2017 Q4 (July, August, September) Work Progress at a Glance

In Progress Completed Behind schedule

Tasks	Q1			Q2			Q 3			Q 4		
	1	2	3	4	5	6	7	8	9	10	11	12
1 Project Management												
1.a Annual ESGF F2F Conference											Agenda & abstracts	Agenda & abstracts
1.b Prioritize agency tasks												DOE, NASA, CRIM
1.e ESGF proposal										DOE proposal		DOE proposal accepted, R&D 100
1.c Strategic plan										Inter-agency strategic plan	DOE Strategic Plan	Inter-agency strategic plan
1.d Implementation plan											DOE implementation plan	Inter-agency implementation plan
2 Data Management (publication, replication, notification, feedback)												
2.a Publishing data to ESGF										Preparing for CMIP6, Obs4Mips, Input4MIPs		
2.a Official release of CMOR										CMOR v3.2.5	CMOR v3.2.6	CMOR v3.2.7
2.b Standardized distributed reanalysis data												
2.c Integrate heterogeneous data												Hydrology data
2.d Translate different archives and data stores												Lightweight data service
2.e Automated data replication & publication											Beta testing w/ DKRZ, CEDA, IPSL	
2.f Notification and feedback												
3 User Interface and Search												
3.a Enhance search												For CMIP6
3.b Enhance CoG UI												For CMIP6
4 Hardware and Network												
4.a New petabyte storage											Purchased 4PB	Integrated 4PB into CSS

4.b Network speed up												High-speed data syncing
4.c Data Transfer Nodes (DTNs)												Performance Tuning for CMIP6
5 Installation												
5.a Official release ongoing and new operations										ESGF major release v2.5	Minor release v2.5.13	Minor release v2.5.14
5.b New Python-base installation script												
5.c Installer UI												Prototype
6 Server-side computing												
6.a CoG UI integration											Development	Deploy & testing
6.b User metrics tracking												
6.c Docker support												
6.d Backend compute engines												EU's Ophida NASA's EDAS
7 Authorization and Authentication												
7.a Manage CA certificates												Managed for LLNL, NCCS, DIASJP, CMCC, others
8 Analysis & Visualization Tools												
8.a Official release of Community Data Analysis Tools (CDAT)												CDAT Minor release v2.12
8.b Community Data Management System (CDMS)										Reads non-CF compliant files	Integrated ViSUS Reader into CDMS	Port to Python 3
8.c Visualization Control System (VCS) Speed up												
8.d VCS additional graphics methods										Streamlines		Error bars
8.e Color maps												Static color maps for CMIP & E3SM
8.f Official release of the Community Diagnostics Package (CDP)										CDP major release v1.0	CDP minor release v1.1	
8.g CDP parallelism												Large scale

												parallelism
8.h CDP input parsing												JSON and CFG
8.i Visual Community Data Analysis Tools (vCDAT)										Development	Development	Development
8.j vCDAT file and dimension editor												
8.k vCDAT VTK-web												
8.l Animations & movies											List and prioritized animation	Animate hurricanes & human fingerprint
8.m test suite										Unit tests		
										Performance test		
8.n Tutorials and online support											Jupyter Notebook for Taylor Diagram	Jupyter Notebook for CDMS 101
9 Outreach												
9.a Publications										8 publications		
9.b Presentations											101 Tech Briefing	DOE ASCR NGNS PI Meeting
10 Machine Learning												
10.a CAM5 tropical cyclone tracking (w/ LBNL)												
10.b Tracking using LSTM												
10.c Connecting multiple locations												

Report summary of AIMS progress for the quarter period: July 1 – September 30, 2017

ESGF –

Members: Dean N. Williams (ESGF Chair and PI); Sasha Ames, Jason Boutte, James Crean, Charles Doutriaux, Cameron Harr, Matthew Harris, William Hill, Tony Hoang, Sookyung Kim, Ji-Woo Lee, Denis Nadeau, Jeffrey Painter, Zeshawn Shaheen, Argonne, JPL, Kitware

Management

- ESGF Scientific Focus Area Triennial Proposal Review, presentations can be seen online: https://esgf.llnl.gov/2017-DOE-Review.html
  - DOE **accepted** 3-year proposed for continued ESGF and some CDAT work
- ESGF is a finalist for a R&D 100 Award
- Presented ESGF and CDAT at LLNL's 101 Technical Briefing
- International collaborations with the Canadian Coalition of Scientific Gateways (ICSG) and EU Copernicus and H2020 Programme
- Preparation for the 7th Annual ESGF Face-to-Face Conference, to be held December 4 – 8 at the Sheraton Fisherman's Hotel in San Francisco, California

Publishing

- Continued publication efforts for CMIP6, E3SM, obs4MIPs, and input4MIPs
- Implemented lightweight data service for non-netCDF files making ESGF more general
- Beta testing automated data publication/replication with DKRZ, CEDA, and IPSL

Installation

- Official release to ESGF version (v2.5) in July 2017; minor releases of ESGF version (v2.5.13) in August 2017 and version (v2.5.14) in September 2017
- Produced new ESGF-Build script and ESGF Installer UI prototypes

Sever-side computing

- Finished CoG integration for CWT WPS
- Deploying Test CoG for CWT WPS on aims2.llnl.gov
- Added CWT WPS user tracking metrics
- Added CWT WPS process metrics (e.g., Average CPU time and memory usage)
- Updated CWT WPS Docker Compose to be released with ESGF version 3
- Exposed NASA's EDAS and EU's Ophidia as a backend compute engine along with CDAT

Security

- Managed the ANL ESGF CA and signed multiple host and intermediate CA certificates for LLNL, NCCS, DIASJP, CMCC, NCI and others

- Wrote ESGF Authentication web application to support both OpenID and OAuth2 authentication protocols for ESGF (see https://github.com/lukaszlacinski/esgf-auth/).

Machine Learning

- Database development of historical CAM5 Tropical cyclone tracking history with LBNL
- Brainstorming of Tropical Cyclone Tracking using LSTM, attention model with Samira Ebrahimi from Microsoft Research for the project collaborated with NERSC LBNL
- Apply "attention is all you need" model for the connecting multiple locations problem of tropical cyclones tracking

CDAT

- Official release of CDAT version (v2.12) in August
- Tutorials using Jupyter Notebook were added on the CDAT website, includes CDMS 101, Mathematical expression, Taylor Diagram, Logo control, color map, and animation
- Community Data Management System (CDMS)
  - Ported CDMS to Python 3; can now build with Python 2.7 or Python 3.0
  - Use of latest GLOBUS "myproxylogin" to get a certificates and set ".dodsrc" file appropriately (for ESGF authentication and authorization)
  - CDMS now allows reading of non-CF-1 compliant file (i.e. GPCP files)
  - Add Python 3 to continuous integration on github for TravisCI (Linux) and CircleCI (Mac OSX).
- Visualization Control System (VCS)
  - Increased CDAT's visualization engine (i.e., VCS) speed and add additional graphics methods (i.e., 1D plots error bars, streamlines fixes, and other enhancements)
  - Added unit test on TravisCI and CircleCI for vcsaddons
  - Visualized all available color maps (CDAT, matplotlib, PCMDI, etc.), including a restoration of IPCC AR4 color maps
- Setup site usage metrics for CDAT
  - Official release of the Community Diagnostics Package (CDP) (v1.1.0) in September
    - Implemented serial and parallel running diagnostics into the CDP and used by E3SM Diagnostics and PCMDI Metrics Package (PMP)
    - Added CDP command line tool and support for input parsing via JSON/CFG files
    - Add additional CDP documentation, examples, and tutorials to the website: https://acme-climate.github.io/acme_diags
- Visual Community Data Analysis Tools (vCDAT) pre-alpha release in July and August
  - Alpha release of vCDAT file and dimension explorer
  - Enhanced VTK-web visualization showing VCS plots in the vCDAT canvas

- Designed CDAT color map editor

- Animations and Movies
  - Initialized the creation of a set of animations and movies for LLNL's Climate Program and the Community (i.e., E3SM, CMIP6, input4MIPs, etc.)

Hardware

- Integrated new expansion hardware into Climate Storage System 03 (CSS03) file system, raising usable space from 1.6PB to 5.3 PB.
- Enhanced security on CSS I/O nodes by integrating them with LC's RSA OTP authentication scheme, instead of static passwords.
- Provided parallelized sync scripts to Jeff P. for high-speed data syncing from older file systems to CSS03.
- Tested transfer performance through aimsdtn nodes and worked with Eli Dart and Jeff P. to diagnose and tune performance issues. End result was one tuning parameter that should help aimsdtn performance and a focus on some European sites to tune their performance.

AIMS Publications –

- Beta release of Publication Hub (PubHub) for entering and tracking AIMS, ESGF, CDAT, etc. journal publications
- The Journal "Scientific Reports" released the article, "Tropospheric Warming Over the Past Two Decades", where AIMS member Jeff Painter was listed as a co-author.
- Sookyung Kim, Sasha Ames, Ji-Woo Lee, Chengzhu Zhang, Aaron C. Wilson, and Dean Williams", have an accepted paper to Climate Informatics (2017), entitled "Massive Scale Deep Learning for Detecting Extreme Climate Events", NCAR/TN536+PROC.
- Added input to LLNL's Science and Technology Review (S&TR) article, entitled "The Atmosphere around Climate Models". The article covers the decades of history involving LLNL's development of predictive climate models (i.e., from Chuck Leith's first ever color animated climate model in 1960 to present day E3SM efforts). The article is in rough draft form and is expected out in the spring of 2018.
- The article: "Use Cases of Computational Reproducibility for Scientific Workflows at Exascale" is a collaborative effort with PNNL, LLNL, and BNL laboratories. Sterling Baldwin is the second author on this provenance paper.
- Sookyung Kim, Sasha Ames, Ji-Woo Lee, Chengzhu Zhang, Aaron C. Wilson and Dean Williams", have an accepted paper to ICDM on IEEE (2017), entitled "Framework for Detection and Localization of Extreme Climate Event with Pixel Recursive Super Resolution". It will be presented at the 2017 Seventh Data Mining Workshop on Earth System Science (DMESS).
- Lee, J.-W., K. Sperber, P. Gleckler, C. Bonfils, and K. Taylor, 2017: Quantifying the Agreement Between Observed and Simulated Extratropical Modes of Interannual Variability. Climate Dynamics (in review)

Conferences –

- Sookyung Kim
  - September 20 - 22, 2017: Climate Informatics 2017
  - September 24 – 39, 2017: Heidelberg Laurate Forum 2017
- Ji-Woo Lee
  - August 10 -20, 2017: Visit South Korean institutions and universities to introduce ESGF, CDAT, PMP, and ongoing researches

Next Quarter –

Management

- Prepare for the 7th Annual ESGF F2F Conference, to be held December 4 – 8 at the Sheraton Fisherman's Hotel in San Francisco, California
- Develop ESGF priority, strategic, and implementation plans for the next three years

Installation

- Prepare for ESGF (v2.5.15) master release in October
- Prepare for ESGF major 3.0-alpha release in November
- Complete the ESGF-Build script and ESGF Installer UI prototypes for ESGF 3.0 major release

Sever-side computing

- Deploy CWT WPS on Federated Testbed
- Add parallel I/O through DASK distributed framework
- Add Ophidia and EDAS support to CWT WPS
- Add more CDAT processes (besides subsetting, regridding, and aggregation)
- Plan/Design CWT WPS federated operation (Security, Data management, Cache sharing)
- Plan/Design ViSUS integration for data access for CWT WPS processes

Machine Learning

- Apply LSTM based attention model for TC tracking given global scaled CAM5 reanalysis data
- Targeting to workshop publication for "Deep Learning for physical science" in NIPS 2018
- Keep working on collaborated project with NERSC LBNL

ESGF

- Continue to prepare for ESGF publication for CMIP6, E3SM, obs4MIPs, input4MIPs

CDAT

- CDAT v3.0 candidate release, to include the port to Python v3.0

- Additional Advance CDAT scientific examples and tutorials with Jupyter Notebook, such as Blue Marble background image control and useful CDMS functions
- CDMS
  - Compile ESMF and ESMPy on Mac OSX
  - Fix ESGF authentication on OSX (netCDF issue)
  - Continue work on CDMS documentation in Sphynx.
- VCDAT
  - Release beta version of vCDAT
  - Implement the color map editor UI for vCDAT.
- CDP
  - Complete CDP journal publication article
- Animations and movies
  - Create animations for LLNL and the climate community that include idealized ocean simulations, water vapor and SST changes in the Gulf of Mexico, superimposing the observed temperature record over a scrolling background of CMIP control time-series, and 3-dimensional high-resolution cloud modeling of shallow cumulus

Publications

- Official release of Publication Hub (PubHub) (v1.0) for entering and tracking AIMS, ESGF, CDAT, etc. journal publications, posters, presentations, software, etc.
- Continue beta testing automated data publication/replication with DKRZ, CEDA, and IPSL
- Prepare to publish CMIP6, E3SM, and other data sets
- Design client app to manage update notification

Hardware

- Investigate new compute cluster, GPU compute node for server-side computing
- Further enhance security and administration automation of CSS03 I/O nodes.