Bridging the Digital Divide to Enhance Access to and Use of NASA Data for the Hydrological Community

William L. Teng[1,2], David R. Maidment[3], Bruce Vollmer[1], Christa D. Peters-Lidard[4], Hualan Rui[1,5], Richard Strub[1,5], Tim Whiteaker[3], David M. Mocko[4,6], and Dalia B. Kirschbaum[4]

[1]Goddard Earth Sciences Data and Information Services Center, Goddard Space Flight Center, NASA, Greenbelt, MD 20771, USA
[2]Wyle Information Systems, Inc., McLean, VA 22102, USA
[3]The University of Texas at Austin, Austin, TX 78758, USA
[4]Hydrological Sciences Laboratory, Goddard Space Flight Center, NASA, Greenbelt, MD 20771, USA
[5]ADNET Systems, Inc., Rockville, MD 20852, USA
[6]Science Applications International Corporation (SAIC)

Hydrology Data and Information Services Center (HDISC)
NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC)
Help Desk: gsfc-help-disc@lists.nasa.gov
http://disc.gsfc.nasa.gov/hydrology
Email: William.L.Teng@nasa.gov

Eleventh Earth Science Data Systems Working Group (ESDSWG) November 13 – 15, 2012

Introduction

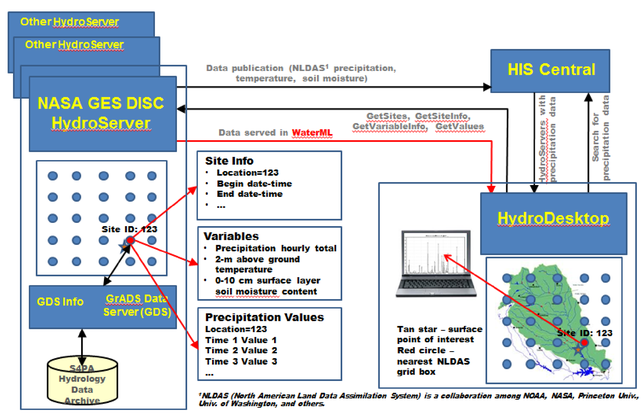
A longstanding and significant "Digital Divide" in data representation exists between hydrology and climatology and meteorology. Typically, in hydrology, earth surface features are expressed as discrete spatial objects such as watersheds, river reaches, and point observation sites; and time varying data are contained in time series associated with these spatial objects. Long time histories of data may be associated with a single point or feature in space. In meteorology and climatology, remotely sensed observations and weather and climate model information are expressed as continuous spatial fields, with data sequenced in time from one data file to the next. Hydrology tends to be narrow in space and deep in time, while meteorology and climatology are broad in space and narrow in time.

This Divide has been an obstacle, specifically, between the hydrologic community, as represented by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) and relevant data sets at the Goddard Earth Sciences Data and Information Services Center (GES DISC). CUAHSI has developed the Hydrologic Information System (HIS), which is built on international geospatial standards, with one of its aims to bridge the Divide. The opportunity costs of the Divide are high. It has largely prevented the routine access and use of NASA Earth sciences data by the hydrological and, more generally, geospatial community.

Two general solutions are possible: (1) retrieve multiple time series for short time periods and stitch the multiple time series into desired single long time series (long-term cost approach) and (2) reprocess (parameter and spatial subsetting) and archive data as one parameter one grid point per file (i.e., one time series per file or "data rod"). The latter is a one- time cost approach. The resultant time series files would be geospatially searchable and could be optimally accessed and retrieved by any user at any time.

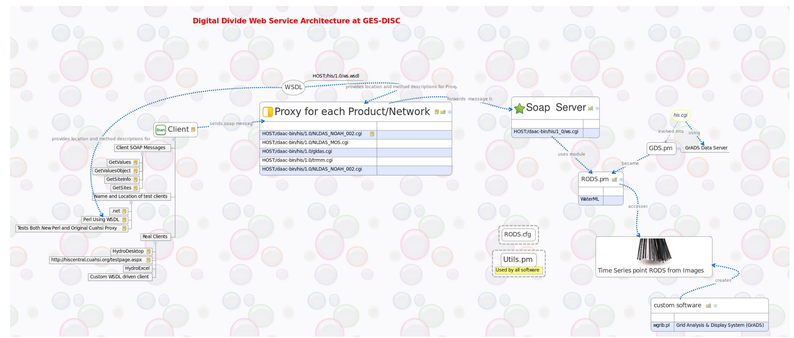
Prototypes

- EPA BASINS (Better Assessment Science Integrating point and Nonpoint Sources)
  - Primary BASINS requirement - On-the-fly retrieval of NLDAS precipitation time series of single grid points for the entire available temporal range in one request. To satisfy requirement, data must be geospatially searchable and amenable to efficient retrieval as time series.
  - GrADS Data Server (GDS) should be able to satisfy the requirement. However, due primarily to the way NLDAS (and generally, NASA) data are archived (one time step per file), time series retrieval for long time periods with adequate performance is hard, if not impossible, to achieve.
  - Solution for prototype: (1) Parameter subsetting; (2) spatial subsetting, but not to single grid points; (3) optimal archive of subset files for time series retrieval; and (4) setting GDS to return time series in ASCII for specified locations (grid point) and time range. After integration of Web service into BASINS tool, 32-year time series of hourly NLDAS data for a single grid point can be instantaneously retrieved and graphed by BASINS and used as inputs to subsequent analyses.
- CUAHSI Hydrologic Information System (HIS) HydroDesktop
  - For this prototype, we developed a Web service that serves NLDAS as time series and corresponding metadata in WaterML. A SOAP proxy service was used for registering the Web service in the CUAHSI-HIS catalog. Following is a schematic of GES DISC hydrologic data access from CUAHSI HIS client, HydroDesktop.

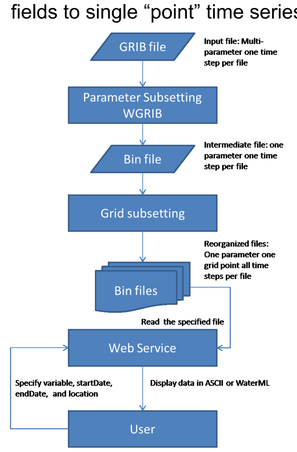


Bridge Over the Digital Divide

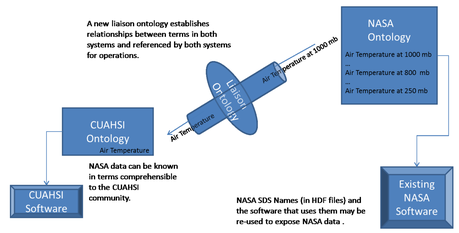
Digital Divide Web service at GES DISC



Converting from time-stepped continuous spatial fields to single "point" time series ("data rods")



Exposing NASA data to CUAHSI community



Data sets to be converted to data

Instr/Model/Param	Spatial Res	Spatial Cov	Temp Res	Temp Cov	Format
NLDAS[a,1]	1/8 °	N. America (53N-25N)	Hourly	1979-Present	GRIB
GLDAS[b,2]	¼ °	Global (90N-60S)	3-hourly	1948-Present	GRIB
TMPA[c] precipitation (TMI, SSM/I, AMSR-E, AMSU-B)[3]	¼ °	Global (50N-50S)	3-hourly, Daily	1998-Present	HDF4
LPRM[d] soil moisture (AMSR-E)[4]	¼ °	Global	Daily	2002-2011	NetCDF

[a]North American Land Data Assimilation System; [b]Global Land Data Assimilation System; [c]TRMM Multi-satellite Precipitation Analysis; [d]Land Parameter Retrieval Model. [1](Mitchell et al., 2004); [2](Rodell et al., 2004); [3](Huffman et al., 2007); [4](Owe et al., 2008; De Jeu et al., 2008).

Selected parameters for conversion to data rods: Precipitation, soil moisture, soil temperature, evapotranspiration, surface runoff, surface air temperature, potential evaporation, and surface wind speed

Total number of files and volume of yearly data rods

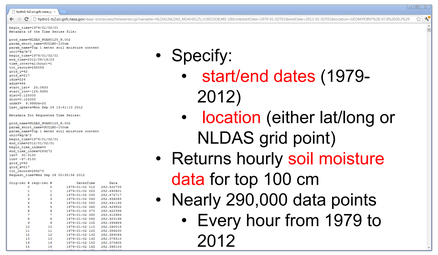
Total # files and volume of yearly data rods

Noah LSM	Lon-dim	Lat-dim	Tot # grids	Original # files/day	Days/yr	Years	Original Tot # files	Land frac	Data rods # files/param/yr	# param	Data rods # files/yr	Data rods Tot # files	Data rods File size	Data rods Tot vol	Original File size	Original Tot vol
NLDAS	464	224	103936	24	365	32	280320	0.9	93542	10	935424	29933568	35040	1.04887E+12	7000000	1.96224E+12
GLDAS	1440	600	864000	8	365	12	35040	0.3	259200	10	2592000	31104000	11680	3.63295E+11	15200000	5.32608E+11
Total							315360					61037568		1.41217E+12		2.49485E+12

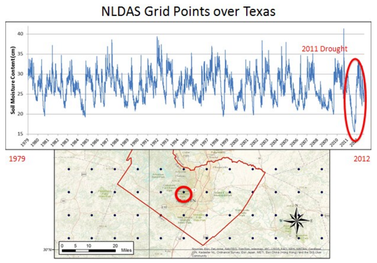
Application

Texas Natural Resources Information System (TNRIS) Drought monitoring

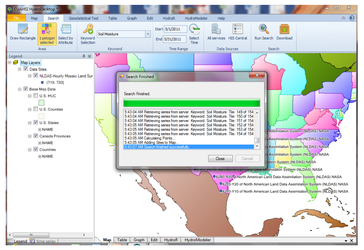
- Early data rods service of NLDAS Noah 0-100 cm soil moisture, e.g., http://hydro1-ts1.sci.gsfc.nasa.gov/daac-bin/access/timeseries.cgi?variable=NLDAS:NLDAS_FORA0125_H.002:APCPsfc&startDate=1979-01-02T01&endDate=2012-11-01T12&location=NLDAS:X304-Y071&type=plot

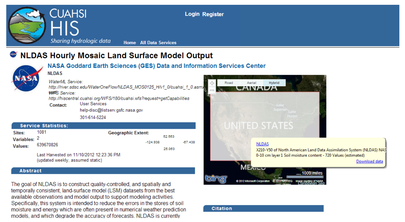


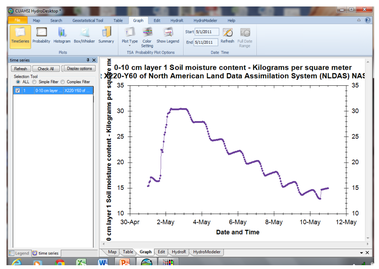
- Specify:
  - start/end dates (1979-2012)
  - location (either lat/long or NLDAS grid point)
- Returns hourly soil moisture data for top 100 cm
- Nearly 290,000 data points
  - Every hour from 1979 to 2012

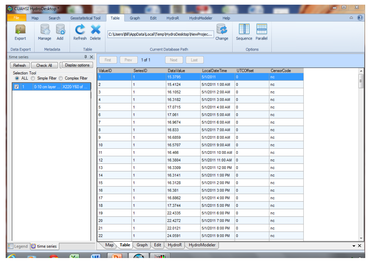


- Registering NLDAS service with HIS Central (catalog) to enable access to data from HIS client, HydroDesktop.









Other Planned Activities

- Enable linking of HIS catalog and Earth Observing System (EOS) Clearinghouse (ECHO) so that any OPeNDAP-enabled data set registered at ECHO could be searchable and accessible by HIS users.
- Planned enhancements to HydroDesktop
  - Modify GetSites to better handle NASA regularly gridded data (with many more "points")
  - Enable accessing multiple time series (e.g., within a polygon such as watershed) and aggregating them into a single time series.

Acknowledgment: This work is supported by NASA ROSES NNH11ZDA001N-ACCESS.