopted, community driven data and metadata standards, as well as **“Managing Geoscience Data, Making Geoscience Data Usable”** by promoting the use of data standards like netCDF with the Climate and Forecast (CF) conventions that allow scientists to quickly understand the shape and provenance of datasets. In general, this work **strategically aligns with the Unidata 2024 focus area “Providing Useful Tools, Creating Modern Data Management Workflows”** through the creation of cloud-native data access services, based around common cloud technologies like object store, hosted databases, and “serverless” technologies.

New Activities

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

- netCDF-Java
    - Deep look into current dependencies and reduce as much as possible.
    - Facilitate migration from netCDF-Java 4.6.x to 5.x where we can.
    - For the most part, this is a “baking” period for 5.x - focus on bug fixes, non-visibile library changes, few new features.
        - Exception - enhance support for S3 storage
    - Complete command line tool creating a WRF intermediate file from a subsetted GRIB dataset.

- TDS

    - Getting TDS v5.0 to a stable release (release candidate targeted for late 2019/ early 2020).
    - 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.
- Outreach
    - AGU 2019 presentation covering UGRID work (supported by NOAA/NOS/IOOS COMT award).

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

- netCDF-Java
    - Begin to modularize (Java Platform Module System) but maintain Java 8 compatibility.
    - Define public API and get 90%+ test coverage of it.
    - API breaks likely as we restructure our current artifacts
    - Initial support for reading / writing Zarr
- TDS
    - Release TDS version 5 (Stable)
    - Create a TDS Registry

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

- netCDF-Java
    - Remove deprecated code
    - Fully support Java 11 and the Java Platform Module System (end of Java 8 support)
    - Commit to semantic versioning
- TDS
    - Reduce dependency footprint
    - Modularize (Java Platform Module System)
    - Create a collection level update notification system
    - Create a collection level metadata search across TDSs
    - Look at re-architecting the TDS to “really” run on the cloud
        - TDS as a collection of autoscalable microservices vs monolithic web application. With our current resource levels, this is a stretch.

Relevant Metrics

NetCDF-Java

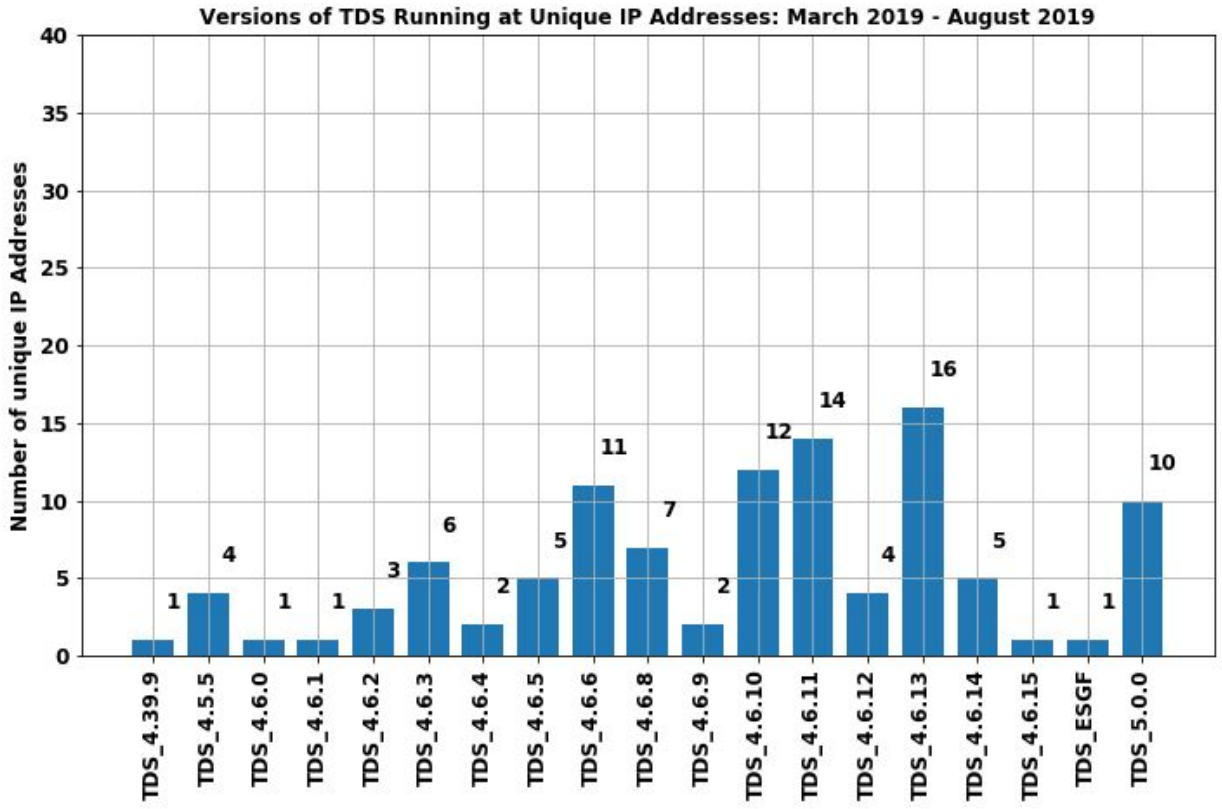
Recently we’ve found that the method by which we count download statistics for the netCDF-Java library has been excluding a significant source of downloads. Traditionally, users have downloaded a single jar (netcdfAll.jar) to use with their packages. However, the way Java developers have been consuming our libraries has changed significantly over the past several years, as many projects now rely on a build system to pull in just the components of the library

they need, and we have not taken that into account.

As an example, if we consider the netCDF-Java version 4.6.13 release, our current method for computing library downloads would show 1717 downloads over the past six months. If we take the previously unaccounted for downloads into account (for just the core component of the library — cdm.jar — not all components), the number of downloads for netCDF-Java version 4.6.13 over the past six months jumps to 16,049.

To put this new number in perspective, at our last meeting we reported that the total **yearly** downloads of **all versions** of netCDF-Java from 2018-02 — 2019-03 was 8389, and our new **six month** download figure for a **single version** of netCDF-Java nearly doubles that. As such, we will be changing the way the download statistics of netCDF-Java are computed to better reflect the community of developers who rely on the library. These changes will be reflected in the download metrics reported at our next meeting (Spring 2020).

THREDDS Data Server



We see that **10,436** unique IPs started up thredds from March 2019 through August 2019, **106** of which are publicly accessible servers. “Publicly accessible” means we could find them using common url patterns. For this plot, the version includes betas and snapshots, not just the official release of that version, for presentation simplicity.

This information is only known for servers running v4.5.3 and above. There are many reasons why these numbers are so different. The differences could be due to:

- People testing the TDS on their local machine, but not actually running a server (most likely the cause for the majority of the difference)
- A TDS running behind a proxy server may not be “seen” in this analysis as publicly reachable at the tested url pattern (e.g. <server>/thredds/catalog.xml). For example, a TDS running behind a proxy might be configured to respond to mytds.<server>/catalog.xml, and so our check for mytds.<server>/thredds/catalog.xml would not work. This can also happen if the TDS has been configured to use a different context without the use of a proxy server.
- The TDS server may be running behind a firewall that does not allow public access.
- A TDS running in the past is no longer running today.

Note 1: the vast majority of the publicly accessible servers are running v4.6.3 or above (v4.6.14 was the most current release during this period, and was released on 26 July 2019).

Note 2: there are some odd looking versions of the TDS being reported in the log files, such as TDS_4.39.x. It is likely these version numbers are actually generated by software that is being built on top of the TDS or applications that bundle the TDS as part of a deployment package.

Furthermore, of the **106** publically accessible servers, **64** 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…).

In the next six months, 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.

Prepared *September 2019*