

NOAA's National Climatic Data Center and Coastal Services Center– Enabling Data Discovery and Interoperability

J. Neal Lott, Rich Baldwin, Jon Burroughs
NOAA National Climatic Data Center, Asheville, North Carolina
Jim Boyd
NOAA Coastal Services Center, Charleston, South Carolina
Glen Reid
I.M. Systems Group at NOAA National Climatic Data Center, Asheville, North Carolina
Jason Marshall
The Baldwin Group at NOAA Coastal Services Center, Charleston, South Carolina
Gabe Sataloff
I.M. Systems Group at NOAA Coastal Services Center, Charleston, South Carolina

ABSTRACT

NOAA's National Climatic Data Center (NCDC) and Coastal Services Center (CSC) have developed a number of products and services over the past several years to provide data and information in forms readily usable by a variety of customers. However, these products can be difficult to locate for users. To address this issue, NCDC and CSC have been partnering to utilize existing metadata standards, Geospatial Information System (GIS) interoperability techniques, and ESRI Geoportal Toolkit software, to enhance data discovery and interoperability for customers. These capabilities have been made available online within products at NCDC, CSC, and the NOAA Climate Services Portal. Examples include GIS Services to locate stations with climatic data, climate monitoring products for various parameters, and ad-hoc search capabilities to allow users to locate datasets based on keywords. In this paper we briefly describe the IT technologies being used, the services available, and a number of examples of the capabilities.

1. Background

One of the key challenges for users of NOAA's climate data and services is finding the appropriate dataset or service for use in any given application. As demands for climate services have increased (for research, private sector, education, etc), the number of datasets, products, and services has grown exponentially. As a result, users are increasingly overwhelmed by the seemingly endless results found in common search engines. For example, search engines such as Google provide massive numbers of results that are ranked in an order that may be irrelevant to data seekers. This adds to the overall confusion as to which product might be the best fit.

Large data-rich agencies, such as NOAA, provide numerous search options from within an organization, but in most cases lack cross-agency integration that is necessary to offer comprehensive results for data seekers. This frustrates users who may need to surf several agency sites to obtain a full understanding of what data is a best fit for a specific use.

Initial efforts to resolve this issue focused on consolidation of data holdings and providing a web mapping interface for locating data of interest. With this method, users can interact with the map service dynamically via a web browser. Usage of common standards, such as the Open Geospatial Consortium (OGC) standards (<http://www.opengeospatial.org/>), formed the basis for many development efforts and provided further momentum to standardize interoperability across NOAA.

Current trends have expanded this paradigm toward allowing users to dynamically search for datasets, products and services of interest using a catalog approach that leverages Federal Geographic Data Committee (FGDC) geospatial metadata standards (<http://www.fgdc.gov/metadata>) as a basis for information discovery.

2. GIS Activities

NOAA's National Climatic Data Center (in Asheville, NC) and Coastal Services Center (in Charleston, SC) have been collaborating in GIS and metadata activities. Each Center provides GIS access to national and global datasets and products. These resources can be found through the NCDC GIS portal page: <http://gis.ncdc.noaa.gov>, and the CSC homepage: <http://www.csc.noaa.gov/>. Access to climate datasets, products, and related services include:

- Global hourly, daily, and monthly climate data
- US hourly and 15 minute precipitation data
- US Climate Reference Network (USCRN) data
- Benthic Habitat data
- NEXRAD level II and III data
- Paleoclimatology data
- US climate indices data

* *Corresponding author address:* J. Neal Lott, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801; e-mail: neal.lott@noaa.gov

- Global summary of the day data
- US long-term normals and extremes
- Historical hurricane tracks
- Coastal zone data (multiple data types)
- Composite Shoreline data products
- Digital elevation data
- Multiple types of aerial imagery
- Land-cover / land-use data
- Hazard-assessment tools
- Many others...

In addition to providing custom mapping interfaces to access data, these websites offer common OGC Web Map Services (WMS) and OGC Web Feature Services (WFS) interfaces which provide images of maps and spatial features respectively. These services may be consumed by applications designed to use these standards and can directly integrate data and metadata. Additional formats, such as Keyhole Markup Language (KML) files used directly within 3-D GIS viewers (e.g., ArcExplorer, Google Earth) are also available for various datasets and products. Datasets hosted within these systems maintain FGDC-compliant metadata, which can be harvested to build metadata catalogs that support data/metadata portals such as Geospatial One-Stop, Data.gov, and the Global Change Master Directory.

Value-added layers are also included with the GIS Service, to allow users to display issue-specific products for agricultural regions, coastal hazards, population density, global ecoregions and wetlands, state and regional climate divisions, etc. These specific products can provide additional functionality as well as more detailed, localized data such as topography digital raster graphics (DRG) and visual imagery digital orthophoto quadrangle (DOQ) data.

Several examples are shown in Figures 1 through 5 below.

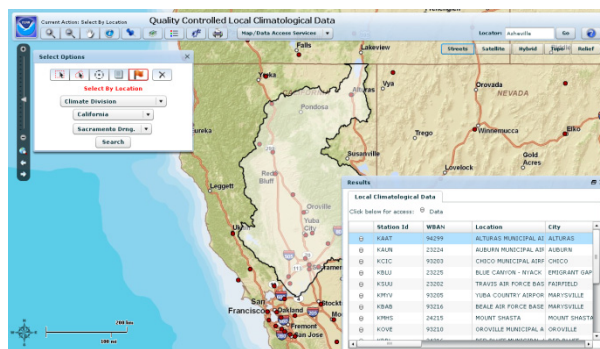


Figure 1. Dynamic Map for “Local Climatological Data” access – one of the more popular climatic products.

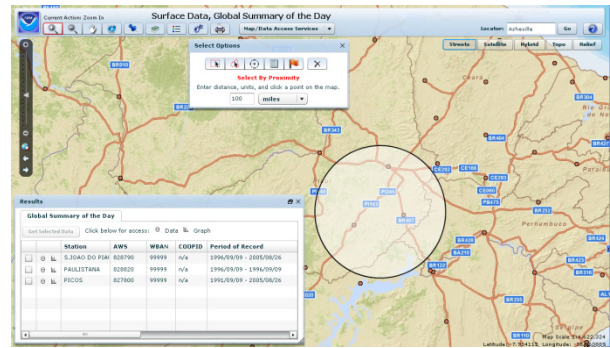


Figure 2. The “Global Summary of the Day” station distribution for a 100 mile radius in Brazil.

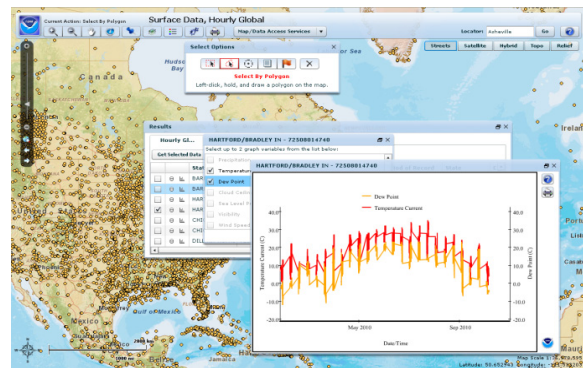


Figure 3. Access to data includes a time-series plotting capability.

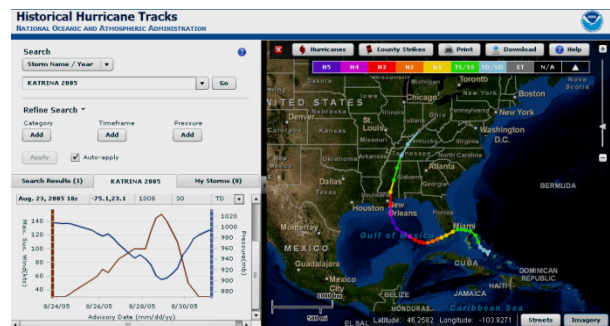


Figure 4. Historical hurricane tracks tool, with Hurricane Katrina (2005) displayed.

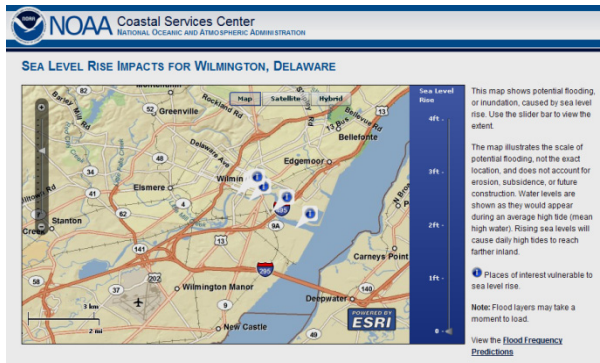


Figure 5. Sea-level rise impacts tool.

3. Data Discovery

While GIS services are one of the cornerstones of robust data access, they alone do not allow NOAA customers to discover the datasets and products which may be most useful in their application. The ability to browse, locate, and download data, whether via a map interface, or simply by downloading the data files, is a critical capability. To achieve this, NCDC and CSC are leveraging FGDC metadata standards as referenced above. Over time, we anticipate the migration of these metadata into the ISO-19115 standard schema (http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020)

These XML-standard formats allow for many different pieces of information about each dataset to be cataloged using Lucene, a common open-source indexing API. Metadata tags, such as time, location, author, theme, etc, can be indexed across data holdings and compiled into catalogs. The consolidated information stored within these catalogs can be leveraged to create Geographic Gazetteers that provide the foundations that allow users to rapidly isolate stations, topics, regions and user-defined areas of interest. With the underlying data and access protocols harmonized, applications making use of the standards can interrogate one another, creating an infrastructure for decentralized catalogs. Since GIS Services are ingesting standard geospatial data, and the catalog systems can be searched by standardized methods, these two technology domains can be unified to provide dynamic discovery and display products that can be used for ad-hoc, dynamic online geo-discovery. For example, the latitude/longitude boundaries of a dataset are included in the metadata record, along with a list of the parameters (eg, temperature, precipitation). This allows external software to search the metadata record – eg, locate all datasets containing temperature data for a particular latitude/longitude point or area.

In order to further enhance the discovery of datasets and services, NCDC and CSC are also leveraging ontological searches. The ontological search links into the General Multilingual Environmental Thesaurus (GEMET), which is exposed through the web

via xml and HTTP (<http://www.eionet.europa.eu/gemet>). The GEMET ontology searches are based not only a specific keyword, but also on associated words or terms to provide more robust result sets.

NCDC and CSC are using the ESRI Geoportal Toolkit as the application to manage the metadata, and to provide various capabilities for users -- most importantly a service endpoint for OGC-compliant Catalog Service for the Web (CS-W). This software package is also being used in the recently released NOAA Climate.Gov website (www.climate.gov). Additionally, ESRI has released the Geoportal Toolkit to the Open Source community. Figures 6 through 8 provide a few samples.

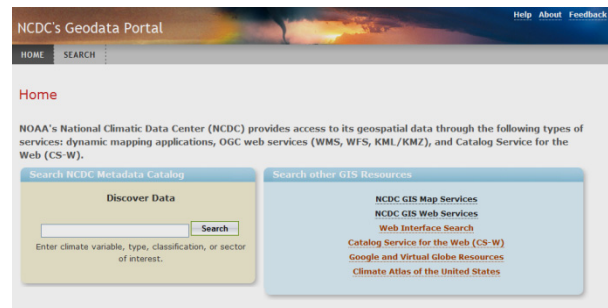


Figure 6. Geoportal "homepage" with various search capabilities displayed.

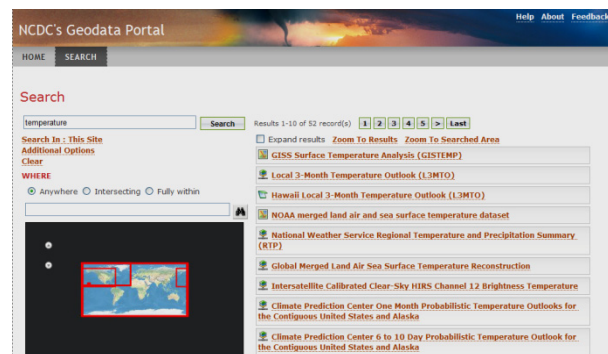


Figure 7. Search results for "temperature" – note that results include products from a number of agencies.

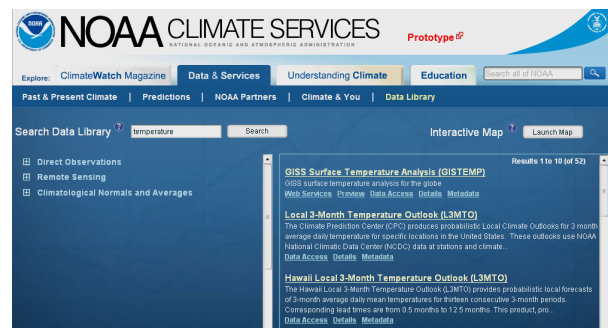


Figure 8. Same result-set ("temperature") displayed in Climate.gov interface.

[illegible]

The screenshot shows the NCEP Climate Data Browser interface. The main map displays the United States with a red rectangular selection box covering the West Coast, from approximately 30°N to 45°N and 120°W to 100°W. A sidebar on the right contains a search results list with the following items:

- 10 matching documents found
- NCEP Precipitation Rate (IR 1A)
- NCEP Precipitation Rate
- WMS Radar Integrated Display with Geospatial...
- San Diego
- GLOBAL Terrain DEM of Los Angeles and Surro...
- GLOBAL Terrain data for the Sacramento Valley...
- Location of Interagency Global Terrain Data
- WORLDWIDE major climate composite reflectivity
- CRS 5000 5000
- FRS3000 NOAA Field 5000 Survey in the snow...

At the bottom of the page, there is a navigation bar with the following links:

- Website
- Contact
- Availability
- Contact
- Details
- Metadata
- Add to Map

NOAA's NCDC and CSC have made some significant strides toward making NOAA data, products, and services more readily available to users, and will continue to improve access methods through the use of standards and improved interoperability. Much work remains to make these capabilities and services as user-friendly as possible to a very wide range of audiences, to expand the number of datasets and products accessible, and to improve interoperability across datasets. We welcome user comments and input, and are expanding our user/sector engagement activities to ensure that all resources expended in these efforts are addressing user requirements.

A key feature of planned improvements includes the ability to search many CS-W services concurrently. What does this mean? Through our geoportal interface we will be able to search any CS-W which we have registered. When a CS-W service is registered (e.g., Data.gov, Global Earth Observation System of Systems, Geospatial One Stop), any number of these resources can be searched simultaneously. This is a dramatic first step in implementing federated metadata searches and eliminates the need to maintain harvesting records from these silos. Searching registered CS-W services will not only broaden the search results, but will also allow users to query repositories that may not have a direct climate focus, such as a geographically focused portal like the NOAA Southeast and Caribbean Data Explorer.