

Jeffrey Painter
Supported Projects: ESGF, CF Conventions

Quarterly Report for January 2018 – March 2018 (Q2, FY18)

Note that I am working part-time.

This quarter's accomplishments:

- I improved my script to migrate information from the old replication database to the new Synda database. Then I ran and tested it. We now have all features needed, and I now routinely use the migrated information in replication. Most important, Synda will not download data which was downloaded with the old system several years ago.
- I patched our copy of Synda so that when a download fails on one server, Synda will try to get the data from other servers. I made a working version available to the Synda development team at IPSL, as a pull request.
- Note that the above two items, plus some smaller tasks, have made it possible to replicate vast amounts of data with a single command (at least several tens of thousands of files). This will be less demanding on user time than the previous system, or the standard Synda.
- Other replication software enhancements: automatically restart a failed daemon, reduce the priority of files retried after errors, work around a problem where ESGF index nodes return urls containing 'None', prevent Synda crashes when two files have been assigned the same download path.
- Production-scale replication of CMIP-5 data: several tens of thousands of files (in a few commands) completing or archive for some popular experiments with monthly atmospheric data.
- Test replication of CMIP-6 data: all available IPSL and GFDL data downloaded. CNRM and CSIRO data have problems, and may be easier to get with another Synda enhancement, in progress.
- I inspected the data on CSS-02 whose non-standard directory paths had prevented automated transfer to CSS-03. I transferred everything worth keeping to CSS-03.

Plans for next quarter:

- Look for a way to ensure that Synda only downloads the latest version of each file. The "latest=true" specification doesn't work very well.
- Inspect data on CSS-01 and transfer all worthwhile data to CSS-03 (this has already been done for data in normal directory paths).
- Re-organize data transferred from CSS-01 and CSS-02 into DRS-compliant paths so that it can be published.
- Continue test replication of CMIP-6 data, and evolve to production replication. This may involve further Synda enhancements.
- Continue replicating CMIP-5 data, until we have all the data we want.
- As needed, support other data management activities.

- The CF Conventions documentation will need a low level of support indefinitely. In particular, the email system needs some filters to send user comments to the mailing list, and system messages to the maintainer.