

Bridging the Gap between NASA Hydrological Data and the Geospatial Community

Hualan Rui^{1,3*}, Bill Teng^{1,4}, Bruce Vollmer¹, David M. Mocko^{2,6}, Hiroko K. Beaudoin^{2,5}, Joseph Nigro^{2,7}, Mark Gray⁸, David Maidment⁹, Richard Hooper¹⁰

¹Goddard Earth Sciences Data and Information Services Center, Goddard Space Flight Center, NASA, Greenbelt, MD 20771, United States

²Hydrological Sciences Branch, Goddard Space Flight Center, NASA, Greenbelt, MD 20771, United States

³ADNET Systems, Inc., 164 Rollins Avenue, Rockville, MD 20852, United States

⁴Wyle Information Systems, Inc., 1651 Old Meadow Road, McLean, VA 22102, United States

⁵Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, United States

⁶SAIC, 4600 Powder Mill Road, Suite 400, Beltsville, MD, 20705, United States

⁷Science System and Applications, Inc., 10210 Greenbelt Road, Suite 600, Lanham, MD 20706

⁸AQUA TERRA Consultants, 150 E. Ponce de Leon Ave., Suite 355, Decatur, GA 30030 United States

⁹University of Texas, Austin, TX, 78712, United States

¹⁰CUAHSI, 196 Boston Avenue, Suite 2100, Medford, MA 02155, United States

Hualan.Rui@nasa.gov; William.L.Teng@nasa.gov; Bruce.E.Vollmer@nasa.gov; David.Mocko@nasa.gov; Hiroko.Kato@nasa.gov; Joseph.D.Nigro@nasa.gov; MarkGray@aquaterra.com; maidment@mail.utexas.edu; rhooper@cuahsi.org

ABSTRACT

There is a vast and ever increasing amount of data on the Earth interconnected energy and hydrological systems, available from NASA remote sensing and modeling systems, and yet, one challenge persists: increasing the usefulness of these data for, and thus their use by, the geospatial communities. The Hydrology Data and Information Services Center (HDISC), part of the Goddard Earth Sciences DISC, has continually worked to better understand the hydrological data needs of the geospatial end users, to thus better able to bridge the gap between NASA data and the geospatial communities. This paper will cover some of the hydrological data sets available from HDISC, and the various tools and services developed for data searching, data subsetting, format conversion, online visualization and analysis, interoperable access, etc., to facilitate the integration of NASA hydrological data by end users. The NASA Goddard data analysis and visualization system, Giovanni, is described. Two case examples of user-customized data services are given, involving the EPA BASINS (Better Assessment Science Integrating point & Non-point Sources) project and the CUAHSI Hydrologic Information System, with the common requirement of on-the-fly retrieval of long duration time series for a geographical point.

KEYWORDS: Hydrological data, NASA, BASINS, CUAHSI, GLDAS/NLDAS, remote sensing, geospatial

INTRODUCTION TO HYDROLOGICAL DATA AT NASA HDISC

Terrestrial hydrological variables are important in global hydrology, climate, and carbon cycle studies. To generate global fields of these variables, however, is still a challenge. The goal of a land data assimilation system (LDAS) is to ingest satellite- and ground-based observational data, using advanced land surface modeling and data assimilation techniques, in order to generate optimal fields of land surface states and fluxes and, thereby, facilitate hydrology and climate modeling, research, and forecast. NLDAS (North American LDAS) (Mitchell et al., 2004) and GLDAS (Global LDAS) (Rodell et al., 2004) data are produced by specific instances of the Land Information System (LIS) software framework for high-performance land-surface modeling and data assimilation. LIS is developed by the Hydrological Sciences Branch at NASA Goddard Space Flight Center. NLDAS is a collaboration project between several groups (NOAA/NCEP/ EMC, NASA/GSFC, Princeton University, University of

* Corresponding Author Address: Hualan Rui, Goddard Earth Sciences Data and Information Services Center, Goddard Space Flight Center, NASA, Greenbelt, MD 20771, United States

Washington, NOAA/OHD, and NOAA/NCEP/CPC) and is a core project of NOAA/CPA. GLDAS is supported by the NASA Energy and Water cycle Study (NEWS). To date, both NLDAS and GLDAS have produced more than 30 years (1979 to present) of quality-controlled, spatially and temporally consistent, land-surface model data (Table 1).

Both NLDAS and GLDAS data sets have recently been improved. With the motivation of creating more climatologically consistent data sets, GLDAS-2 data have been generated by using upgraded versions of Land Surface Models (LSMs). Compared with GLDAS-1, **GLDAS-2** data have been enhanced by using the global meteorological forcing data set from Princeton University (Sheffield et al., 2006); initialization of soil moisture over desert; updated model specific parameter files; and advanced snow assimilation scheme (in NOAA0.25). By using the Princeton forcing data set, the GLDAS-2 temporal coverage is extended back to 1948. In GLDAS-1, forcing sources switched several times throughout the record from 1979 to present, which introduced unnatural trends and exhibited highly uncertain forcing fields in 1995-1997. GLDAS-2 has two streams of simulations, Princeton-based from 1948 to present and observation-based from 2001 to present. The GLDAS-2 NOAA model provides 3-hourly and monthly data at 1 and 0.25 degree resolutions and covers a period of more than 60 years (from 1948 to present). Other GLDAS-2 models (CLM, Catchment, and VIC) provide 3-hourly and monthly data at 1 degree resolution from 2001 onwards. With the extended temporal coverage, the GLDAS-2 data are expected to play an even more important role in global hydrology and climate studies. Table 2 shows the model version of GLDAS-1 and GLDAS-2 data. The 30-year hourly 0.125 degree resolution **NLDAS Phase 2 (NLDAS-2)** data were released at the end of 2009 and continue to be produced with a typically 2-4 day lag from the present. Additionally, the NLDAS Phase 1 data (1996 – 2007) were recently added to the GES DISC archives and released to the public, to continue to serve users of NLDAS-1 data and allow easier comparisons between the two phases of NLDAS.

Table 1. Basic characteristics of the NLDAS and GLDAS products.

	NLDAS	GLDAS
Content	Water and energy budget data, forcing data	
Spatial coverage	Conterminous US, parts of southern Canada and northern Mexico	All land north of 60° South
Spatial resolution	0.125°	0.25° and 1.0°
Temporal coverage	Phase-1: Aug. 1, 1996 - Dec. 31, 2007 Phase-2: Jan. 1, 1979 – present	Version-1 1.0°: Jan. 1, 1979 - present 0.25°: Feb. 24, 2000 - present Version-2: Jan. 1, 1948 - present
Temporal resolution	Hourly and monthly	3-hourly and monthly
Forcing	Multiple data sets derived from satellite measurements, radar estimation, precipitation gauges, and atmospheric analyses	Multiple data sets derived from satellite measurements and atmospheric analyses
Land surface models	Mosaic, Noah, SAC, VIC	CLM, Mosaic, Noah, VIC
Output format	GRIdded Binary (GRIB)	
Elevation definition	GTOPO 30	
Vegetation definition	University of Maryland, 1 km	

Table 2. LSM model versions for GLDAS-1 and GLDAS-2.

Model	Resolution	GLDAS-1	GLDAS-2	Remarks
NOAH	1.0°	Version 2.7	Version 2.7.1	Updated model parameters that specify the initial soil temperature
CLM	1.0°	Version 2.0	Version 3.5	Used MODIS based parameter data sets, stand alone

VIC	1.0°	Version 4.4 Water balance mode	Energy balance mode	Includes all variables
Catchment	1.0°	Mosaic model	Catchment	Model switch
NOAH	0.25°	Version 2.7, Snow DA (data assimilation): direct insertion	Version 2.7.1, Snow DA: forward-looking	Updated model parameters that specify the initial soil temperature

All NLDAS and GLDAS data are accessible from the Hydrology Data and Information Services Center (HDISC) at the NASA GES DISC (<http://disc.sci.gsfc.nasa.gov/hydrology>). To facilitate access to these data by various user communities, GES DISC has implemented several convenient data access methods (Table 3):

1. **Mirador** searching and downloading (Lynnes et al., 2009) - Includes keyword searching, hierarchical navigation based on projects and on Science Areas. Provides spatial and parameter subsetting and data format conversion.
2. **GrADS Data Server (GDS)** access - Provides parameter and spatial subsetting. Outputs data in binary, ASCII, or image. Performs any operation that can be expressed in a single GrADS expression.
3. **Anonymous ftp** downloading - Navigation based on data product, year, and Julian date, which provides simple and fast direct data downloading.

Table 3. GLDAS and NLDAS data access at <http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings>.

Data Type (Short Name)	Description	FTP	GDS	Mirador	
				Navigation	Search
NLDAS-1, 0.125 degree, North America					
NLDAS_FOA0125_H.001	Hourly forcing	✓	✓	✓	✓
NLDAS-2, 0.125 degree, North America					
NLDAS_FORA0125_H.002	Hourly primary forcing	✓	✓	✓	✓
NLDAS_FORB0125_H.002	Hourly secondary forcing	✓	✓	✓	✓
NLDAS_MOS0125_H.002	Hourly Mosaic	✓	✓	✓	✓
GLDAS-2, 1.0 degree, Global					
GLDAS_NOAH10_3H_E1.002	3 hourly Noah experiment 1	✓	✓	✓	✓
GLDAS_NOAH10_M_E1.002	Monthly Noah experiment 1	✓	✓	✓	✓
GLDAS-1, 0.25 degree, Global					
GLDAS_NOAH025SUBP_3H	3 hourly Noah	✓	✓	✓	✓
GLDAS_NOAH025_M	Monthly Noah	✓	✓	✓	✓
GLDAS-1, 1.0 degree, Global					
GLDAS_CLM10SUBP_3H	3 hourly CLM	✓	✓	✓	✓
GLDAS_CLM10_M	Monthly CLM	✓	✓	✓	✓
GLDAS_MOS10SUBP_3H	3 hourly Mosaic	✓	✓	✓	✓
GLDAS_MOS10_M	Monthly Mosaic	✓	✓	✓	✓
GLDAS_NOAH10SUBP_3H	3 hourly Noah	✓	✓	✓	✓
GLDAS_NOAH10_M	Monthly Noah	✓	✓	✓	✓
GLDAS_VIC10_3H	3 hourly VIC	✓	✓	✓	✓
GLDAS_VIC10_M	Monthly VIC	✓	✓	✓	✓

BRIDGING THE GAP (“DIGITAL DIVIDE”)

The NLDAS and GLDAS data sets are part of a vast and ever increasing amount of data on the Earth’s interconnected energy and hydrological systems available from NASA remote sensing and modeling systems. Yet, one challenge persists: increasing the usefulness of these data for, and thus their use by, the geospatial communities.

NLDAS and GLDAS were created and archived as individual model products, one time step per file in GRIB format. GRIB (GRIdded Binary) is a mathematically concise data format commonly used in meteorology to store historical and forecast weather data. It is standardized by the World Meteorological Organization. Although there are many well-developed software packages and tools for handling GRIB files, some users from communities other than meteorology still have difficulty in handling data in GRIB format. Furthermore, hydrologic information science commonly requires data to be at specific geo-locations, often as time series. In order to retrieve a single point complete time series for one parameter, e.g., a user has to go through the entire data archive, often of volumes in the Terabytes. This is the "Digital Divide," described by Maidment et al. (2010), that exists between the world of discrete spatial objects in geographical information systems (GIS) and associated time series and the world of continuous space-time arrays as used in weather and climate science.

The HDISC has continually worked to better understand the hydrological data needs of the geospatial end users, to thus better able to bridge the gap between NASA data and the geospatial communities. Here, we provide three examples of this effort.

1. Online Visualization and Analysis in Giovanni

Giovanni is a NASA Goddard data analysis and visualization system that provides a simple and intuitive way to visualize, analyze, and access vast amounts of Earth science remote sensing data, without having to download the data (Acker and Leptoukh, 2007; Berrick et al., 2009). It is an online application that allows researchers to rapidly explore data, so that spatial-temporal variability, anomalous conditions, and patterns of interest can be directly analyzed online before optional downloading of data (<http://disc.sci.gsfc.nasa.gov/techlab/giovanni/>). Giovanni comprises a number of interfaces, called instances, tailored to meet the needs of different Earth science research communities. Giovanni has contributed to many users’ science research efforts and applications (<http://disc.sci.gsfc.nasa.gov/giovanni/additional/publications>).

To facilitate the hydrological user community’s use of GLDAS data, the GES DISC has created and made available a customized monthly GLDAS Giovanni instance (Figs. 1 and 2). With the latter, users can simply select a specific region or location point, one or more parameters, spatial and temporal ranges, and the visualization function (e.g., latitude-longitude map, time series), and get the result in image, ASCII, netCDF, or HDF format. ASCII outputs can be directly input to GIS, Excel, and other software packages for further studies, together (possibly) with data from other sources. The GES DISC is currently working to incorporate GLDAS 3-hourly data and NLDAS hourly data into Giovanni.

Other hydrological data from NASA are also available via Giovanni, such as the TRMM Online Visualization and Analysis Systems (TOVAS) for TRMM and other precipitation data, Giovanni for Collaborative Energy and Water Cycle Information Services (CEWIS), and Giovanni for International Precipitation Working Group (IPWG) (Liu et al., 2007; 2009).

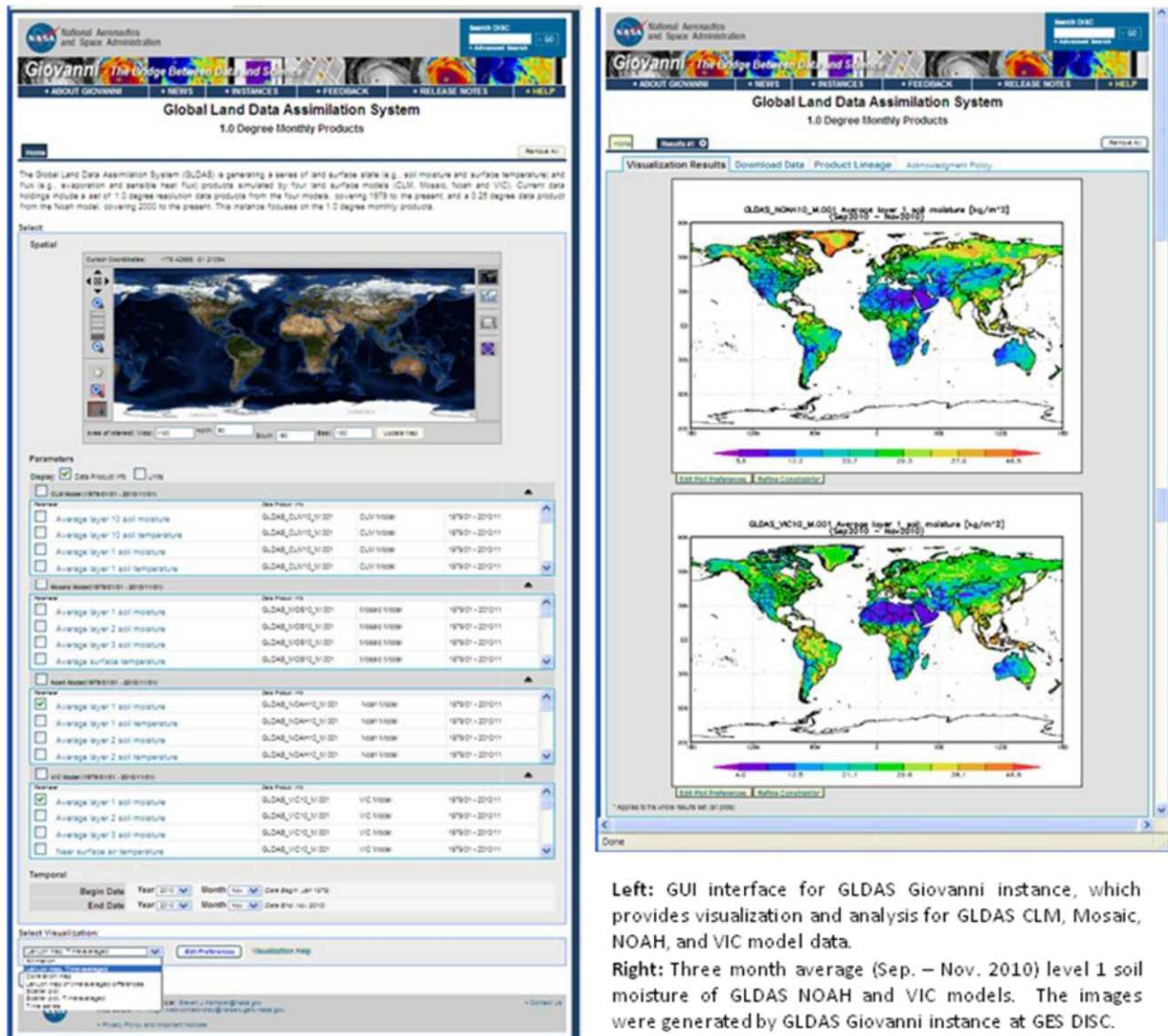


Figure 1. GLDAS Giovanni GUI interface and sample results.

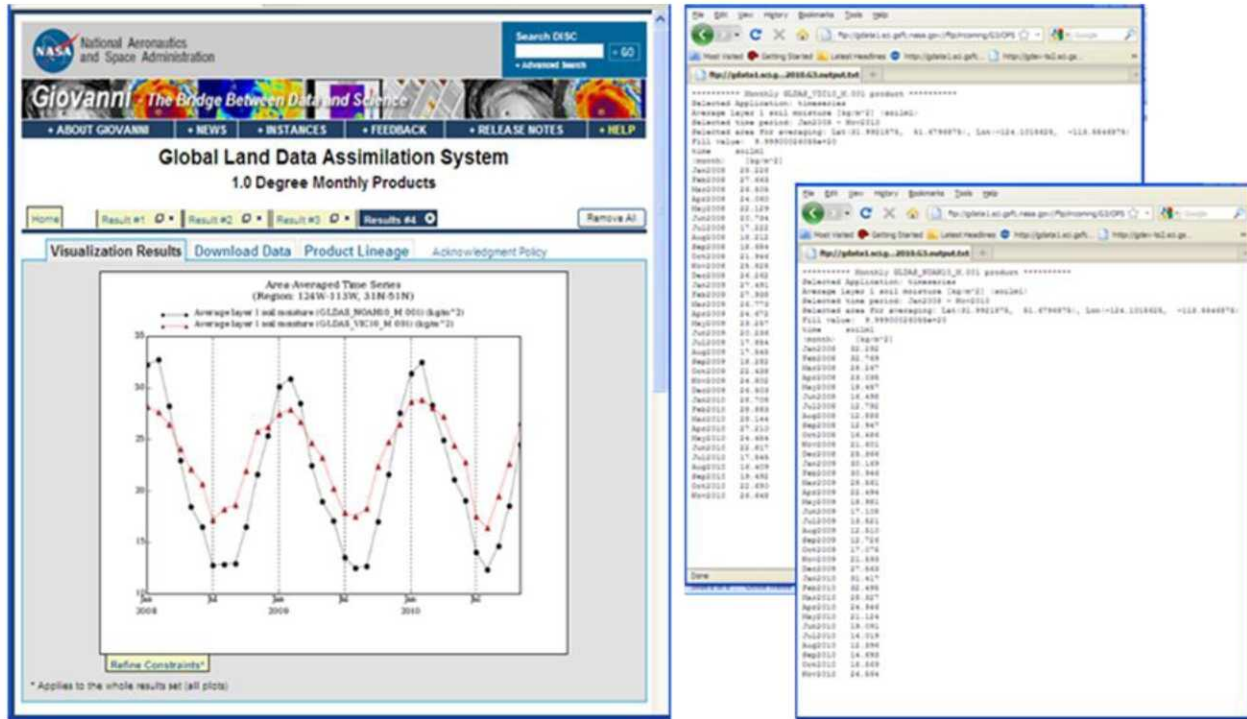


Figure 2. Average level-1 soil moisture intercomparison between GLDAS NOAH and VIC models. The time series plots were generated by GLDAS Giovanni instance. The ASCII outputs were also generated by the instance and available as one of the “Download Data” options.

2. Integrate NASA NLDAS precipitation data into BASINS

The Better Assessment Science Integrating Point & Nonpoint Sources (BASINS), <http://water.epa.gov/scitech/datat/models/basins/index.cfm>, created by the U.S. Environmental Protection Agency (EPA), is a multi-purpose environmental analysis system that integrates a GIS, national watershed data, and state-of-the-art environmental assessment and modeling tools into one convenient package. The latest BASINS 4.0 makes the tool universally available to anyone interested in facilitating examination of environmental information, supporting analysis of environmental systems, and providing a framework for examining management alternatives for watersheds. This system makes it possible to quickly assess large amounts of point source and non-point source data in a format that is easy to use and understand. BASINS supports the development of cost-effective approaches to watershed management and environmental protection, including the monitoring of water quality using Total Maximum Daily Loads (TMDLs).

In collaboration with the BASINS project, NASA HDISC has made the NLDAS precipitation data available via the BASINS download tool, thus making NLDAS precipitation data spatially searchable, via the BASINS interface, and allowing BASINS users access to single point 30-year time series of hourly NLDAS precipitation data in a single request (Fig. 3). To enable these capabilities and ensure operational performance, the NLDAS GRIB data were reprocessed (reorganized) by parameter and spatial subsetting, archived in a way optimized for time series retrieval, and incorporated into the GrADS Data Server (GDS).

Because NLDAS data provide complete coverage for the entire North American region, with high temporal and spatial resolutions (hourly and $0.125^\circ \times 0.125^\circ$), the data are in high demand by the hydrological community. For example, a recent EPA project on storm water management will need 30-year time series of NLDAS precipitation, evapotranspiration (ET), temperature, wind speed, etc. With the integration into BASINS, the NLDAS precipitation data can now be easily retrieved and analyzed with other hydrological data in BASINS. As resources permit, HDISC will continue to work with BASINS to incorporate more NLDAS parameters and other hydrological data.

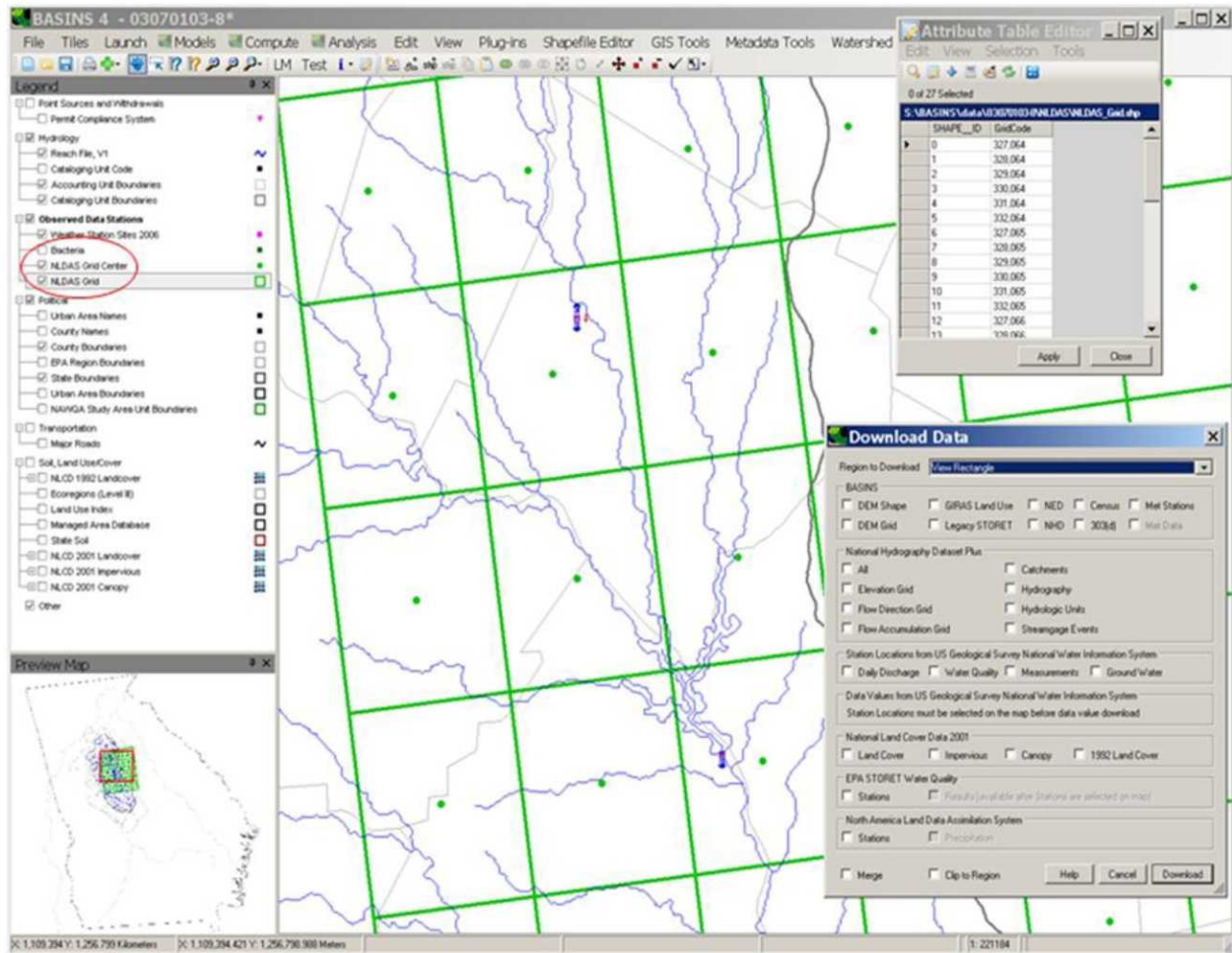


Figure 3. Screen capture of the BASINS interface, showing availability of NLDAS data.

3. Integrate NASA hydrological data into CUAHSI HIS

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS), <http://his.cuahsi.org/>, is an internet-based system for sharing hydrologic data. It comprises databases and servers, connected through web services to client applications, allowing for the publication, discovery, and access of data.

In collaboration with the Hydrologic Sciences Branch at NASA's Goddard Space Flight Center and CUAHSI HIS, NASA HDISC has been working to integrate NLDAS, GLDAS, and other NASA hydrological data into CUAHSI HIS. Work has focused on developing a Web service that serves NASA hydrological data as time series and corresponding metadata in WaterML (an XML Language for Communicating Water Observations Data, <http://adsabs.harvard.edu/abs/2007AGUFM.H11K.04M>). The Web service provides four basic functions, GetSites, GetSiteInfo, GetVariableInfo, GetValues. Figure 4 shows a schematic of NASA hydrological data access from CUAHSI HIS' client, HydroDesktop.

NASA Hydrologic Data Access from HydroDesktop (an example)¹

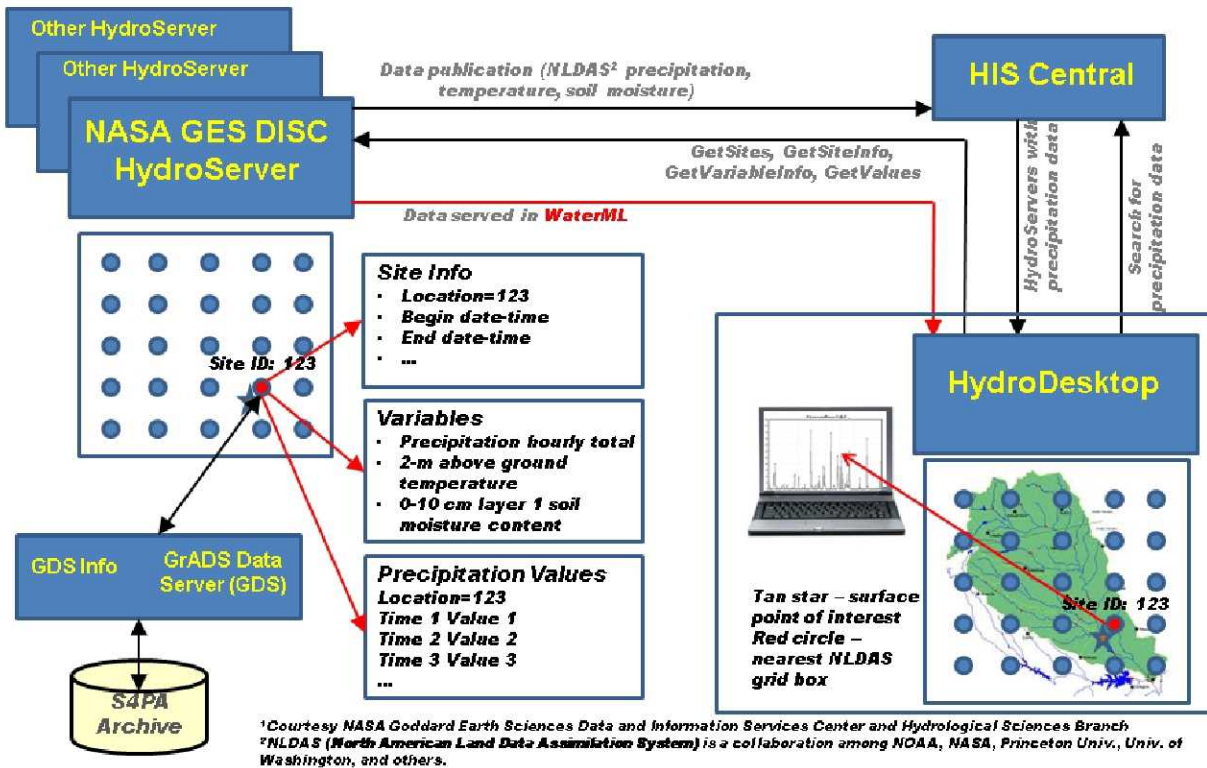


Figure 4. Schematic of NASA hydrological data access from CUAHSI HIS' client, HydroDesktop.

With the HDISC Web service registered in CUAHSI HIS, the NLDAS, GLDAS, and other NASA hydrological data could become searchable, retrievable, and analyzable, along with hydrologic data from other data sources available via HIS. This enhanced data access will facilitate, for the broad CUAHSI HIS user community, the use of NASA hydrological data.

CONCLUSIONS

The gap, or “Digital Divide,” between NASA hydrological data and the geospatial community is a longstanding one and still to be bridged. The key to bridging this gap is a better understanding of the hydrological data needs of the geospatial end users, which is a central focus of the NASA HDISC. The availability of the GLDAS (and, in the future, NLDAS) data set via NASA Goddard Giovanni is one approach to facilitate the integration of NASA data by end users. The two ongoing, collaborative efforts between HDISC and EPA BASINS and CUAHSI HIS have already demonstrated the potential of user-customized Web services for enhanced access and use of NASA data, in these two case, in the form of on-the-fly retrieval of long duration time series for a geographical point.

REFERENCES

Acker, J.G. and G. Leptoukh, 2007. Online analysis enhances use of NASA earth science data, Eos, Trans., Amer. Geophysical Union, 88(2), 14, 17.

- Berrick, S.W., G. Leptoukh, J.D. Farley, and H. Rui, 2009. Giovanni: A Web services workflow-based data visualization and analysis system, *IEEE Trans. Geosci. Remote Sensing*, 47(1), 106-113, doi: 10.1109/TGRS.2008.2003183.
- Liu, Z., H. Rui, W.L. Teng, L. Chiu, G. Leptoukh, and S. Kempler, 2009. Developing an online information system prototype for global satellite precipitation algorithm validation and intercomparison. *J. Appl. Meteor. Climatol.*, 48, 2581–2589, doi: 10.1175/2009JAMC2244.1.
- Liu, Z., H. Rui, W.L. Teng, L.S. Chiu, G. Leptoukh, and G. Vicente, 2007. Online visualization and analysis: A new avenue to use satellite data for weather, climate, and interdisciplinary research and applications, *In Measuring Precipitation from Space—EURAINSAT and the Future*, V. Levizzani, P. Bauer, and F. Joseph Turk, Eds., *Advances in Global Change Research Series*, Springer, 28, 549–558.
- Lynnes, C., R. Strub, E. Seiler, T. Joshi, and P. MacHarrie, 2009. Mirador: A simple fast search interface for global remote sensing data sets, *IEEE Trans. Geosci. Remote Sensing*, 47, 92-96.
- Maidment, D.R., F. Salas, B. Domenico, and S. Nativi, 2010. Crossing the Digital Divide: Connecting GIS, time series and space-time arrays, Abstract IN13A-1095 presented at 2010 Fall Meeting, AGU, San Francisco, CA.
- Mitchell, K.E., D. Lohmann, P.R. Houser, E.F. Wood, J.C. Schaake, A. Robock, B.A. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R.W. Higgins, R.T. Pinker, J.D. Tarpley, D.P. Lettenmaier, C.H. Marshall, J.K. Entin, M. Pan, W. Shi, V. Koren, J. Meng, B. H. Ramsay, and A.A. Bailey, 2004. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, 109, D07S90, doi: 10.1029/2003JD003823.
- Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, 2004. The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, 85(3), 381-394.