Facilitate Visualization and Distribution of NASA's Environmental Science Data through Open Standards and Open Source Software for Geospatial

Yaxing Wei, Zhaoying (Angie) Wei, Suresh Vannan

Environmental Sciences Division (ESD), Oak Ridge National Laboratory (ORNL)
P.O. Box 2008, MS 6407, Oak Ridge, TN 37831, USA
weiy@ornl.gov, weiz@ornl.gov, santhanavans@ornl.gov

ABSTRACT

This paper introduces the utilization of open standards and open source software for visualization and distribution of geospatial environmental science data at the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). The ORNL DAAC is one of the NASA Earth Observing System Data and Information System (EOSDIS) data centers. A big challenge for the ORNL DAAC (https://daac.ornl.gov) is to efficiently manage over a thousand heterogeneous environmental data, collected through field campaigns, aircraft/satellite observations, and model simulations. ORNL DAAC also has to provide tools to easily find, visualize, and access the heterogeneous data. To address this challenge, the ORNL DAAC has leveraged Open Geospatial Consortium (OGC) standards and open source software to develop the Spatial Data Access Tool (SDAT, https://webmap.ornl.gov/ogc). SDAT is a suite of open standards-based web mapping, subsetting, and transformation services and applications that allow users to visualize and download geospatial data in customized spatial/temporal extents, formats, and projections. The open source MapServer/Geospatial Data Abstraction Library (GDAL) powers the backend OGC Web services of SDAT. Open source Javascript libraries, including OpenLayers, GeoExt, and proj4js, were used to create the SDAT Web User Interface and MapWidget, a light-weight Javascript library that allows SDAT visualization to be easily embedded on any webpage. SDAT also provides KML files to enable interactive data visualization in the popular Google Earth application or any KML-compatible client.

SDAT provides a common framework and standard service interfaces for ORNL DAAC data holdings. SDAT user interface hides their heterogeneity from end users, and promotes their usage. SDAT facilitates integration of ORNL DAAC data resources with other related data systems. In 2016, SDAT served more than 2 million mapping requests and 72 thousand customized data downloads from over 2500 distinct data users.

1. INTRODUCTION

National Aeronautics and Space Administration (NASA)'s investments on Earth observation provide a rich set of resources open to a wide variety of users, including Earth and environmental science researchers, decision makers, and public users, to increase our understanding of Earth processes and enhance predictive capabilities that underpin sound decision-making. NASA commits to full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms, and field campaigns with all users as soon as such data become available[1]. Standardization and interoperability are keys to ensure these valuable but heterogeneous data resources can be easily accessed, used, and fused together to address Earth

[1] NASA Earth Science. https://science.nasa.gov/earth-science.

and environmental issues. Free and open source software for geospatial contributed by the community enable data systems and services to be easily created, maintained, and rapidly evolved with relatively low cost. NASA is an active contributor and adopter of such free and open source software (https://software.nasa.gov).

1.1. ORNL DAAC and NASA EOSDIS

The ORNL DAAC for biogeochemical dynamics (https://daac.ornl.gov) is one of NASA Earth Observing System Data and Information System (EOSDIS) data centers managed by the Earth Science Data and Information System (ESDIS) Project. The EOSDIS is a key core capability in NASA's Earth Science Data Systems Program. It provides end-to-end capabilities for managing NASA's Earth science data from various sources – satellites, aircrafts, field measurements, and various other programs. The EOSDIS science operations are performed within a distributed system of many interconnected and distributed, discipline-specific DAACs. DAACs are responsible for archive and distribution of Earth science data products. The DAACs serve a large and diverse user community by providing capabilities to search and access science data products and specialized services[2]. Seamless integration across these distributed nodes is a key to providing consistent service experience to a broad range of users.

The mission of the ORNL DAAC is to assemble, distribute, and provide data services for a comprehensive archive of terrestrial biogeochemistry and ecological dynamics observations and models to facilitate research, education, and decision-making in support of NASA's Earth Science[3].

1.2. **Challenges Faced by the ORNL DAAC and NASA EOSDIS**

Within NASA EOSDIS, the ORNL DAAC is unique in terms of the diversity of data it archives, manages, and distributes. ORNL DAAC serves as a repository for data relatively small in size but disparate in variety from principal investigators of field campaigns, airborne missions, remote sensing validation projects, and modeling activities. The ORNL DAAC was formed in 1994 and has archived 1,258 data sets (as of May 26, 2017). These 1000+ data sets contain 373 variables captured/produced from 282 sensors of 125 different sources (satellites, flux towers, airplanes, etc.) by 1410 researchers (Vannan, et al., 2016). As illustrated by Vannan, et al. (2016), the number of files per data set at the ORNL DAAC ranges from 1 to 92,619 and data volume per data set ranges from 1 KB to more than 10 GB (as of Jan 31, 2017). More importantly, environmental data formats have changed significantly along with the evolution of data capture mechanisms in the past 20 years. Figure 3 of Vannan et al., 2016 shows binary (*.dat) and compressed zip (*.z and*.gz) files constitute 95% of the data files in the ORNL DAAC archive, mainly because binary formats were popular in the 1990s when geospatial file format standards were immature. Data in old binary formats are not readily usable in common data analysis tools and require special data reader programs, often written in Fortran. Total volume of ORNL DAAC data holdings has been increasing more rapidly in recent years and had doubled in 3 years from ~90 TB in 2013 to ~183 TB in 2015. All these factors raised significant challenges for the ORNL DAAC to serve its user community, including data providers and data

[2] An Overview of EOSDIS. https://earthdata.nasa.gov/about. Accessed on May 28, 2017.
[3] Mission Statement and Goals of ORNL DAAC. http://daac.ornl.gov/who.shtml. Accessed on May 28, 2017.

users. It is critical to improve the usability of these rich sets of heterogeneous data so that their usage barrier can be further lowered for end users and their value can be maximized and preserved longer into the future.

Expanding the scope from the ORNL DAAC to NASA EOSDIS, data heterogeneity and volume become even bigger challenges. There is an increasing need of improving the interoperability of both data products and data systems so that distributed data archives (DAACs) under the NASA EOSDIS umbrella can be seamlessly integrated to ensure that data providers and users have consistent experience with NASA EOSDIS to archive data they create and find, access, and use NASA Earth science data products.

1.3. **Motivation**

To address the above described challenges and to improve the usability of increasingly growing ORNL DAAC-archived data products, there is need to:

1) convert data files in outdated formats into current open and commonly-used standard formats
2) enhance data files to be self-descriptive by embedding metadata information following open standards
3) allow users to interactively visualize and explore data prior to download
4) enable users to subset data they need in Geographic Information System (GIS)-ready forms instead of entire original data files
5) improve interoperability of data products and associated visualization and distribution services by following open standards
6) enable the ORNL DAAC data system to be easily integrated into broader data systems

The Spatial Data Access Tool (SDAT) is an effort at the ORNL DAAC targeting the above six goals.

2. **OGC STANDARDS**

Interoperability is a major goal of the ORNL DAAC SDAT effort. Interoperability between geospatial data systems requires a common set of rules and concepts that define a consistent understanding of the information and operations available in every participating system. Standardization processes and interoperability initiatives such as those of the Open Geospatial Consortium (OGC) try to provide an agreed-on set of such rules and concepts (Vckovski, A., 1999). OGC is an international consortium of companies, government agencies, research organizations, and universities participating in a consensus process to develop publicly available interface specifications. The objective of OGC standards is to promote geospatial information interoperability. OGC Standards adopted by the SDAT, as listed below, provide the foundation for the ORNL DAAC to develop geospatial data visualization and distribution Web services and applications that can be used by end data users and common software and tools that support OGC standards.

- Web Coverage Service (WCS)
- Web Feature Service (WFS)

- Web Map Service (WMS)
- Keyhole Markup Language (KML)
- Styled Layer Descriptor (SLD)
- Well-Known Text Coordinate Reference System (WKT CRS)

WCS supports electronic interchange of geospatial data as "coverages" – that is, digital geospatial information representing space-varying phenomena that relate a spatio-temporal domain to a range of properties. A WCS provides access to coverage data in forms that are useful for client-side rendering, as input into scientific models, and for other clients (OGC, 2005). At ORNL DAAC, the WCS standard is used to distribute geospatial imagery data files like Land Use/Land Coverage Change data and ecosystem model simulation data (Wei, et al., 2009).

WFS provides an interface allowing access to discrete geospatial feature data encoded in Geography Markup Language (GML) (OGC, 2007a). Unlike coverages, geospatial features are described by a set of properties where each property can be thought of as a {name, type, value} tuple (OGC, 2010). The vector data files archived at ORNL DAAC such as field campaign site location information, Global River Discharge field data fall into this category and can leverage WFS for data distribution.

WMS, unlike WCS or WFS that provides access to geospatial data itself, produces styled maps of georeferenced data, including both coverage and feature data. A "map" is defined as a visual representation of geospatial data (OGC, 2001). WMS can be used to visualize data prior to downloading. SLD defines an encoding that extends the WMS standard to allow user-defined symbolization and coloring of geographic feature and coverage data (OGC, 2007b). SDAT utilizes SLD to allow users to easily provide their own or import existing map styles when making WMS requests.

KML provides an XML notation for expressing geographic annotation and visualization within Internet-based, two-dimensional maps and three-dimensional Earth browsers, e.g. Google Earth and NASA World Wind (OGC, 2015a).

WKT CRS is an OGC standard jointly developed by International Organization for Standardization (ISO) and OGC. It provides well-known text string representation of coordinate reference systems that follows the provisions of ISO 19111:2007 and ISO 19111-2:2009 (OGC, 2015b). WKT CRS is leveraged by SDAT to provide standard representation of the CRSes associated with geospatial data and maps generated by SDAT, so that they can be automatically recognized and imported into majority of commercial and open source GIS software.

Among those OGC standards supported by SDAT, WMS and KML have been adopted by NASA ESDIS Standards Office (ESO) and recommended for use in NASA Earth Science Data Systems.

3. SPATIAL DATA ACCESS TOOL

3.1. Overview

SDAT is a Web-based framework developed at the ORNL DAAC to ensure archived geospatial data are in open and standard formats. It provides easy and interoperable OGC Web services for users and compatible clients to retrieve geospatial data and maps based on their specific needs. The SDAT framework as illustrated by Fig. 1 contains a family of Web-based services and applications for geospatial data visualization and exploration, transformation (i.e. subset, reformat, resampling, and reprojection), and data download.

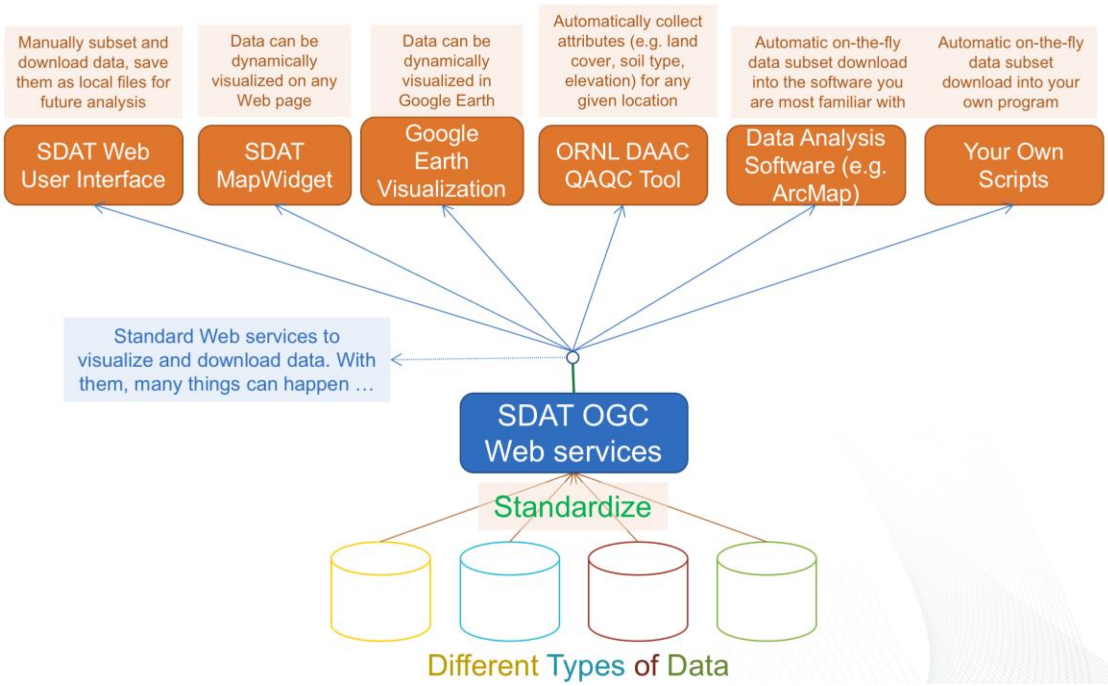

Figure 1. Spatial Data Access Tool (SDAT) framework

At the foundation of SDAT are well-curated geospatial data in open standard formats, which are then ingested into SDAT OGC Web services (i.e. WCS, WFS, and WMS) and made available to users through standard service interfaces. The SDAT Web User Interface (UI), the SDAT MapWidget, and Google Earth Visualization are the three major Web applications in the SDAT framework that provide advanced visualization, transformation, and data download capabilities to users who are not familiar with OGC standards. Interoperability achieved through SDAT OGC Web services allows other applications, such as the ORNL DAAC Identify Tool[4], to be easily developed utilizing SDAT's OGC Web services. Of course, many data analysis software, including common commercial and open source GIS tools, already supports OGC standards and can consume SDAT OGC Web services as remote geospatial data resources directly inside analysis workspace.

At this moment, SDAT WFS is in testing phase. The ORNL DAAC is investigating the benefits that WFS can bring to users and the amount of effort needed to curate feature-based geospatial data, such as field and airborne observations, and make them available through WFS. As a data center focusing on field campaigns and land validation, OGC WFS may have a lot of potentials for the ORNL DAAC and its users.

[4] ORNL DAAC Identify Tool. https://webmap.ornl.gov/qaqc/identify.

3.2. SDAT Implementation

3.2.1. Data Curation

The SDAT workflow starts with curating a wide variety of geospatial data into open standard formats. GeoTIFF is the most popularly-used open format for geospatial imagery data in the GIS community. It was selected as the primary format for imagery data prepared for SDAT WCS and WMS. When SDAT first started, it focused on data sets that had been archived at the ORNL DAAC in the past. Important data sets (e.g. most downloaded), especially those stored in outdated formats, were prioritized for the SDAT effort. Data files in pure binary, text, or other legacy formats were converted into the GeoTIFF format and enhanced with rich metadata information, including standard CRS representation, variable names, and data units. In recent years, with the maturing of the Climate & Forecast (CF) convention (Eaton, et al., 2011), the NetCDF format is gaining popularity rapidly. Well-curated, especially spatially and temporally, CF-compatible netCDF data format is another option for preparing imagery data for SDAT. Recently, all SDAT data preparation requirements have been incorporated into ORNL DAAC's standard data publication workflow. This ensures every geospatial data set prepared by the ORNL DAAC's data publication team is ready for SDAT without any further processing, which helped SDAT to shift its focus from past data sets to present and future ones.

As of September 2017, 147 ORNL DAAC data sets (containing 8,185 data files) have been curated and added to the SDAT and OGC Web services.

3.2.2. Metadata Database

Implementation of SDAT relies on a centralized metadata database, which stores metadata information for both data sets (data set table) and data files (data granule table), to automate the configuration of data layers and the mapping between layer identifiers and data files. "Data layer" is a concept used in OGC Web services. In SDAT, a data set contains multiple data files, or granules, and each data granule is further mapped onto one or more data layers, depending on the number of variables a data file contains.

The data set table contains field "dataset_id", which stores the unique ID for each data set. Data set IDs are consistent with ORNL DAAC's core database and provide a logical linkage between SDAT and the ORNL DAAC data system. Other general metadata, such as dataset name, description, citation, and keywords, of the data set table are also synchronized with the ORNL DAAC's core database. To automate the generation of a common color scheme for all layers in a dataset, dataset-level data value statistics, i.e. maximum, minimum, mean, and standard deviation values, are included in the data set table.

In the data granule table, the "granule_id" field, specifically created for SDAT, contains the unique identifier of each data granule, which is used to construct data layer descriptors in the SDAT OGC Web services. Granule-level metadata contain descriptive information (filename, variables, data type, data unit, resolution etc.), file-level statistical information (maximum, minimum, mean, standard deviation, and missing values), as well as the spatial and temporal extent information (i.e. bounding box, native projection, start time, end time, time series). The statistical metadata is used to automatically generate a color scheme for each individual data layer in OGC Web services. The spatial and temporal extent metadata facilitate the subsetting based on geographical/temporal extent of the data.

3.2.3. **SDAT OGC Web Services**

The core components of SDAT are the OGC Web services (i.e. WCS, WFS, and WMS) that provide geospatial data visualization, transformation, and access capabilities. To maximize the usage of these Web Services, we chose to support the most widely-used versions of OGC standards, which are not necessarily the latest versions. WCS version 1.0.0, WMS version 1.1.1, and WFS version 2.0 are currently supported in SDAT.

Implementation of the SDAT OGC Web services is based on MapServer, which is an Open Source platform for publishing spatial data and interactive mapping applications to the web. Originally developed in the mid-1990's at the University of Minnesota, MapServer is released under an MIT-style license[5]. When the SDAT effort started, there were multiple mapping platform options to choose from, including the commercial ArcIMS/ArcGIS Server and the open source MapServer and GeoServer. Commercial platforms were not preferred since a non-open platform will significantly increase development and maintenance cost in the long run. MapServer was eventually selected as the mapping engine for SDAT for the following reasons: 1) MapServer had outstanding geospatial mapping performance, especially with FastCGI enabled on Linux operating systems (OSGeo, 2011); 2) It supported a large array of imagery and feature data formats using the open source GDAL/OGR library[6]; and 3) Even though MapServer lacked a nice Web Graphic UI for managing data layers, its mapfile-based configuration mechanism allowed data layers to be easily managed in MapServer and automated through shell scripts or MapScript API.

MapServer, along with the open source GDAL/OGR and PROJ libraries, provide the geospatial data mapping and transformation capabilities of the SDAT OGC WMS and WCS. The SDAT WCS supports spatial and temporal subsetting. 6 different popular data formats (ArcInfo/ASCII Grid, Erdas Imagine, GeoTIFF, NetCDF, NITF, and XYZ) are supported in the SDAT WCS as output formats for users/clients to choose from. For each data format, users/clients can choose what internal data type (floating point, integer, or byte) to use. These 6 data formats include both open and proprietary formats to better support user needs. At this moment, each data layer available through SDAT OGC WMS and WCS supports a fixed list of target CRSes, which include the data layer's native CRS, commonly-used CRSes (e.g. EPSG:4326 and Web Mercator) across all data layers, and project-specific CRSes (for example, the ABoVE project[7] has defined CRSes suitable for data sets produced by ABoVE). In the future, SDAT OGC Web services will be enhanced to allow users/clients to provide customized CRSes instead of choosing from a list of pre-defined CRSes.

3.2.4. **SDAT Manager: automate data configuration into SDAT**

Making well-curated data available through SDAT OGC Web services with low effort is one major design goal of SDAT. To minimize human effort needed, the SDAT Manager was developed to automate majority housekeeping tasks. The SDAT Manager is a Web-based dashboard application accessible only by SDAT team members. Supported by backend server-side scripts and the SDAT metadata database, the SDAT Manager provides team members easy

---

[5] MapServer, open source web mapping. http://mapserver.org.
[6] Geospatial Data Abstraction Library. http://www.gdal.org.
[7] NASA Arctic-Boreal Vulnerability Experiment (ABoVE). https://above.nasa.gov.

controls for automatic creation of map color schemes and mapfiles needed by MapServer. The Manager also supports the generation of static preview images for each data layer, update of OGC Capabilities XML files when data layers are added or updated, creation of KMZ packages for Google Earth visualization, production of SLD files containing default color schemes that users can customize for their own needs, and miscellaneous metadata database management tasks (Fig. 2). The SDAT Manager has significantly streamlined the data publication process for SDAT and its OGC Web services.

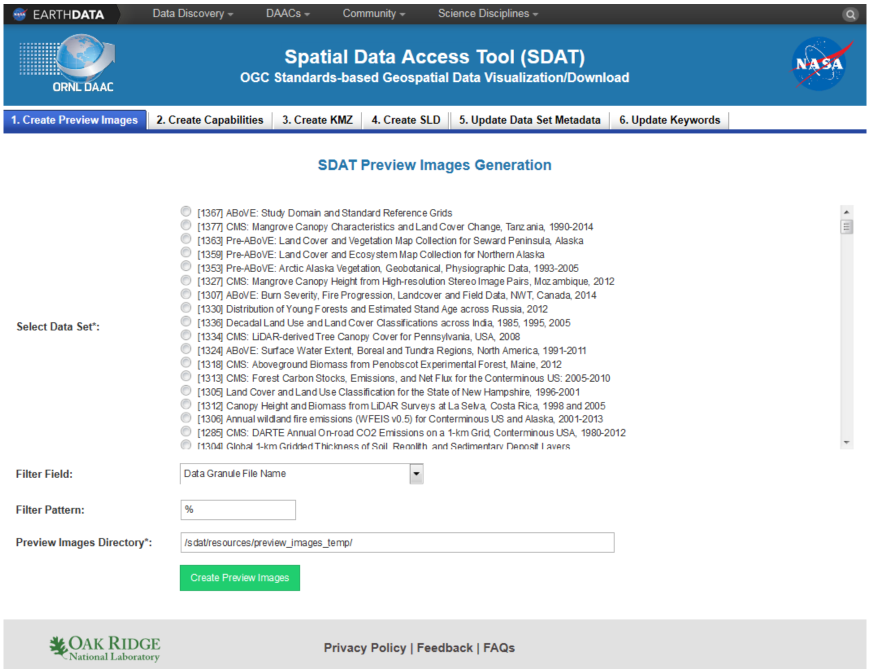

Figure 2. SDAT Manager dashboard

3.2.5. SDAT Key Applications and Use Scenarios

3.2.5.1. SDAT Web UI

To provide advanced capabilities of SDAT OGC Web services to non-expert users who are not familiar with OGC standards, the SDAT Web UI (https://webmap.ornl.gov) was developed to allow users to search, visualize, explore, and download transformed data they need through an intuitive Web user interface. The SDAT Web UI organizes data in two levels: data set level and data granule level. A data set contains one or more data granules (or files), following the same data organization as the ORNL DAAC data archive.

As shown in Fig. 3, the home page of SDAT Web UI presents users a paginated list of data sets. Users can filter data sets through two groups of filters: 1) science disciplines, which are topic-level keywords in Earth Science category of Global Change Master Directory (GCMD); and project names. These 167 data sets in SDAT are categorized into 8 science disciplines and 12 projects. The home page also offers keyword-based search functionality, which can be combined with the science disciplines and project filters to further refine search results. For example, Fig. 3 displays 12 matched data sets (out of 167) after applying the "Biosphere" science discipline filter, the "LBA" project filter, as well as keyword "land". Order of data sets in the search results is based on when they were published into SDAT and most recently added data sets are displayed on top.

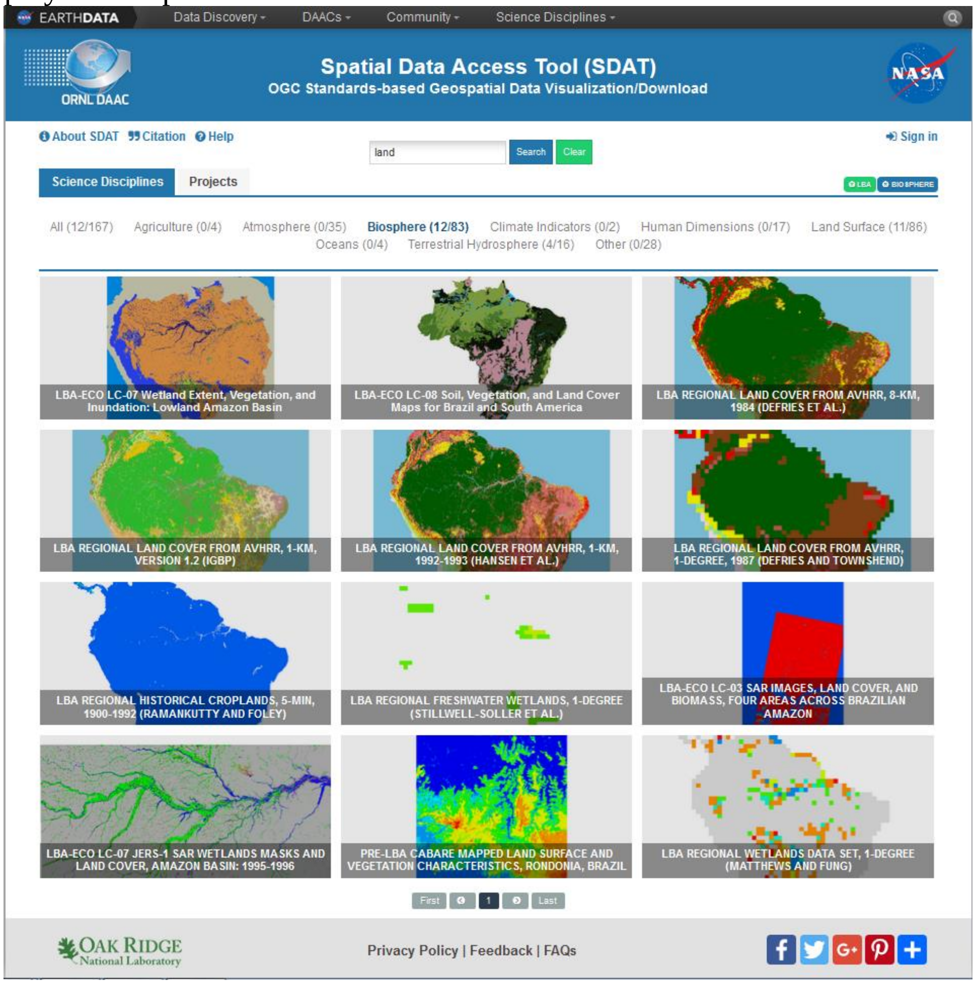

Figure 3. Home page of SDAT Web UI: List of data sets with filters applied

The data download page (Fig. 4) contains 3 sections. The first section displays granule-level metadata, including spatial extent and available time steps. The second section contains a map viewer, dynamically generated by the SDAT OGC WMS service, for users to interactively visualize the geospatial data. The third section gives users control to customize geospatial data download by setting WCS (or WFS) parameters, which provide a series of data transformation options, such as subsetting, reformatting, resampling, and reprojection.

Data transformation options offered by SDAT allow users to download data in ready-to-use forms. For example, to analyze surface water extent changes in boreal and tundra regions, users can compare data of ABoVE surface water extent in two adjacent decades: 2000s and 2010s (Carroll, et al., 2016)[8]. They first choose the CRS (e.g. Web Mercator) needed. Then to specify subset region, users can draw a bounding box over the map, say an area surrounding the Beaver Hill lake in Canada, or enter the spatial extent manually in the Spatial Extent boxes. After selecting the output format and other transformation parameters, users can request data

[8] ABoVE: Surface Water Extent, Boreal and Tundra Regions, North America, 1991-2011. https://webmap.ornl.gov/ogc/dataset.jsp?ds_id=1324.

transformation through SDAT WCS and download data with a single click, once signing in via Earthdata login (https://urs.earthdata.nasa.gov). Data downloaded through SDAT are in a ready-to-use data format and can be directly imported into data analysis software, such as QGIS (Fig. 5), to calculate differences between surface water extents in the 2000s and 2010s in the Beaver Hill lake area. Areas in red color represent shrinking of surface water extent during the last two decades. SDAT facilitates such exploratory data visualization and analysis through standards-based data delivery and visualization.

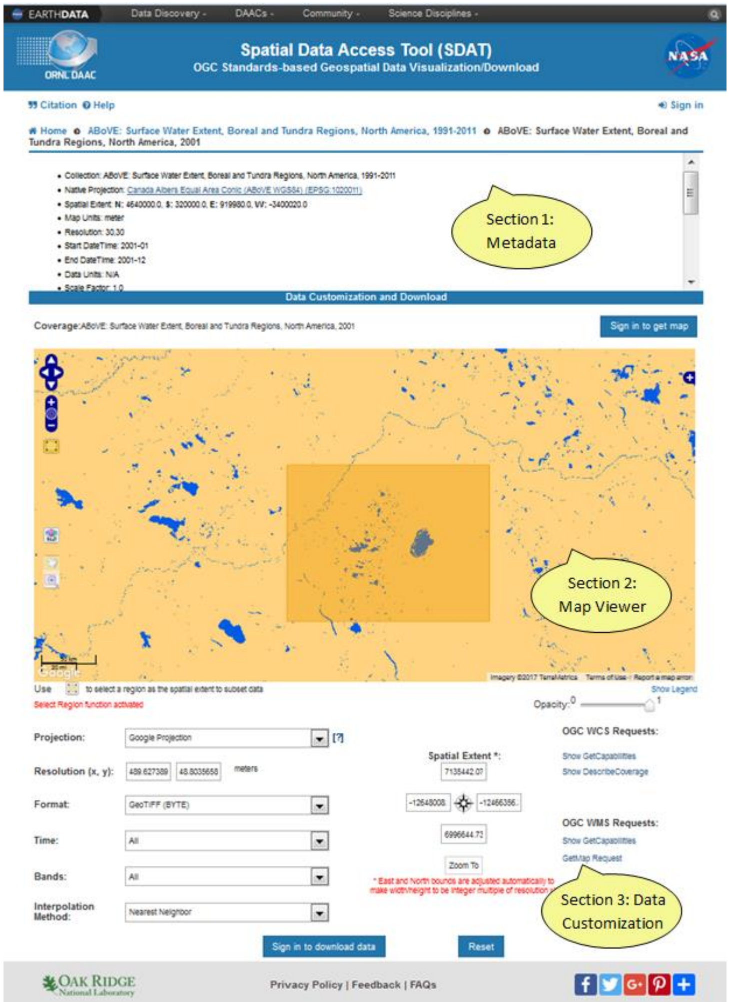

Figure 4. SDAT Web UI - data visualization and download page for ABoVE surface water extent in the 2000s

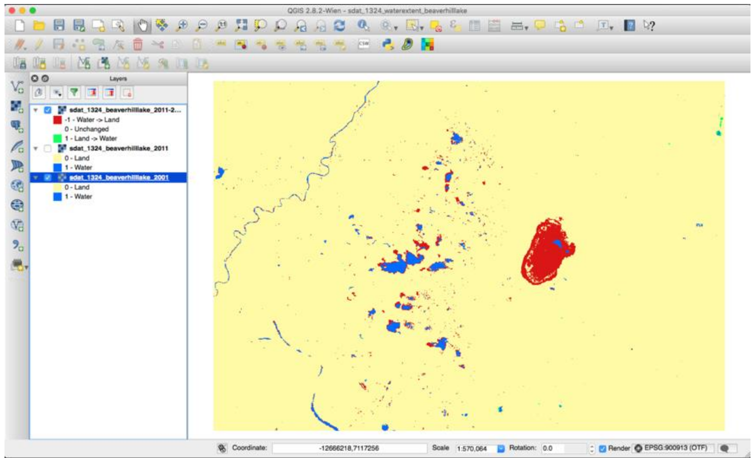

Figure 5. Analysis of surface water extent changes in the Beaver Hill Lake area in QGIS

Implementation of the SDAT Web UI utilizes open source software and libraries, including JQuery, OpenLayers, and Proj4js.

3.2.5.2. SDAT MapWidget

SDAT MapWidget[9] is a JavaScript library that allows a light-weight interactive map viewer to be easily embedded onto any Web page to provide live data visualization, very similar to how Google Map viewer works. SDAT MapWidget is built on top of the open source OpenLayers and GeoExt libraries. The ORNL DAAC has incorporated SDAT MapWidget as a standard component on its data set landing pages to allow users interactively to visualize and explore data without leaving landing pages. This helps users to better evaluate the fitness-to-use of geospatial data through deeper, intuitive, and closer interaction. Fig. 6 shows the SDAT MapWidget viewer on the landing page (https://doi.org/10.3334/ORNLDAAC/1290) of data set "NACP NAFD Project: Forest Disturbance History from Landsat, 1986-2010" (Goward, et al., 2015).

[9] SDAT MapWidget Help. https://webmap.ornl.gov/ogc/help/wms_widget.html.

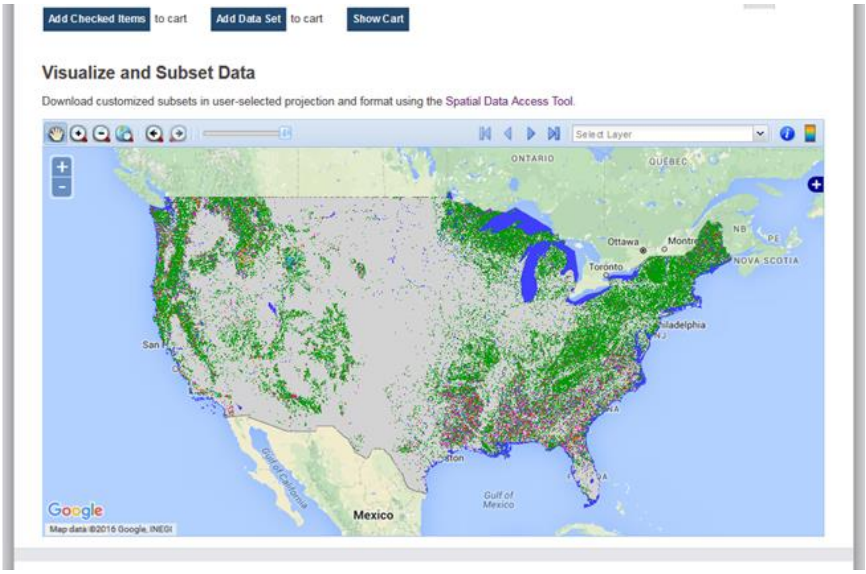

Figure 6. SDAT MapWidget viewer for NACP NAFD Project: Forest Disturbance History from Landsat, 1986-2010

3.2.5.3. Google Earth Visualization

For each data set, SDAT provides a KML/KMZ file that can be dynamically visualized in KML-compatible clients, such as Google Earth. We leveraged an XSLT WMS to KML converter (Benedict and Sanchez-Silva, https://github.com/rnz0/wms2kml) and enhanced it to create KML/KMZ files from WMS Capabilities documents. These KML/KMZ files are configured to retrieve new maps on-the-fly from SDAT OGC WMS service whenever Google Earth viewport changes. Google Earth provides nice features and resources, for example, time-series animation, high-resolution background imageries, and historical Landsat imageries, that enhance users' data exploration experience. Fig. 7 presents visualization of DARTE Annual On-road CO2 Emissions on a 1-km Grid (Gately, et al., 2015) in Google Earth. The time series animation in Google Earth allows users to easily explore changes of annual CO2 emissions from on-road transportation in the Knoxville, TN area from 1980 to 2012. Comparing 3 maps in Fig. 7, on-road CO2 emissions in 1990 were lower than in 1980 and 2000, which could be possibly affected by the early 1990s economic recession.

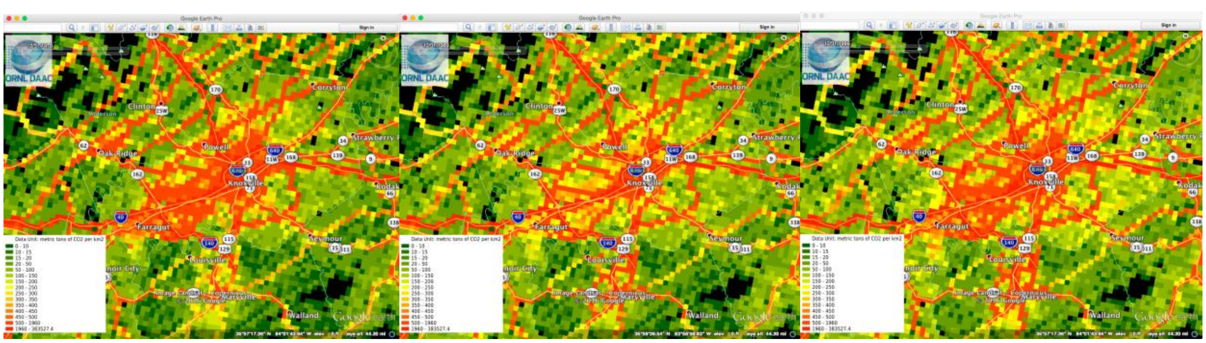

Figure 7. Visualization of annual on-road CO2 emissions in Knoxville, TN area in 1980, 1990, and 2000 in Google Earth

3.2.6. Use SDAT/OGC Data in Common GIS software and programs

For advance users who have knowledge of OGC standards, they can directly access SDAT OGC Web services from many popular GIS software (e.g. ESRI ArcMap and QGIS) and their own programs (e.g. Matlab and Python). In this way, there is no need to download data onto local machines. Remote data available through SDAT OGC Web services can be consumed directly in these software and programs as if they exist locally.

3.3. SDAT Citation

Citation of data products, especially those dynamically transformed through Web services, is being actively discussed within NASA ESDIS and broader data communities (Hausman, et al., 2017) (Smith, et al., 2016) (Rauber, et al., 2015). The ORNL DAAC is a pioneer on providing citations for not only data products but also data services it offers. In the past, data users cited data downloaded via SDAT through footnote or acknowledgement. In early 2017, the ORNL DAAC registered a DOI and created a formal citation for SDAT.

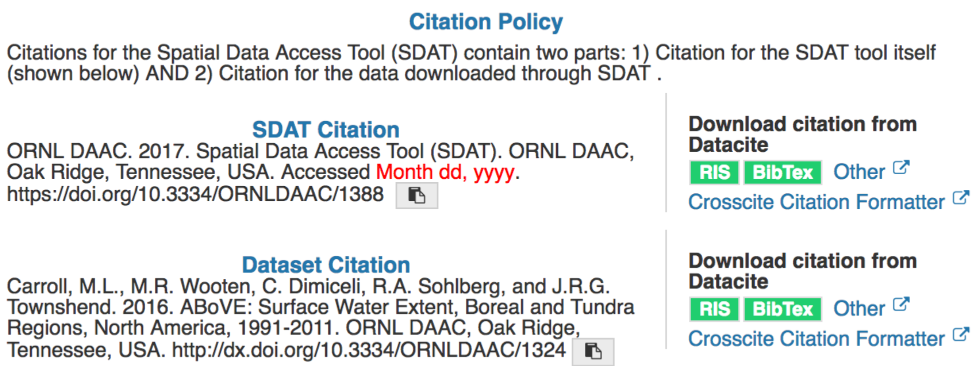

Figure 8. SDAT citation policy popup for the ABoVE Surface Water Extent data set

An initial SDAT citation policy has been set up and provided to users on the SDAT Web UI. As shown in Fig. 8, the SDAT citation policy suggests including citations for SDAT itself and the data set downloaded. SDAT also leverages DataCite and CrossCite Citation Formatters to provide DOI citations in other formats.

4. INTEGRATION OF SDAT/OGC WITH NASA EOSDIS AND BROADER DATA SYSTEMS

4.1. SDAT/OGC - NASA EOSDIS

NASA EOSDIS provides an enterprise data infrastructure that can seamlessly integrate the distributed data systems of all data processing systems and DAACs. This integration involves many different aspects. Metadata integration is achieved through the Unified Metadata Model (UMM)[10] and the Common Metadata Repository (CMR)[11]. The SDAT metadata database will be fully integrated with the ORNL DAAC core database in the future so that SDAT metadata

[10] Unified Metadata Model (UMM), https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository/unified-metadata-model-umm. Accessed on May 28, 2017.
[11] Common Metadata Repository (CMR), https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository. Accessed on May 28, 2017.

elements will be incorporated into UMM and managed by the CMR. In return, SDAT Web UI will be able to leverage the Earthdata Search API to provide data search capability, on both data set and file levels. Efficient data access and transformation is key to utilization of EOSDIS data. The CMR and the Earthdata Search application[12] are capable of brokering data orders to avoid ordering applications or users needing to know DAAC specific ordering systems [24]. This brokering capability has been enabled for OGC WMS services to dynamically display maps on the search result map viewer. NASA EOSDIS is actively seeking solutions to add brokering capability to CMR and Earthdata Search application for OGC WCS services hosted at each individual DAACs. Data in SDAT are freely available to any user, but require user identification per NASA policy to help track detailed data usage at the individual user level, enabling both accurate metrics on data usage and the ability to send critical data update messages (NASA, 2016). SDAT utilizes NASA EOSDIS Earthdata Login for user registration, authentication, and identification and provides single sign-on support across NASA EOSDIS.

4.2. SDAT/OGC - GEOSS

Interoperability enabled by SDAT OGC Web services allows ORNL DAAC-archived data to be easily integrated into broader data systems, such as the Global Earth Observation System of Systems (GEOSS), without any additional effort on the ORNL DAAC side. This is a huge benefit of OGC standards and maximizes the usage and value of ORNL DAAC data. GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. GEOSS adopted a Service-Oriented Architecture (SOA) and set interoperability as its foundation (Khalsa, et al., 2009). The GEOSS Portal offers a single Internet access point for users seeking data, imagery and analytical software packages relevant to all parts of the globe. The GEOSS Portal has integrated all data layers of SDAT WMS through dynamically pulling and parsing Capabilities XML of SDAT WMS. These SDAT WMS layers are available for users to search and visualize directly within the GEOSS Portal (http://www.geoportal.org).

5. SDAT/OGC USAGE ANALYSIS AND FUTURE IMPROVEMENTS

SDAT has been gaining popularity among a broad range of users in the past few years. One reason is the improvements, on the aspects of both performance and usability, we have been continuously making to SDAT. The other reason is the increased support of OGC standards in common GIS and data analysis tools, such as ArcMap, QGIS, Matlab, and R, that people use on a daily basis.

As shown in Fig. 9, usage (number of data downloads, volume of data downloaded, and number of unique user IP addresses) of SDAT WCS has been consistently increasing in the past 6 years (2011-2016). Number of unique WMS user IP addresses has also been consistently increasing in 2011-2016. But the usage of SDAT WMS had a significant jump in 2014 and remained high usage rate thereafter. This was possibly caused by increased visibility of SDAT WMS through integration into broader data systems, which attracted some power geospatial mapping users, especially those based on Cloud computing platforms.

[12] Earthdata Search. https://search.earthdata.nasa.gov.

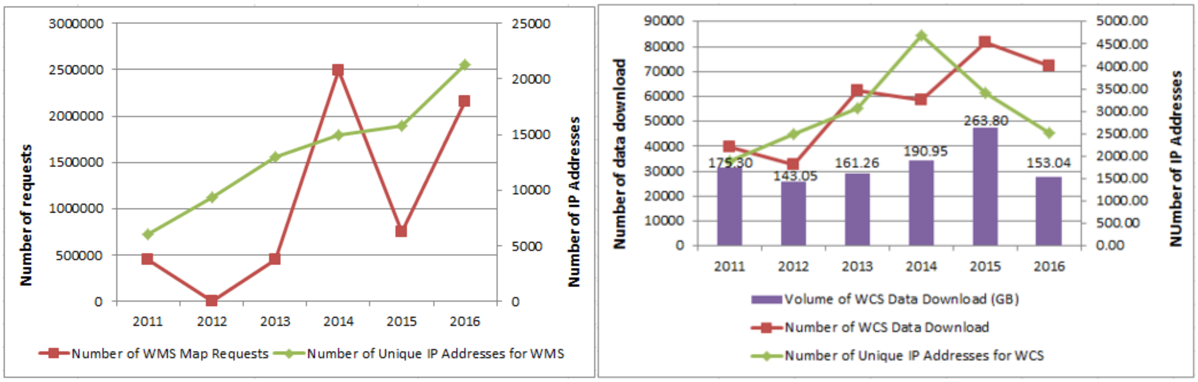

Figure 9. SDAT WCS/WMS usage statistics

We performed further analysis on the 72,000 SDAT WCS data downloads in 2016 to discover to what extent users had utilized the transformation capabilities provided by SDAT WCS. Table 1. shows that GeoTIFF, Arc/Info ASCII Grid, and NetCDF were the most popular data formats users selected. Majority users (~85.1%) leveraged the subsetting capability. Close to half (43.9%) users chose to request data in different spatial resolutions. But only 3.7% users chose to reproject data onto a CRS different from their native CRSes. Reprojection is a critical issue to fuse heterogeneous geospatial data from different sources. Even though SDAT WCS provides reprojection capability, but it was not fully utilized. The main reason is that only limited and fixed target CRS options are available for users. We will enhance SDAT OGC Web services to allow users/clients to provide customized CRSes to request data and maps in the future.

Table 1. Statistics on requested data formats of SDAT WCS service

Arc/Info ASCII Grid	GeoTIFF	Erdas Imagine	NetCDF	NITF	XYZ
12.4%	77.8%	0.65%	7.1%	0.14%	1.91%

One most requested feature that will be added to SDAT in the future is the SDAT Workspace for the SDAT Web UI. The SDAT Workspace is aimed to solve multiple issues. It will allow users to select multiple data files from multiple data sets and add them onto a common viewer for visualization. Also, users can specify a single set of data transformation criteria, e.g. target CRS, spatial extent/resolution, temporal range, and data format, and use them to batch download multiple data files through SDAT WCS with one single click. Results data files, along with metadata and transformation provenance information, will be packed together and accessible by users.

OGC WFS will be further investigated and fully incorporated into the SDAT framework to provide visualization and download capabilities for feature-based geospatial data products, including observations made at field locations and along aircraft flying trajectories.

6. CONCLUSION

NASA's investments on Earth observation provide a rich set of valuable resources open to a wide variety of users, including Earth and environmental science researchers, decision makers, and public users, to increase our understanding of Earth processes and enhances predictive capabilities that underpin sound decision-making. NASA EOSDIS provides a distributed infrastructure that links distributed DAACs, including the ORNL DAAC, through interoperable interfaces and allows a wide variety of users to easily find, visualize, explore, and access Earth science data resources. The open standards and open source software-based SDAT increases usability of the environmental science data products archived at the ORNL DAAC by assuring data are stored in open and self-descriptive standard formats, enabling users to interactively visualize and explore data prior to downloading, allowing users to download data they need in GIS-ready format instead of entire original data files. More importantly, SDAT and the OGC Web services improve the interoperability of the environmental science data, enable the ORNL DAAC data system to be easily integrated into broader data systems, and maximize the usage and value of NASA's environmental science data. The ORNL DAAC SDAT has gained great popularity in the past few years. But there are still many aspects needing improvement, including dynamic data citation, improved data transformation capabilities, better integration with the NASA EOSDIS infrastructure, and the SDAT Workspace concept that allows users to overlay multiple data layers for visualization and batch download multiple data with the same set of transformation criteria.

7. REFERENCES

Carroll, M.L., Wooten, M.R., Dimiceli, C., Sohlberg, R.A., and Townshend, J.R.G., 2016. ABoVE: Surface Water Extent, Boreal and Tundra Regions, North America, 1991-2011. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1324.

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., and Juckes, M., 2011. NetCDF Climate and Forecast (CF) Metadata Conventions, version 1.6, 5 December, 2011.

Gately, C. K., Hutyra, L.R., and Wing, I.S., 2015. CMS: DARTE Annual On-road CO2 Emissions on a 1-km Grid, Conterminous USA, 1980-2012. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1285.

Goward, S.N., Huang, C., Zhao, F., Schleeweis, K., Rishmawi, K., Lindsey, M., Dungan, J.L., and Michaelis, A., 2015. NACP NAFD Project: Forest Disturbance History from Landsat, 1986-2010. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1290.

Hausman, J., Stall, S., Gallagher, J., et al., 2017. Recommendations of NASA ESDSWG 2016-2017 Software Citations Working Group.

Khalsa, S.J.S., Nativi, S., and Geller, G.N., 2009. The GEOSS Interoperability Process Pilot Project (IP3), in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 80-91, Jan. 2009. doi: 10.1109/TGRS.2008.2007304

NASA, 2016. ESDIS Project, *EOSDIS Handbook v1*, November 2016.

OGC, 2005. OGC 05-076, *OpenGIS Web Coverage Service (WCS) Implementation Specification*, version 1.0.0.

OGC, 2007a. OGC 05-078r4, *Styled Layer Descriptor profile of the Web Map Service Implementation Specification*, Version: 1.1.0 (revision 4).

OGC, 2007b. OGC 07-036, *OpenGIS Geography Markup Language (GML) Encoding Standard*, version: 3.2.1.

OGC, 2010. OGC 09-025r1, *OpenGIS Web Feature Service 2.0 Interface Standard*, Version 2.0.0.

OGC, 2001. OGC 01-068r3, *OpenGIS® Web Map Service Implementation Specification*, Version 1.1.1.

OGC, 2015a. OGC 12-007r2, *OGC KML 2.3*.

OGC, 2015b. OGC 12-063r5, *Geographic information - Well-known text representation of coordinate reference systems*.

OSGeo, 2011. WMS Performance Shootout 2011. https://www.slideshare.net/gatewaygeomatics.com/wms-performance-shootout-2011. Accessed May 21, 2017.

Rauber, A., Pröll, S., and RDA Data Citation WG, 2015. Scalable Dynamic Data Citation Approaches, Reference Architectures and Applications. Draft Version - 2015-03-23.

Smith, A.M., Katz, D.S., and Niemeyer, K.E., 2016. FORCE11 Software Citation Working Group. Software Citation Principles. PeerJ Computer Science 2:e86. DOI: 10.7717/peerj-cs.86

Vannan, S., Beaty, T.W., Cook, R.B., Wright, D.M., Devarakonda, R., Wei, Y., Hook, L.A., and McMurry, B.F., 2016. A Semi-Automated Workflow Solution for Data Set Publication. *ISPRS Int. J. Geo-Inf*. 2016, 5, 30. doi: 10.3390/ijgi5030030.

Vckovski, A., 1999. “Interoperability and spatial information theory”. *In Interoperating geographic information systems*, Edited by: Goodchild, M., Egenhofer, M., Fegeas, R. and Kottman, C. Volume 495, pp 31–37. Boston: Kluwer.

Wei, Y., Santhana-Vannan, S. K., and Cook, R. B., 2009. Discover, visualize, and deliver geospatial data through OGC standards-based WebGIS system, 17th International Conference on Geoinformatics, Fairfax, VA, 2009, pp. 1-6. doi: 10.1109/GEOINFORMATICS.2009.5293520.