

[Data reorganization for optimal time series data access, analysis, and visualization](#) [1]

Submitted by Wteng on Thu, 2012-12-27 12:57 **Event:** [Winter Meeting 2013](#) [2]

Abstract:

The way data are archived is often not optimal for their access by many user communities (e.g., hydrological), particularly if the data volumes and/or number of data files are large. The number of data records of a non-static data set generally increases with time. Therefore, most data sets are commonly archived by time steps, one step per file, often containing multiple variables. However, many research and application efforts need time series data for a given geographical location or area, i.e., a data organization that is orthogonal to the way the data are archived. The retrieval of a time series of the entire temporal coverage of a data set for a single variable at a single data point, in an optimal way, is an important and longstanding challenge, especially for large science data sets (i.e., with volumes greater than 100 GB). Two examples of such large data sets are the North American Land Data Assimilation System (NLDAS) and Global Land Data Assimilation System (GLDAS), archived at the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; Hydrology Data Holdings Portal, <http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings> [3]). To date, the NLDAS data set, hourly $0.125^\circ \times 0.125^\circ$ from Jan. 1, 1979 to present, has a total volume greater than 3 TB (compressed). The GLDAS data set, 3-hourly and monthly $0.25^\circ \times 0.25^\circ$ and $1.0^\circ \times 1.0^\circ$ Jan. 1948 to present, has a total volume greater than 1 TB (compressed). Both data sets are accessible, in the archived time step format, via several convenient methods, including Mirador search and download (<http://mirador.gsfc.nasa.gov/> [4]), GrADS Data Server (GDS; <http://hydro1.sci.gsfc.nasa.gov/dods/> [5]), direct FTP (<ftp://hydro1.sci.gsfc.nasa.gov/data/s4pa/> [6]), and Giovanni Online Visualization and Analysis (<http://disc.sci.gsfc.nasa.gov/giovanni> [7]). However, users who need long time series currently have no efficient way to retrieve them. Continuing a longstanding tradition of facilitating data access, analysis, and visualization that contribute to knowledge discovery from large science data sets, the GES DISC recently begun a NASA ACCESS-funded project to, in part, optimally reorganize selected large data sets for access and use by the hydrological user community. This presentation discusses the implementation of reorganizing data, including data processing (parameter and spatial subsetting), metadata and file structure of reorganized time series data (true "Data Rod," single variable, single grid point, and entire data range per file), and production and quality control. The reorganized time series data will be integrated into several broadly used data tools, such as NASA Giovanni and those provided by CUAHSI HIS (<http://his.cuahsi.org/> [8]) and EPA BASINS (<http://water.epa.gov/scitech/datait/models/basins/> [9]), as well as accessible via direct FTP, along with documentation and sample reading software. The data reorganization is initially, as part of the project, applied to selected popular hydrology-related parameters, with other parameters to be added, as resources permit.

Author(s):

Name: [Hualan Rui](#) [10]
Organization(s): [NASA GES DISC](#) [11],
[ADNET Systems Inc](#) [12]
Email: hualan.rui@nasa.gov [13]

Name: [Bill Teng](#) [14]
Organization(s): [Ag DSS](#) [15]
Email: william.l.teng@nasa.gov [16]

Name: [Richard Strub](#) [17]
Organization(s): [NASA GES DISC](#) [11], [ADNET Systems Inc](#) [12]
Email: richard.f.strub@nasa.gov [18]

Name: [Bruce Vollmer](#) [19]
Organization(s): [NASA GES DISC](#) [11]
Email: Bruce.E.Vollmer@nasa.gov [20]

Keywords: [hydrology](#) [21]

[time series](#) [22]

Source URL: <https://commons.esipfed.org/node/1174>

Links

- [1] <https://commons.esipfed.org/node/1174>
- [2] <https://commons.esipfed.org/taxonomy/term/464>
- [3] <http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings>
- [4] <http://mirador.gsfc.nasa.gov/>
- [5] <http://hydro1.sci.gsfc.nasa.gov/dods/>
- [6] <ftp://hydro1.sci.gsfc.nasa.gov/data/s4pa/>
- [7] <http://disc.sci.gsfc.nasa.gov/giovanni>
- [8] <http://his.cuahsi.org/>
- [9] <http://water.epa.gov/scitech/datait/models/basins/>
- [10] <https://commons.esipfed.org/node/1170>
- [11] <https://commons.esipfed.org/taxonomy/term/210>
- [12] <https://commons.esipfed.org/taxonomy/term/612>
- [13] <mailto:hualan.rui@nasa.gov>
- [14] <https://commons.esipfed.org/node/1171>
- [15] <https://commons.esipfed.org/taxonomy/term/658>
- [16] <mailto:william.l.teng@nasa.gov>
- [17] <https://commons.esipfed.org/node/1172>
- [18] <mailto:richard.f.strub@nasa.gov>
- [19] <https://commons.esipfed.org/node/1173>
- [20] <mailto:Bruce.E.Vollmer@nasa.gov>
- [21] <https://commons.esipfed.org/taxonomy/term/614>
- [22] <https://commons.esipfed.org/taxonomy/term/615>