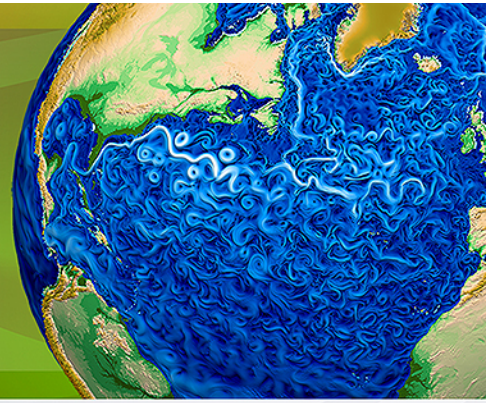


V1 Process for data management, infrastructure and diagnostics

Ben Mayer, Sasha Ames, Rachana Ananthakrishnan, Raju Bibi, John Harney, Charles Doutriaux, Aashish Chaudhary, Jeff Painter, Brian Smith, James McEnerney, Charlie Zender, Jerry Potter

ACME Workflow Team Leads

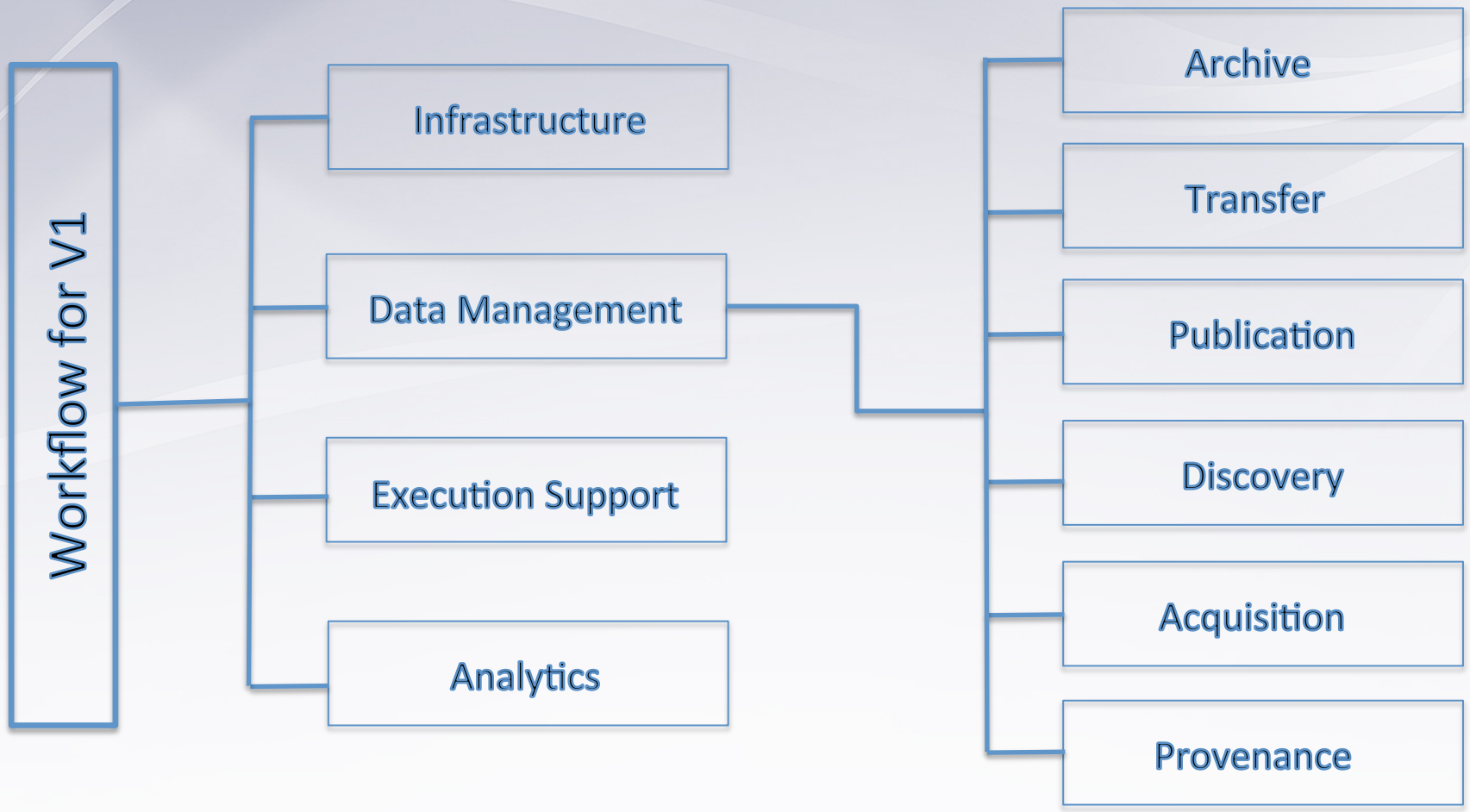
Dean N. Williams and Valentine Anantharaj

ACME Workflow Group Leads

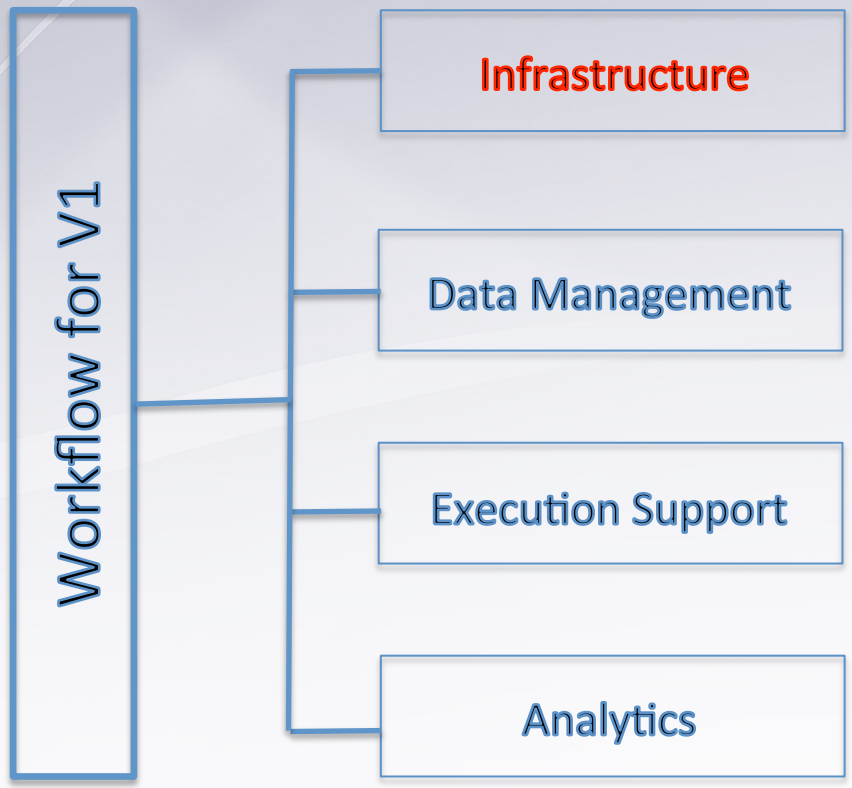
Rockville, MD

June 7-9, 2016

Manual workflow process for V1

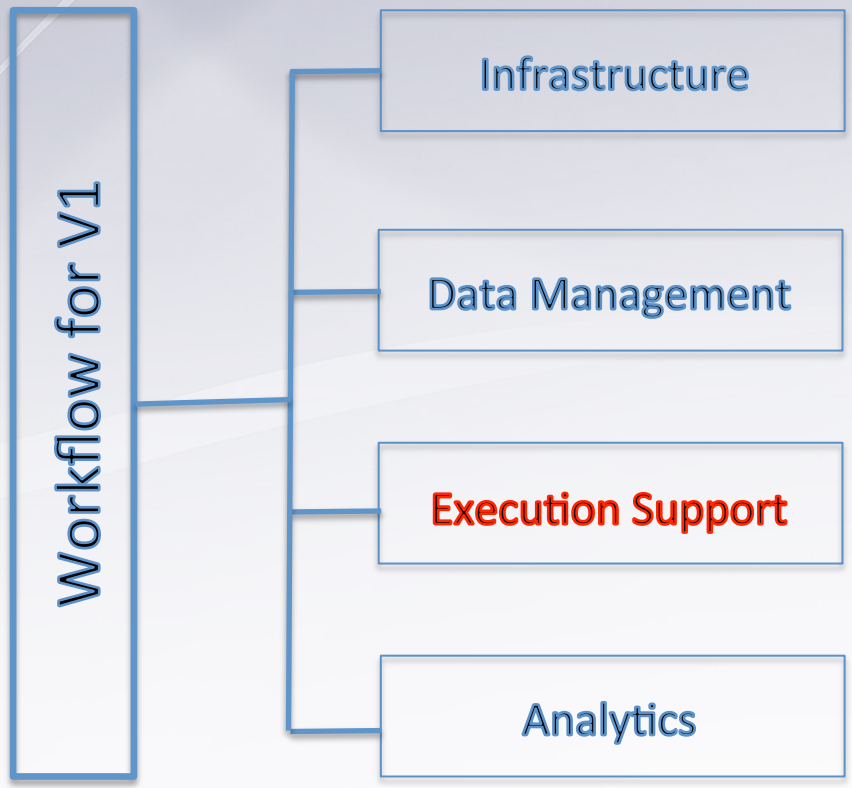


Manual workflow process for V1



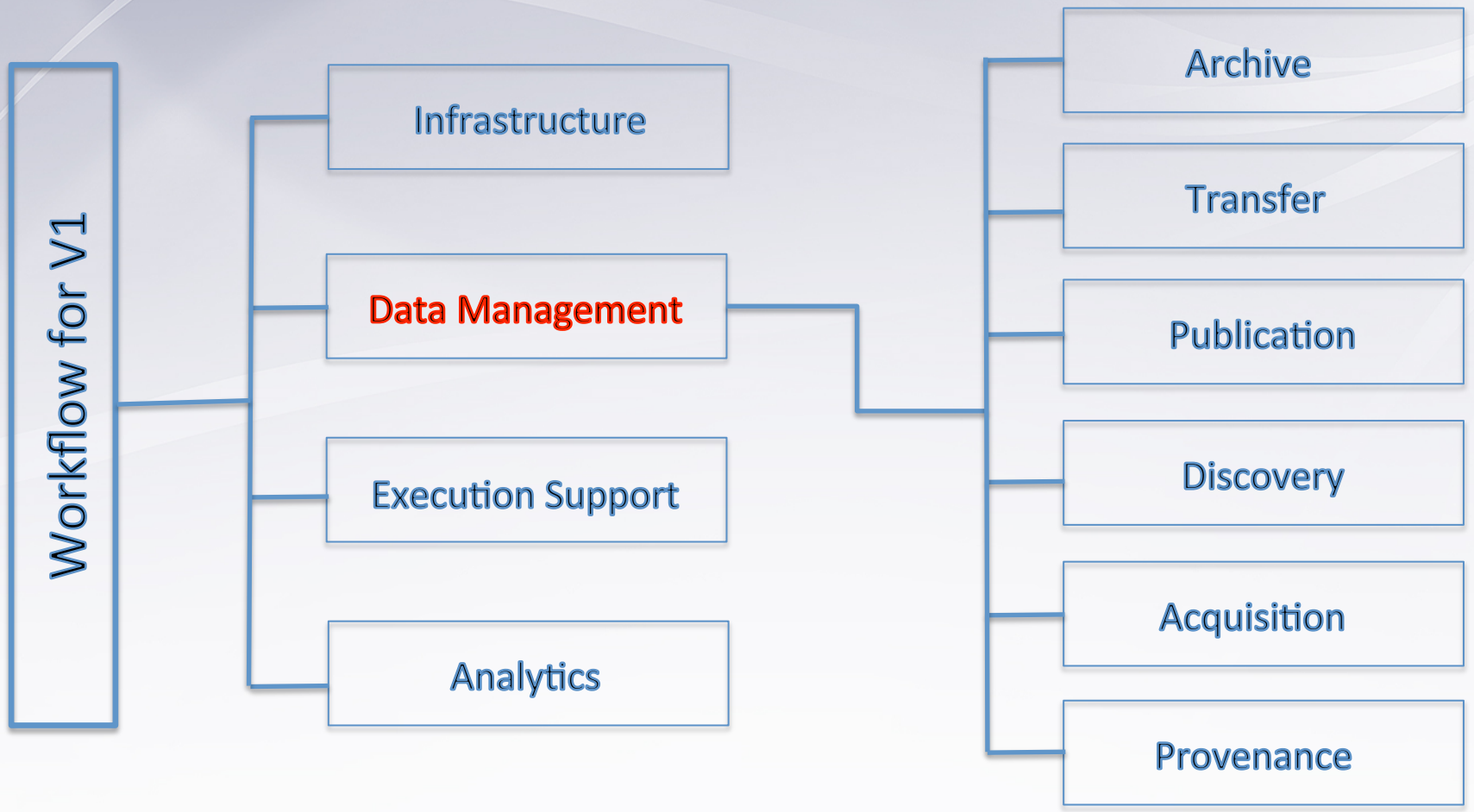
- Facilities for data analysis and management
  - NERSC, OLCF, ALCF, ORNL/ CADES, LLNL/AIMS, condo investment (TBD)
- Services
  - Short-term and longer term archive; data transfer; compute and analysis; publication; provenance; discovery, access and collaboration
- ACME ESGF nodes
  - ORNL/CADES; LLNL/AIMS; NERSC & ALCF (future)

Manual workflow process for V1

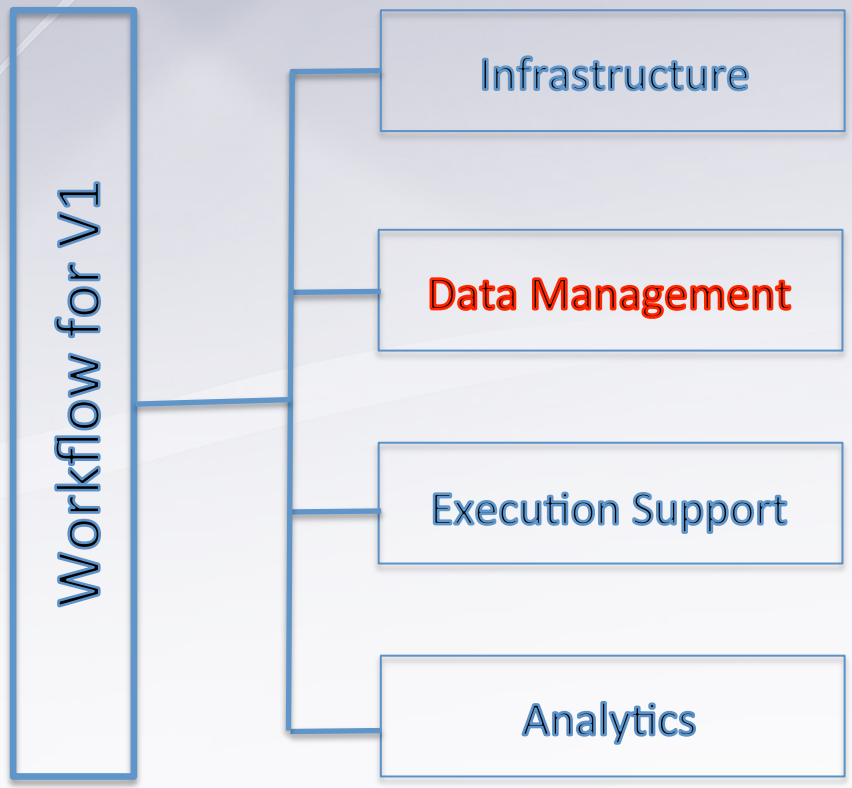


- Manual workflow will be managed by individual science teams with workflow team assisting as necessary
- Will provide guidance on best practices
- Independent effort to help with recovery & improve resiliency
- Automated diagnostics (TBD)

Manual workflow process for V1

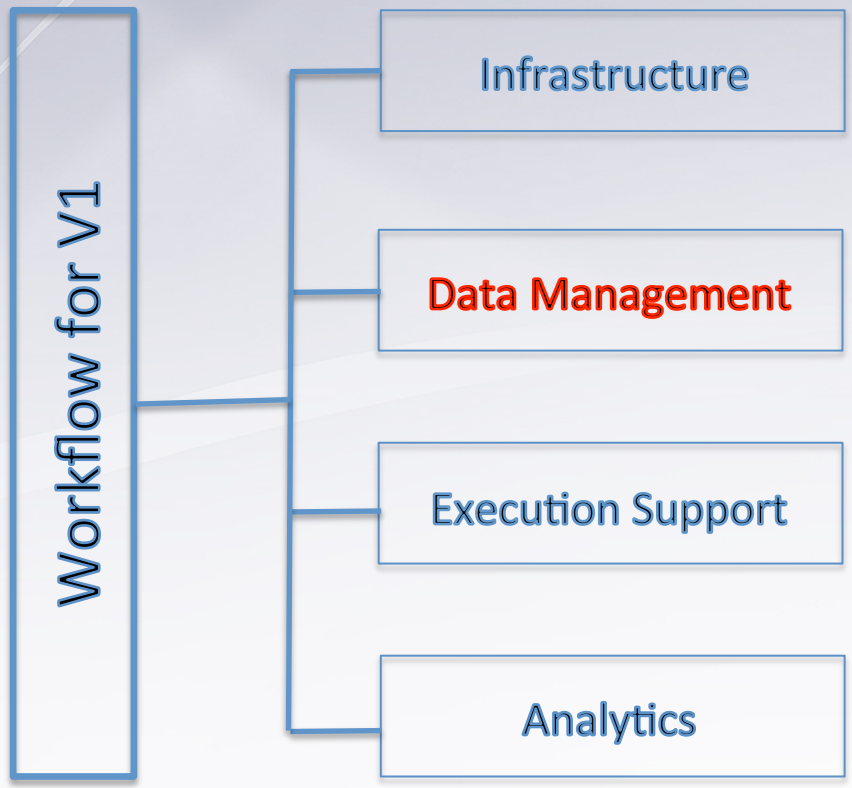


Manual workflow process for V1



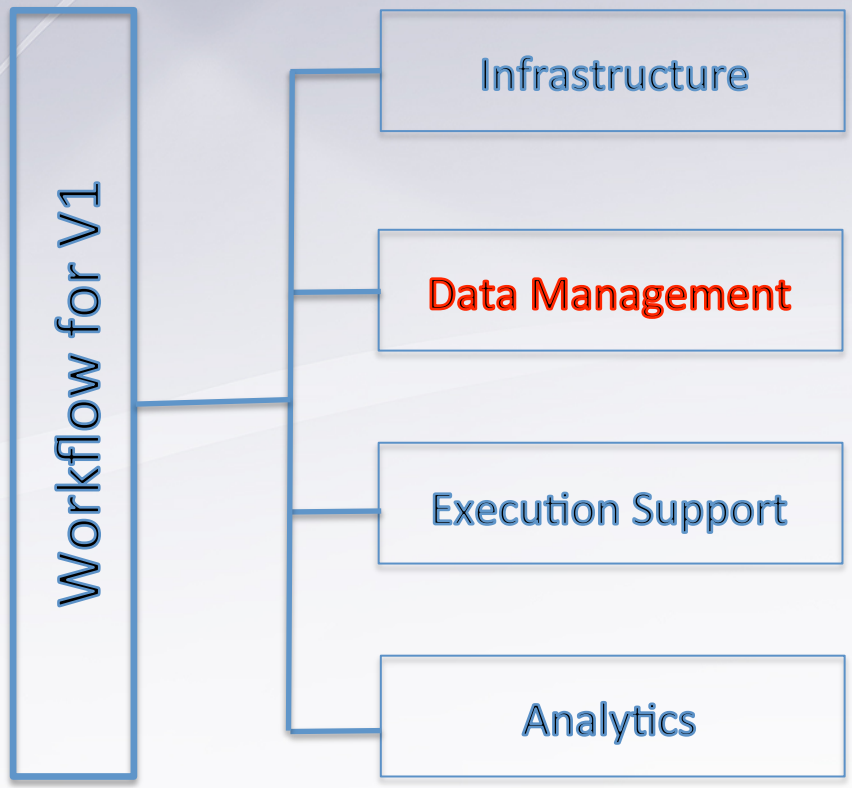
- Archive
  - ACME HPSS (mass store) utilities to store & retrieve files; use python wrapper around system interface.
  - Native model output
    - archived at source
  - Climatologies
    - created and published incrementally (say every 5 years); and published via federated set of ACME ESGF nodes.

Manual workflow process for V1



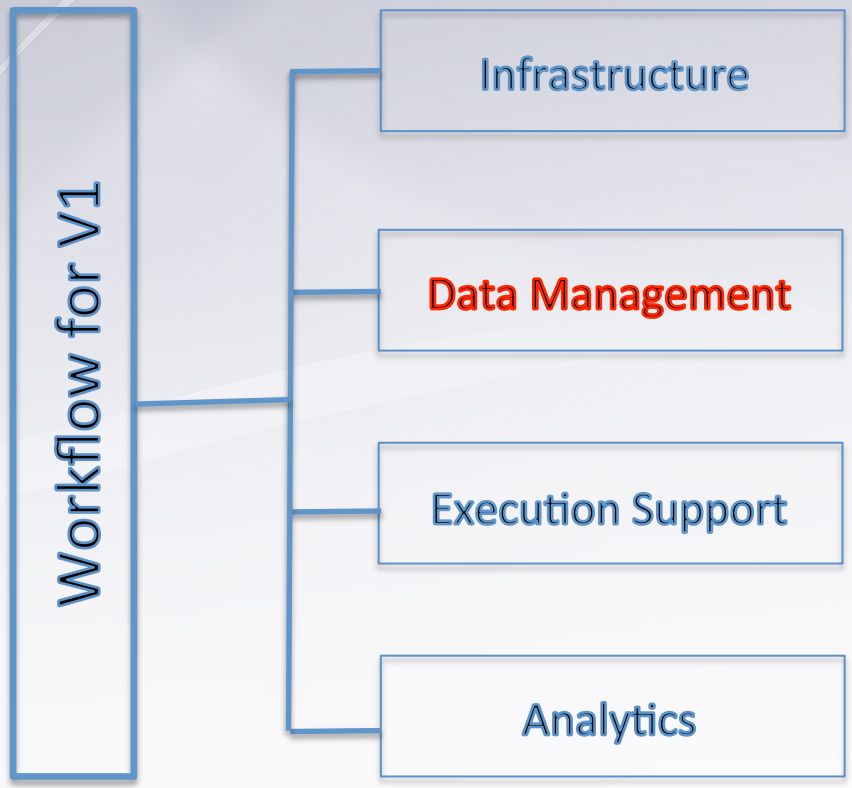
- Transfer
  - Globus transfer utilities
    - Endpoints available at all key sites
  - Able to transfer 1 PB / week across LCF
  - Incremental transfers possible
  - REST API available for integration with other scripts
  - Globus SDK being tested for ACME

Manual workflow process for V1



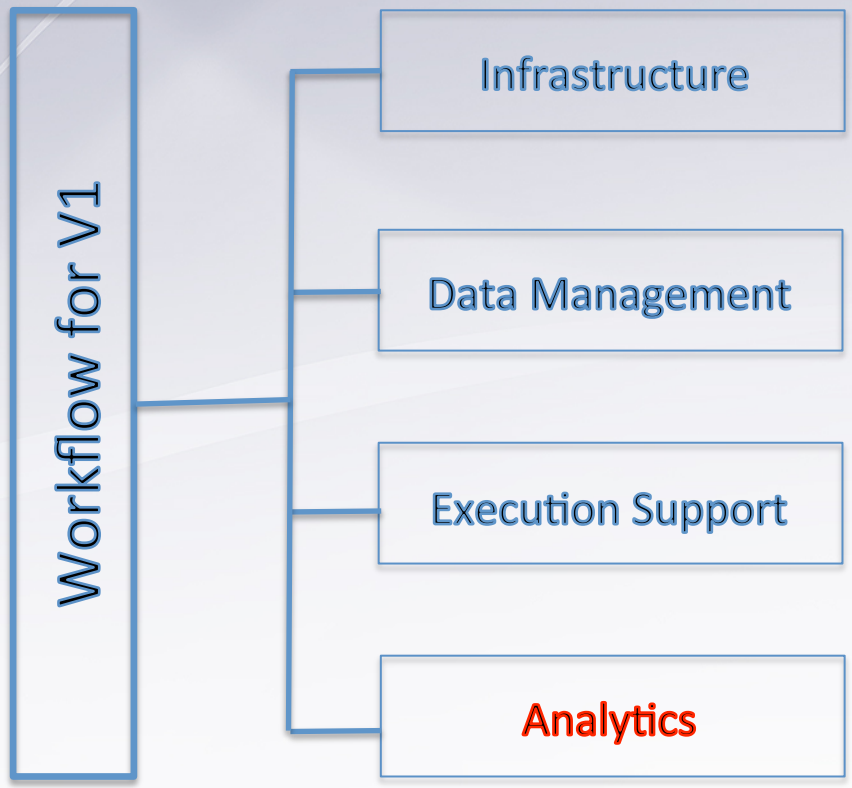
- Publication
  - Globus publisher tested & deployed
  - Tutorial available online
  - Publisher API
  - Work in progress to publish files residing in HPSS archive (tested at NERSC; in progress at OLCF)

Manual workflow process for V1



- ESGF
  - Nodes available at ORNL/ CADES; LLNL/AIMS; NERSC & ALCF (future)
  - Climatology and 2D time series files are published & have direct access
  - Work in progress to publish files residing in HPSS archive (tested at NERSC; in progress at OLCF)

Manual workflow process for V1



- NCO (remap; grid generation)
- UV-CDAT (I/O, analysis, regridding, plotting, etc.)
- UVCmetrics (AMWG, LMWG)
- Classic Viewer (.png, .svg)
- Automated diagnostics *(to be supported)*

Archive Infrastructure: OLCF HPSS ~2013

- 2 x DDN SFA10K (10 GB/s ea)
  - 2 PB disk cache capacity
- 1 x NetApp E5560
  - Files < 16 MB
  - 330 TB of capacity, 150 TB utilized
- 8 Disk movers
  - Responsible for data ingress and migration to tape
  - 10 GbE Networking (~1GB/s each)
- 40 GbE Network switch to Disk movers
- 120 Oracle T10K-{A,B,C} tape drives
  - ~150 MB/s each
- 32 Oracle T10K-D tape drives
  - 252 MB/s each

Archive Infrastructure: OLCF HPSS ~2014

Disk

- 2 x DDN SFA10K (10 GB/s ea)
  - 2 PB disk cache capacity
- 3 x DDN SFA12K (40 GB/s)
  - ~12 PB raw capacity for cache
- 1 x NetApp E5560
  - Files < 16 MB
  - 330 TB of capacity, 150 TB utilized
- 20 Disk movers
  - Responsible for data ingress and migration to tape
  - 40 GbE Networking

Tape

- 120 Oracle T10K-{A,B,C} tape drives
  - ~150 MB/s each
- 32 Oracle T10K-D tape drives
  - 252 MB/s each

Network

- 2 x Arista 7508 40 GbE switches
  - Connectivity to Disk movers
  - Connectivity between Disk and Tape movers
  - 13 x 100 GbE ISL's

Archive Infrastructure: OLCF HPSS ~2016

Disk

- 5 x DDN SFA12K (40 GB/s)
  - ~20 PB raw capacity for cache
- 1 x NetApp E5560
  - Files < 16 MB
  - 330 TB of capacity, 150 TB utilized
- 40 Disk movers
  - 40 GbE Networking

Tape

- 112 Oracle T10K-D tape drives
  - 252 MB/s each

Metadata

- NetApp EF560
  - SAS connected; All Flash

Network

- 2 x Arista 7508 40 GbE switches
  - Connectivity to Disk movers
  - Connectivity between Disk and Tape movers
  - 13 x 100 GbE ISL's

Archive Infrastructure: OLCF HPSS ~2013

- 2 x DDN SFA10K (10 GB/s ea)
  - 2 PB disk cache capacity
- 1 x NetApp E5560
  - Files < 16 MB
  - 330 TB of capacity, 150 TB utilized
- 8 Disk movers
  - Responsible for data ingress and migration to tape
  - 10 GbE Networking (~1GB/s each)
- 40 GbE Network switch to Disk movers
- 120 Oracle T10K-{A,B,C} tape drives
  - ~150 MB/s each
- 32 Oracle T10K-D tape drives
  - 252 MB/s each

Archive Infrastructure: OLCF HPSS ~2016

Courtesy: Jason Hill, OLCF

Disk

- 5 x DDN SFA12K (40 GB/s)
  - ~20 PB raw capacity for cache
- 1 x NetApp E5560
  - Files < 16 MB
  - 330 TB of capacity, 150 TB utilized
- 40 Disk movers
  - 40 GbE Networking

Tape

- 112 Oracle T10K-D tape drives
  - 252 MB/s each

Metadata

- NetApp EF560
  - SAS connected; All Flash

Network

- 2 x Arista 7508 40 GbE switches
  - Connectivity to Disk movers
  - Connectivity between Disk and Tape movers
  - 13 x 100 GbE ISL's

Performance analysis (2013)

Courtesy: Jason Hill, OLCF

- In 2013 asked how long to put in 1PB of data to HPSS
  - Getting it to disk took 21 days.
  - Migration to tape took another 35 days.
  - Single directory, multiple files
    - Processed serially from single node
- This seems inefficient, right?
- Drive investments in hardware and user experience

Performance Analysis (2016)

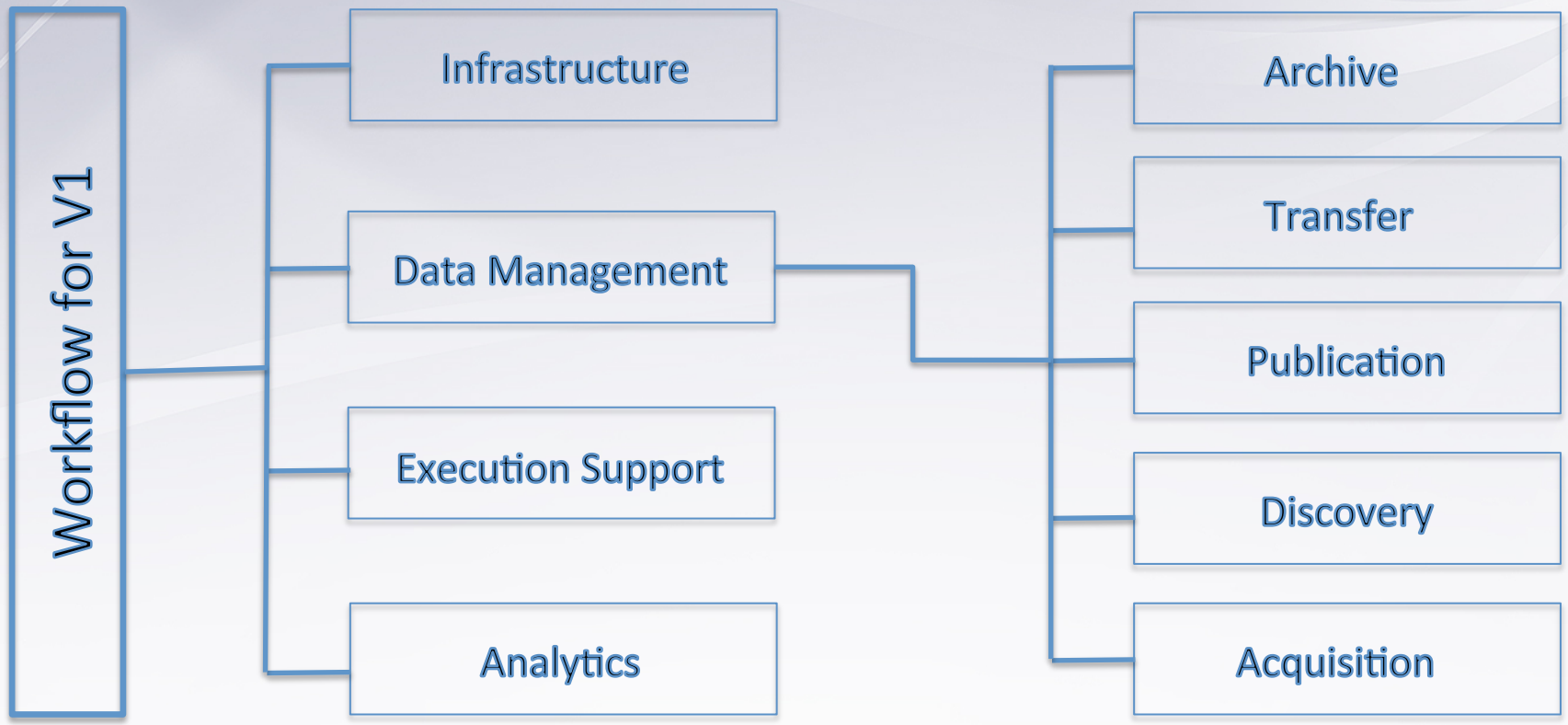
Courtesy: Jason Hill, OLCF

- In February, a user moved 1.3 PB of data into HPSS in 8 days.
  - Utilized DTN scheduled queue
  - Requested extended wall time (not needed in the end)
  - Used HTAR utility
- Migration to tape took ~12 additional days to complete
- Significant improvement over 2013 data point
- Other success stories out there

Date	# Files	TB Transferred	Daily Total
2/10/16	10	155	162
2/11/16	8	125	126
2/12/16	14	218	220
2/13/16	12	186	189
2/14/16	18	279	281
2/15/16	12	186	199
2/16/16	14	218	279
		1367	1456


ACME
Accelerated Climate Modeling
for Energy

Manual workflow process for V1



Workflow new features

Process Flow

- The Pegasus workflow manager is being tested at OLCF and NERSC
- The ACME configure, build, and run process under Pegasus is working at OLCF and NERSC
- The HPSS storage wrapping software is completed and being tested
- Service with REST API for programmatic access
- Web front-end for users to browses and prepare and review models
- Review and create visualizations with CDATWeb
- Refining technical requirements for ACME Workflow Integration Framework

Data Management

- Set up additional ACME ESGF nodes and work environment:
  - LLNL
- Publish additional data from model runs
- Track a few outstanding issues or limitations, such as:
- Need additional storage from the ONRL's CADES storage infrastructure
- Work with publication team to allow individual ACME scientist to publish data to the ACME archive

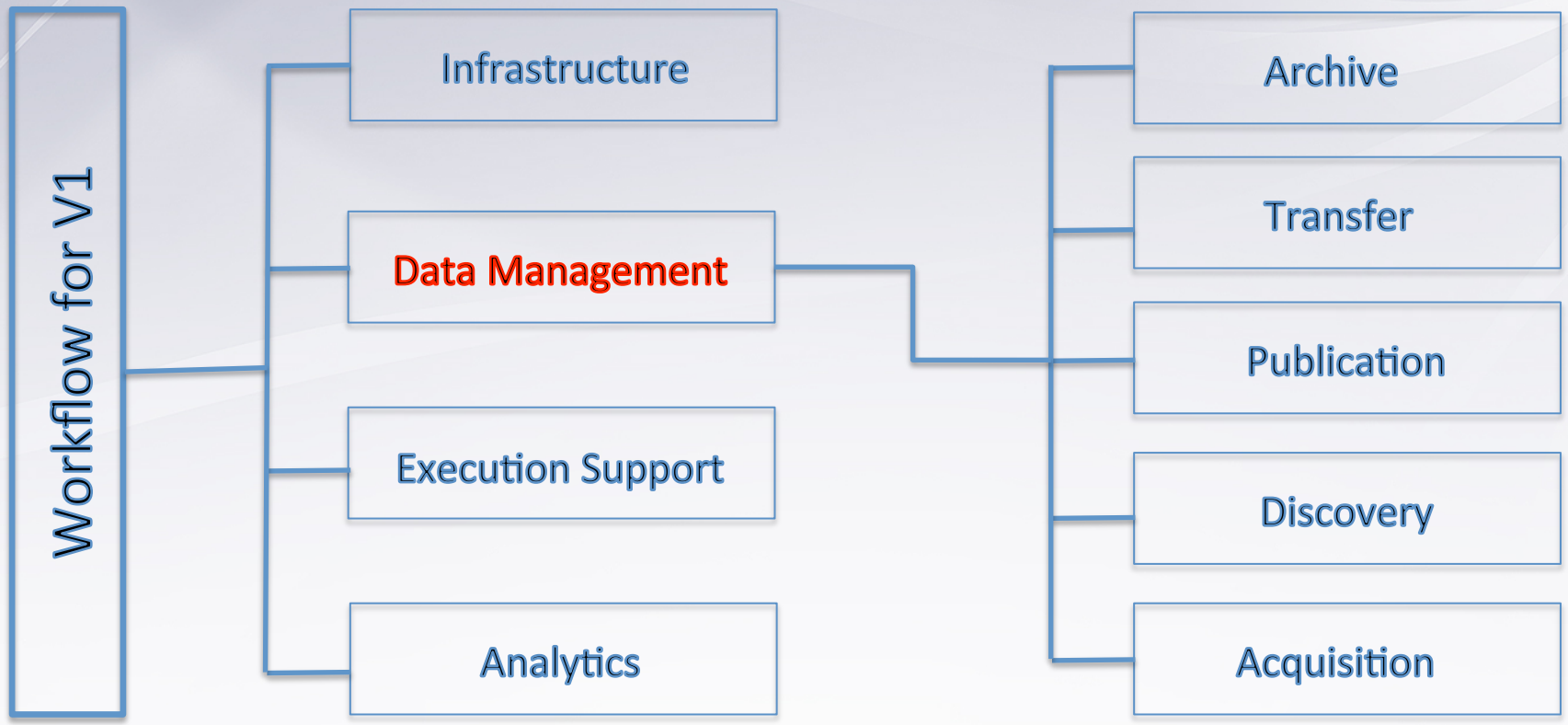
Publication

- Moved authentication from Globus Nexus to Globus Authentication
- Improvements to user interface, e.g.,
  - remembering last selected facet
  - rearranged widgets, etc.
- Globus endpoints
  - Users can authenticate to ORNL/OLCF Rhea and Titan Globus endpoints using OSG certificates (instead of only PIN +SecureRSA) what makes scripting data transfer from/to Rhea and Titan possible now.
- Webinar tutorials

NCO

- Works uniformly on all ACME, CESM, and observation components
- More accurate and Parallel mode ~25x faster than AMWG
- Remapper and grid/map-generator ncremap:
  - Infers grids from SCRUD (Swath, Curvilinear, Rectangular, Unstructured Data), or creates rectangular grids *de novo,* and remaps data in parallel
  - Generates weights with *ESMF_RegridWeightGen* or *TempestRemap*

Manual workflow process for V1



Workflow new features cont.

UV-CDAT

- Anaconda build
- Linux, OSX
- Comes with: Matplotlib, VCS, CDMS2, cdtime, NumPy, iPython, etc.
- Interactive point selection
- Continue work on cleaner API
- UV-CDAT new user's documentation

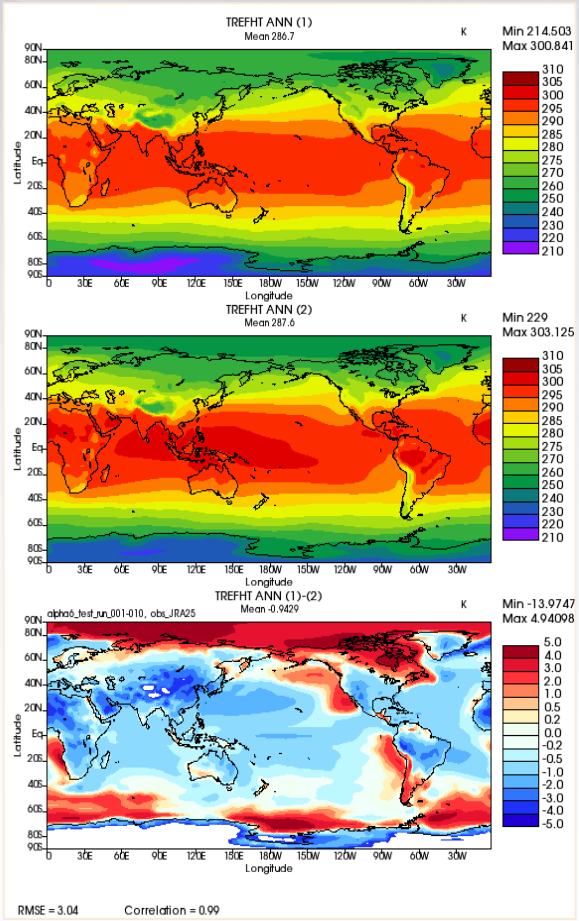
UVCmetrics

- AMWG reproduction
- Output customizable for ACME needs
- Generates own "climo" files or uses those generated by NCO
- Diags.py
- Metadiags.py
- Built-in Viewer
- MPI testing (mcenerney paper)
- Vector (.svg) and Raster (.png) graphical output
- ACME model variable name on plots

Test suite

- Continuous integration and code testing using CMake/CDash (https://cmake.org) and Buildbot (http://buildbot.net)
- Increased overall code coverage and new tests

DIAGS



AMWG