Giovanni - the Bridge between Data and Science

Zhong Liu[a,b] and James Acker[a,c]

[a]NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC)

[b]Center for Spatial Information Science and Systems (CSISS), George Mason University

[c]Adnet Systems, Inc.

Capsule: Using satellite remote sensing data sets can be a daunting task. A Web-based tool, Giovanni, has been developed to facilitate access, visualization, and exploration of NASA Earth science data sets

Abstract

This article describes new features in the Geospatial Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni), a user-friendly online tool that enables visualization, analysis, and assessment of NASA Earth science data sets without downloading data and software. Since the satellite era began, data collected from Earth-observing satellites have been widely used in research and applications; however, using satellite-based data sets can still be a challenge to many. To facilitate data access and evaluation, as well as scientific exploration and discovery, the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) has developed Giovanni for a wide range of users around the world. This article describes the latest capabilities of Giovanni with examples, and discusses future plans for this innovative system.

Over Earth's vast oceans and remote continents, traditional large-scale observations of the atmosphere, ocean, and land surface can be difficult and costly to both deploy and maintain, and are therefore impractical to provide adequate long-term observational data for research and applications. This is particularly true as global observations become increasingly important for understanding global change processes. Satellite instruments can overcome surface observation limitations by synoptically acquiring Earth science data. NASA's Earth Observing System (EOS) project [NASA 2016] is a global observation campaign, consisting of a coordinated series of polar-orbiting, low-inclination satellites intended for long-term global observations, potentially enabling improved understanding of Earth's geophysical systems.

[Break-up text: Giovanni simplifies satellite data access. Downloading data and software is not required.]

Accessing and using NASA data is a challenge to many researchers, due to issues such as heterogeneous data formats, complex data structures, large volume data storage, special programming requirements, diverse analytical software options, etc. that often require a significant investment in time and resources, especially for novices. By facilitating data access and evaluation, as well as promoting open access to create a level playing field for non-funded scientists, NASA data can be more readily used for scientific discovery and societal benefits. To advance this goal, the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) developed the Geospatial Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni) [GES DISC 2016]. Giovanni has assisted researchers around the world publish over 1300 peer-reviewed papers [Acker 2016] in a wide range of Earth science disciplines and other areas.

A Brief History

Giovanni [Liu et al. 2007, Acker and Leptoukh 2007, Berrick et al. 2009] was initiated and developed for faster and easier access to, and evaluation of, data sets at the GES DISC after collecting and analyzing feedback from user support. The first Giovanni instance was the Tropical Rainfall Measuring Mission (TRMM) Online Visualization and Analysis System (TOVAS). TOVAS served TRMM Multi-satellite Precipitation Analysis (TMPA) and other global precipitation data sets. As TOVAS gained popularity, inclusion of more satellite data sets in Giovanni was requested. To address this demand, multiple discipline- or mission-based data portals were created. The current Giovanni has evolved further, featuring a new unified Web interface to support interdisciplinary Earth system research, allowing synergistic use of data sets from different satellite missions.

[Break-up text: Giovanni has helped publish over 1300 peer-reviewed papers]

Giovanni Data Set Summary

Giovanni provides access to numerous satellite data sets archived and distributed by the GES DISC, concentrated primarily in the areas of atmospheric composition, atmospheric dynamics, global precipitation, hydrology, and solar irradiance. The data are from NASA heritage and EOS missions including Terra, Aqua, Aura, Solar Radiation and Climate Experiment (SORCE), TRMM, Global Precipitation Measurement (GPM), UARS (Upper Atmosphere Research Satellite), Earth Probe Total Ozone Mapping Spectrometers (TOMS), AIRS (Atmospheric Infrared Sounder), MISR (Multi-angle Imaging SpectroRadiometer), SeaWiFS (Sea-viewing Wide Field-of-view Sensor), and data assimilation collections (including

Modern-Era Retrospective Analysis for Research and Applications (MERRA), Global Land Data Assimilation System (GLDAS), North American Land Data Assimilation System (NLDAS)). The Federated Giovanni partnership provides access to data sets archived at other NASA data centers, thereby broadening data access opportunities.

[Break-up text: Giovanni offers over 1300 variables covering a wide range of Earth science disciplines]

Over 1300 variables are currently available in Giovanni. The Web interface has both Keyword and Faceted Search capabilities (Fig. 1) for locating variables of interest. For example, a search for 'precipitation' returns over 100 related variables (Fig. 1). By using facets, one can filter for variables based on satellite missions (TRMM, GPM) (Fig. 1), instruments, spatial or temporal resolution, etc.

The operating lifetimes of low Earth orbiting (LEO) satellites are often quite limited (~5 years), far less than the thirty years defined by WMO (the World Meteorological Organization) for developing climatology data sets. Some users, however, may still wish to conduct preliminary studies with these satellite data sets to obtain information on spatial distribution and inter-seasonal variation. Giovanni provides the capability to derive climatological maps and time series based on user-defined time periods.

Analytical Features

[Break-up text: Commonly used analytical methods are available in Giovanni]

Many commonly used analytical and plotting capabilities, for capturing spatial and temporal characteristics of data sets, are available in Giovanni. Mapping options include time-averaging, animation, accumulation (precipitation), time-averaged overlay of two data sets, and user-defined climatology. For time series, options include area-averaged, differences, seasonal, and Hovmöller diagrams. Cross-sections, applicable to 3-D data sets from AIRS and MERRA, include latitude-pressure, longitude-pressure, time-pressure, and vertical profile. For data comparison, Giovanni has built-in processing code for data sets that require measurement unit conversion and regridding. Commonly used comparison functions include map and time-series differences, as well as correlation maps and X-Y scatter plots (area-averaged or time-averaged). Zonal means and histogram distributions can also be plotted. Samples of the analytical and plotting features are shown in Fig. 2.

Visualization Features

Visualization features include interactive map area adjustment; animation; interactive scatter plots; data range adjustment; choice of color palette; contouring; and scaling (linear or log). The on-the-fly area adjustment feature allows an interactive and detailed examination of a result map without re-plotting data. Animations are helpful to track evolution of an event or seasonal changes. Interactive scatter plots allow identification and geolocation of a point of interest in a scatter plot. Adjustments of any of these plots provide customized options to users.

Other Features

To support increasing socioeconomic and GIS activities in Earth sciences, shapefiles have been added for countries, states in the United States, and major watersheds around the world. Available functions for shapefiles are time-averaged and accumulated maps (Fig. 2b), area-averaged time series, and histogram. Land-sea masks have been recently added.

[Break-up text: Data in Giovanni can be downloaded for further analysis]

All data files involved in Giovanni processing are listed and available in the lineage page. Available image formats are PNG, GEOTIFF, and KMZ (Keyhole Markup Language) that can be used for different applications and software packages; for example, KMZ files are conveniently imported into Google Earth (Fig. 2b) where a rich collection of overlays are available. All input and output data are available in NetCDF, which can be handled by many off-the-shelf software packages. Furthermore, users can bookmark URLs generated by Giovanni processing for reference, documentation, or sharing with other colleagues.

Conclusion and Future Plans

With the latest features and examples described in this overview article, we have shown how Giovanni simplifies access, evaluation, and exploration of NASA satellite data sets. More examples are available in the Giovanni publication list [Acker 2016]. Despite these achievements, improvements in Giovanni are both in development and under consideration, to accommodate increasing demand for additional analytical and plotting capabilities, more data sets, and advanced information technologies to make data exploration simple and productive.

Potential enhancements to the system, currently being evaluated, include visualization and analysis of satellite orbital data sets (Level 2 data), acquisition of complementary data sets from other NASA data centers, and analysis methods specifically for multi-satellite and multi-sensor measurements. Currently, data sets in Giovanni consist of variables mapped on uniform space-time grid scales, so non-gridded or satellite orbital data sets (commonly providing higher spatial and temporal resolution) are not currently available in the system. Adding orbital data sets to Giovanni could contribute to research requiring increased data resolution and coverage. Data sets from other data centers, including current and future satellite missions, would further enhance Giovanni’s capabilities for analysis of the integrated Earth system. Note that barriers still exist in the optimization of Giovanni for interdisciplinary studies and data intercomparison. For example, terminologies in data sets can be significantly different between Earth science communities, requiring coordinated efforts to reach consensus and develop standards for uniform data products.

The NASA-wide User Registration System (URS) will be another useful feature in Giovanni. It is expected to enhance the Giovanni user experience; for example, with the URS, users can set frequently-used preferences in their profiles, record and retrieve their personal history of data set exploration, and establish their own data collections. For data product developers, their test data can be uploaded and compared with observations and other well-established data sets in Giovanni to identify issues in their products, a useful capability to improve data quality. Giovanni developers will also be able better understand their users through profiles and other statistics collected from the URS, in order to develop more user-friendly services.

Acknowledgements:

We would like to recognize the team effort of all past and current members at the GES DISC for their contributions to the development of Giovanni. Thanks extend to data set algorithm developers and many users for their feedback and suggestions. The GES DISC is funded by NASA's Science Mission Directorate (SMD)

References:

Acker, J. G., and G. Leptoukh, 2007: Online analysis enhances use of NASA earth science data. *Eos. Trans. Amer. Geophys. Union*, ***88*** *(2)*, 14–17. DOI: 10.1029/2007EO020003

Acker, J.G., 2016: GIOVANNI Publications by Year, available online: http://disc.sci.gsfc.nasa.gov/giovanni/additional/publications

Berrick, S. W., G. Leptoukh, J. D. Farley, and H. Rui, 2009: Giovanni: A Web service workflow-based data visualization and analysis system. *IEEE Trans. Geosci. Remote Sens.*, **47** (1), 106–113. DOI: 10.1109/TGRS.2008.2003183

GES DISC, 2016, Giovanni, available online: http://giovanni.gsfc.nasa.gov/

Liu, Z., H. Rui, W. Teng, L. 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. *Measuring Precipitation from Space - EURAINSAT and the future, Advances in Global Change Research, 28,* 549-558. DOI: 10.1007/978-1-4020-5835-6

NASA, 2016, NASA’s Earth Observing System (EOS), available online: http://eospso.nasa.gov/

Figure 1. The Giovanni web interface. Over 1300 variables are available for visualization and analysis. Commonly used analytical methods and visualization are available. Keyword and faceted search capabilities allow locating variables of interest with ease. Shapefiles are available for supporting GIS data exploration. Both input and output data can be downloaded for further analysis.

Figure 2. Examples of maps and plots generated from Giovanni: a) July 2016, the hottest month ever - MODIS day surface temperatures (in Kelvin); b) Accumulated rainfall (in mm) from IMERG-Late Run, imported in Google Earth as KMZ, showing a record breaking flood event in Louisiana in August 2016; c) The TMPA precipitation climatology (1998-2015, in mm/day) for the boreal summer (June, July, and August); d) El Niño reduces the phytoplankton (mg m^{-3}) productivity of the Pacific coastal waters off Central America during the 2015-16 winter; e) Hovmöller diagram of TMPA precipitation (in mm/day) in the tropical region (5°S - 5°N) showing ENSO events between 1998 - 2015; f) NASA's Aura satellite views nitrogen dioxide (NO_2, 1/cm) from Fort McMurray wildfires in Alberta, Canada in May 2016; and g) Quasi-Biennial Oscillation (QBO) seen from AIRS daily tropical temperature (in Kelvin) time series between 2002-08 at 100 hPa.

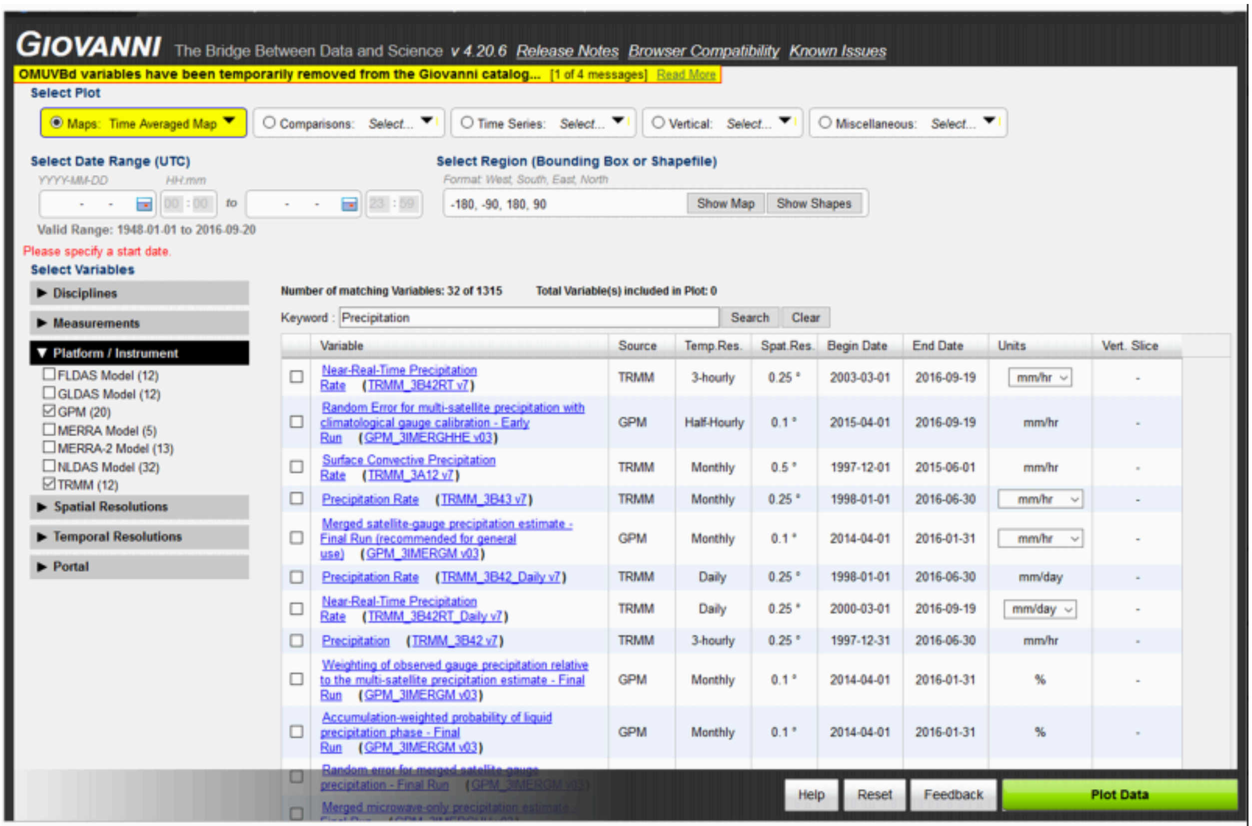

Figure 1. The Giovanni web interface. Over 1300 variables are available for visualization and analysis. Commonly used analytical methods and visualization are available. Keyword and faceted search capabilities allow locating variables of interest with ease. Shapefiles are available for supporting GIS data exploration. Both input and output data can be downloaded for further analysis.

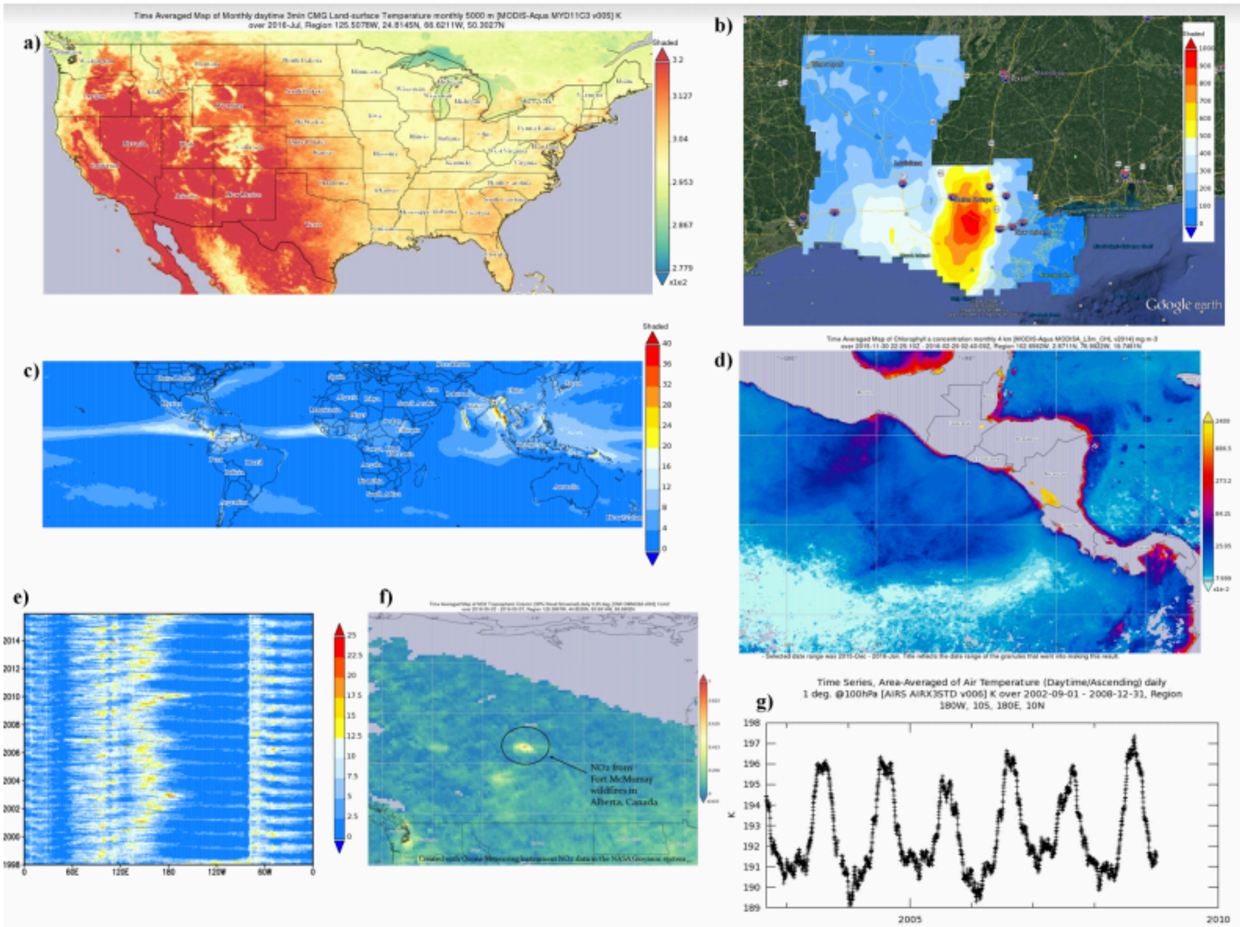

Figure 2. Examples of maps and plots generated from Giovanni: a) July 2016, the hottest month ever - MODIS day surface temperatures (in Kelvin); b) Accumulated rainfall (in mm) from IMERG-Late Run, imported in Google Earth as KMZ, showing a record breaking flood event in Louisiana in August 2016; c) The TMPA precipitation climatology (1998-2015, in mm/day) for the boreal summer (June, July, and August); d) El Niño reduces the phytoplankton (mg m^{-3}) productivity of the Pacific coastal waters off Central America during the 2015-16 winter; e) Hovmöller diagram of TMPA precipitation (in mm/day) in the tropical region (5°S - 5°N) showing ENSO events between 1998 - 2015; f) NASA's Aura satellite views nitrogen dioxide (NO_2, 1/cm) from Fort McMurray wildfires in Alberta, Canada in May 2016; and g) Quasi-Biennial Oscillation (QBO) seen from AIRS daily tropical temperature (in Kelvin) time series between 2002-08 at 100 hPa.