ESGF Executive Committee Findings

- Dean N. Williams (Chair, DOE)
- Michael Lautenschlager (Co-Chair, DKRZ)
- Luca Cinquini (NASA/NOAA)
- Sébastien Denvil (IPSL)
- Robert Ferraro (NASA)
- Daniel Duffy (NASA)
- V. Balaji (NOAA)
- Claire Trenham (NCI)

ESGF F2F findings from conference (readiness for CMIP6)

- **Estimated Size of CMIP6:** The lack of solid agreed-upon figures for how much CMIP6 data there is and when it is coming makes tools and infrastructure planning difficult. It is clear that enough data volume will be stored at, analyzed at, and/or transmitted by various nodes that changes are needed in the infrastructure.
- **Tier 1 and Tier 2 Nodes:** It is clear that enough data volume will be stored at, analyzed at, and/or transmitted by various nodes that changes are needed in the infrastructure.
- **Software Security:** Overall, it seems like ESGF has recovered from security challenges in 2015, but we are not out of the woods yet. There's going to be a lot of last-minute work to get the desired services and features ready in time for CMIP6. We need to get the right folks in place for a robust/routine software security scan (to include **risk assessment of code base**). Security process is in place – initial concern: missing a manifest file for each release.
- **Server-side Computing:** There's a desire to make it easier for scientists to download and use just the data portions they need rather than bulky data sets (i.e., **data aggregation, sub-setting, regridding**). There is also a strong interest in moving some of the computing to where the data is stored, but these tools are really in testing phase and need **resource management strategies** along with determining the needed **hardware** and which sites will deploy the capabilities (circle back to Tier 1 vs. Tier 2 nodes). Additionally there is a clear need for cross-validation across the many implementations of the WPS API.

ESGF F2F findings from conference (readiness for CMIP6)

- **Search and Metadata:** There is a strong interest in search and metadata and in enhancing those capabilities. What they are TBD.
- **Provenance Capture:** Stress the need for maturing our capabilities to capture provenance. Start a new team (Provenance Capture/integration/usability Working Team – Eric Stephan).
- **Metrics:** Dashboard team will be able to display hard metrics in next year's release of ESGF, in time for CMIP6 (e.g., number of users, number of downloads, size of current archive, etc.). However, must incorporate different metrics into the dashboard such as how many subsets or averages (**computing resources**) took place and on which machines. Must also track resource allocations used by users. More **surveys** should be helpful in capturing metrics as well.
- **Modularity:** There's definitely an interest in making some of the open-source tools and interfaces different groups are developing more broadly available to the ESGF community (i.e., modularity and APIs). Birdhouse is an example of using components of ESGF tools.
- **Installation/Docker:** Installing ESGF software continues to be a challenge, various working teams are helping to addressing this. Working to remove the monolithic installation to work with something similar to Docker (does not have to be Docker). Docker has security concerns. Zed and Luca think all security concerns can be addressed (i.e. monitoring and timely updates).

ESGF F2F findings from conference (readiness for CMIP6)

- **Data Replication:** Plans and progress for data replication are not where we want/expect them to be for CMIP6. Still in test phase.
- **Network:** Network speeds between some nodes are also not where we want/expect them to be, either. Seems like this might not be a priority, given all the other preparations for CMIP groups are making.
- **PIDs:** Interest in versioning/citation tools and services (PIDs, errata, "unpublish" features, etc.) remain strong, and progress is being made. Should be ready for CMIP6
- **Architecture Design:** Need overarching detailed design of current and future ESGF architecture—show connectivity.
- **Test Platform:** There's strong interest in a ESGF test platform to test how tools perform on different nodes and performance between nodes before new tools and features go live. That is, every new development runs through the test environment before going live—internal back end test environment.
- **Notification:** Automatic process to notify the user community when data has been changed, added, etc.
- **Documentation:** Documentation are need for all components.
- **Cloud or not to Cloud:** There's a strong interest in cloud storage and services, but there are question in using it: Who owns the data? Who pays for the storage? How will cloud storage options evolve over the next few years? Etc.