Publication Working Team:
Progress Update and Future Roadmap

ESGF 7th F2F Conference
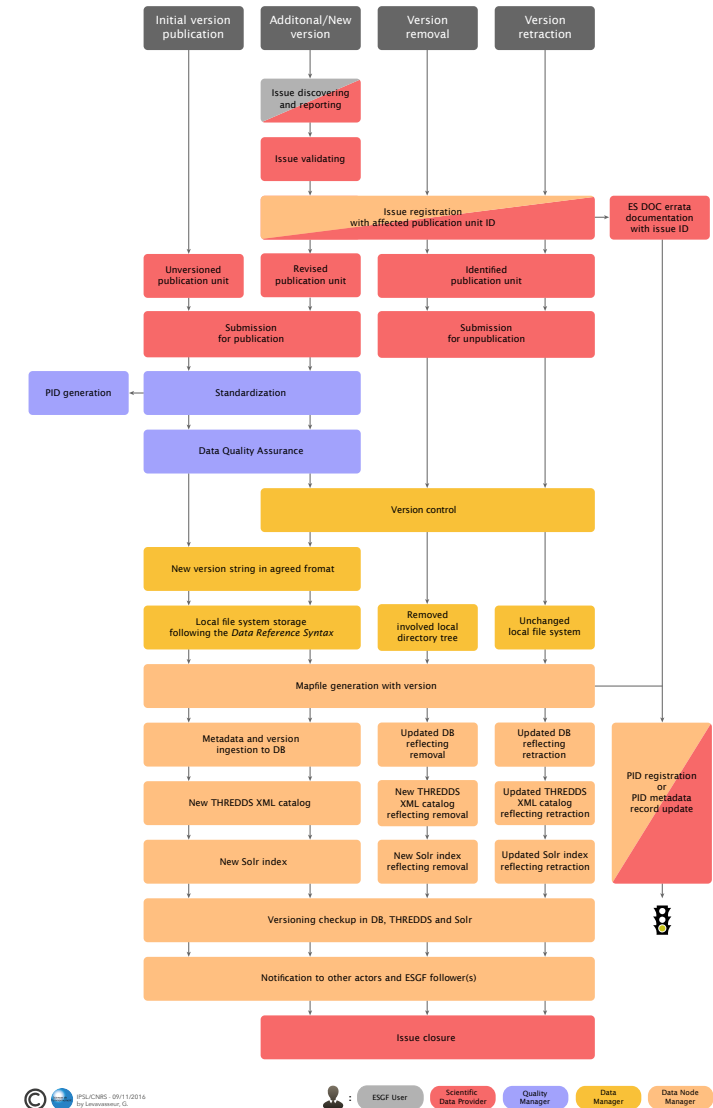San Francisco, CA,USA, December 2017

Team lead: Sasha Ames (LLNL)

Lukasz Lacinski, Rick Wagner(ANL), Guillaume Levavasseur (IPSL), Ag Stevens, Alan Iwi. (CEDA), Katharina Berger, Hans Rathmun (DKRZ), Serguei Nikonov (GFDL), Luca Cinquini (JPL/ESRL), Jon Smilie (NCI/ANU)

Publication Working Team Overview

- Manage the esg-publisher repo
  - "legacy code" we keep updating
- Quality control for publication process
- Configuration
  - Implementing controlled vocabulary
- Publication process
- Publishing operations
- Documentation
- Site support
- Ingestion service for remote GUI/scripted publication

- Conda-based esg-publisher released in v2.5
  - new environment (esg.env deprecated, no longer used for publishing)
  - Publisher issues was #1 contributor to lengthy release timeframe (May-October)
- esgprep fetch-ini » simplified
  - Simply download or update your local esg.\<project\>.ini files from the official GitHub repo
- esgprep drs » now available
  - Migrate your CMORized files into the corresponding project DRS
  - Manage the data versioning easily
- esgprep mapfile » refactored
  - Performance improved
  - Choose your mapfile management

- PID service running in "—test" mode
  - Datasets populate in DKRZ server
  - Citation urls
- PrePARE updated and tested
  - Some iteration
  - Challenge in configuring for particular data-specs version
    - read from data file vs manual configure
    - IPSL contributes custom procedure
- IPSL and GFDL test model output published in test federation
- CDF2CIM integrated into the publisher for ES-DOC creation triggered by publisher (CEDA)

More 2017 accomplishments

- NCI republication of cmip5, cordex, pmip3, geomip
  - Used esgprep 2.7.9
  - Complete reorganization – correct previous DRS
- input4MIPs
  - exercising multiple variable issues, list facets
  - nonNetCDF data
  - PID testing (non default unlike CMIP6)
- More new projects: eg. PCMDI – PSIPPS, E3SM Process ordering
- Use of metadata update API to make SOLR changes without full republish
- Automated publication test
- Switch to SOLR REST API (testing now)

- esg-publisher refresh
  - Python 3.0
  - Switch to requests
  - Notification service hook ?
  - Parallelization (using dask)
  - Switch CV part of .ini to service (an option)
- Ingest service (ANL) revisit and expansion to more projects
- HPSS – LBL is resuming the base-jumper effort
- Finalize cmip5 replica republication
  - LLNL has more unpublished
  - CEDA and LLNL to flag down missing originals and correct
  - (plan republication as one variable-per-dataset?)
- Republish input4MIPs with consistent DRS, single variable, PID and citation url
- 31 GitHub issues – prioritize and close outdated ones
- **Start CMIP6 publishing**

Still needed for CMIP6…?

- PrePARE performance
- CDF2CIM readiness – testing adequate?
- DOIs will be available? End-to-end demo
- ES-DOC
- model_cohort updates – use API
- CV stability
  - all groups, etc. finally accounted for
- Adequate documentation
- Need vYYYYMMDD_xx format?
- Production rehearsal planning?

- Mailing list: esgf-pwt@llnl.gov

- Code: https://github.com/ESGF/esg-publisher

- Documentation: https://github.com/ESGF/esgf.github.io/esg-publisher

- Bi-weekly telcos - Wednesday 8 AM PT / 1700 CET