Gstore - A Scalable Platform for Discovery and Access of Earth Science Data

Karl Benedict & Renzo Sanchez Silva
Earth Data Analysis Center, UNM

Abstract

In order to support scalable access to geo-temporal Earth Science and related data, the Earth Data Analysis Center (with support provided by the NSF EPSCoR and NM RGIS Programs) has developed a platform that enables flexible discovery and use of these data. Called Gstore, the developed system provides a unified data model within which heterogeneous vector data products (i.e. attributes associated with points, linear feature or polygonal areas) may be stored, allowing rapid retrieval of values from thousands of separately generated datasets, representing tens of millions of individual features (currently we have over 45 million features stored in the system). The data stored within Gstore are accessible through a variety of standards-compliant interfaces, including REST-based query and meta/data access services and OGC (WxS) services. End products generated by these services include data in a variety of formats including: zip archives containing the original source data for specific datasets, GML and KML files containing features across multiple datasets, ESRI shapefiles for individual datasets, and tabular representations both as CSV files and Excel spreadsheets. The system has been designed for maximum flexibility in forward evolution as all client interactions with the system are through web service interfaces that are extensible through the addition of new interaction models, data generation capabilities, and eventually analytic tools.

Sample Requests

List of theme-subtheme-groupname values (JSON)
http://rgis.unm.edu/gstore/search/datasets.json

List of datasets for all categories (JSON)
http://rgis.unm.edu/gstore/search/datasets.json?offset=0&sort=lastupdate&dir=desc&limit=15

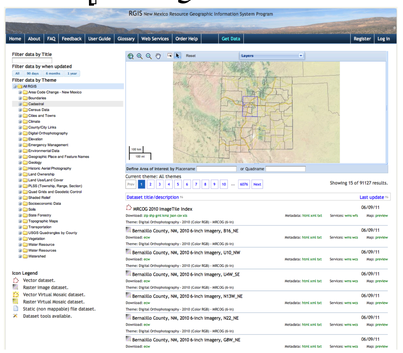
List of datasets for a specific theme-subtheme-groupname combination (JSON)
http://rgis.unm.edu/gstore/search/datasets.json?offset=0&sort=lastupdate&dir=desc&limit=15&theme=Vegetation&subtheme=General&groupname=New+Mexico

Request for a specific file download (KMZ)
http://rgis.unm.edu/gstore/datasets/58330.kmz

GetCapabilities (WMS) Request for a specific dataset
http://gstore.unm.edu/apps/rgis/datasets/58330/services/ogc/wms?VERSION=1.1.1&SERVICE=WMS&REQUEST=GetCapabilities

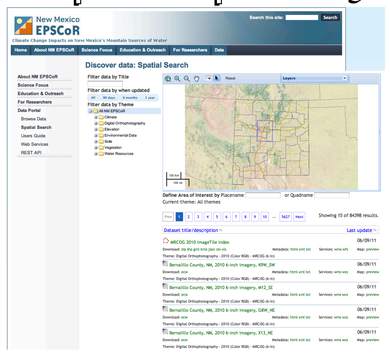
Metadata Request for a specific dataset (HTML)
http://rgis.unm.edu/gstore/datasets/58330/metadata/58330.html

http://rgis.unm.edu



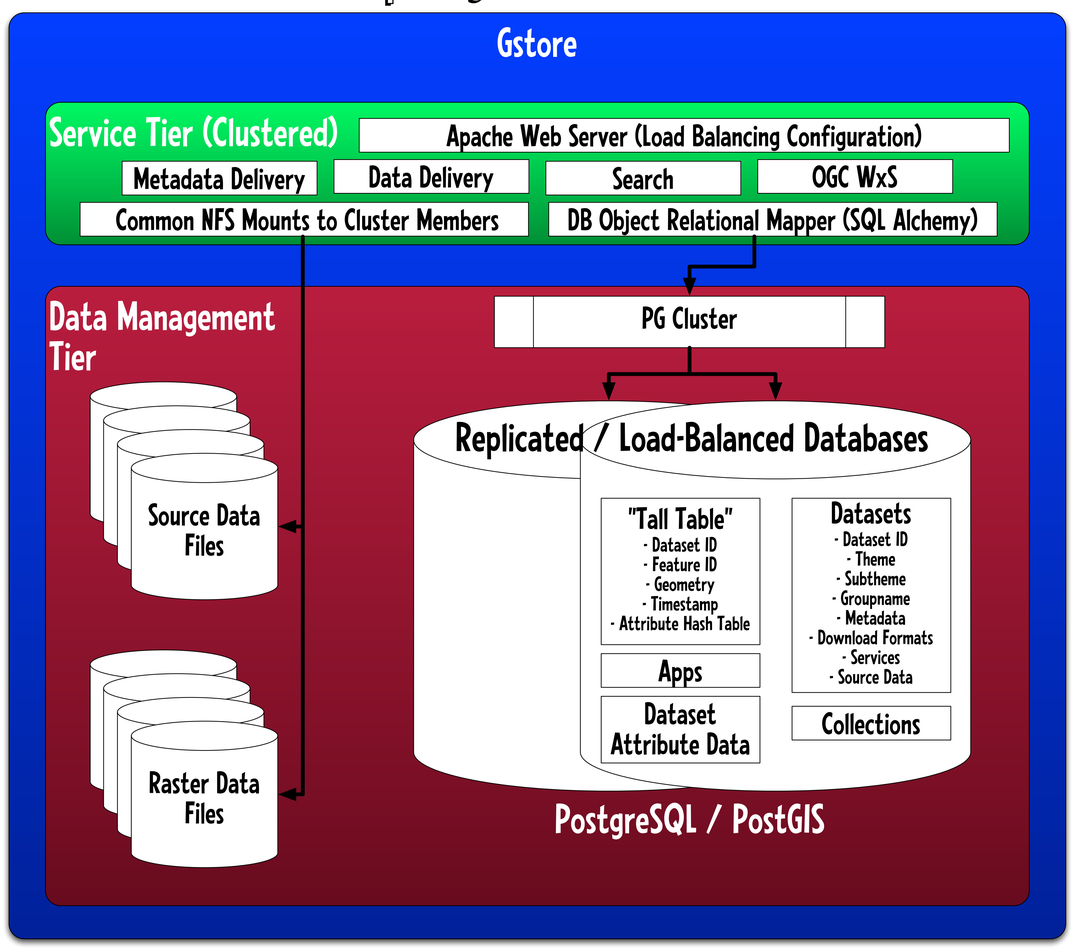
HTML / CSS / Javascript / GeoExt

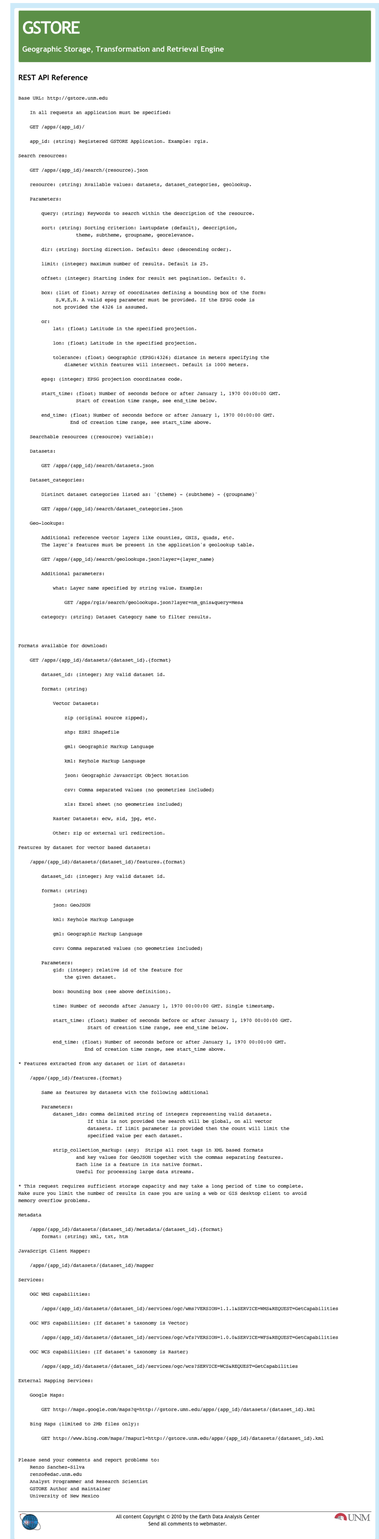
http://nmepscor.org



Drupal / HTML / CSS / Javascript / GeoExt

http://gstore.unm.edu





GSTORE

Geographic Storage, Transformation and Retrieval Engine

REST API Reference

```
base URL: http://gstore.unm.edu

    In all requests an application must be specified:

    GET /apps/{app_id}/

    app_id: (string) Registered GSTORE Application. Example: rgis.

Search resources:

    GET /apps/{app_id}/search/{resource}.json

    resource: (string) Available values: datasets, dataset_categories, geolookup.

    Parameters:

        query: (string) Keywords to search within the description of the resource.

        sort: (string) Sorting criterion: lastupdate (default), description,
                theme, subtheme, groupname, georelevance.

        dir: (string) Sorting direction. Default: desc (descending order).

        limit: (integer) maximum number of results. Default is 25.

        offset: (integer) Starting index for result set pagination. Default: 0.

        bbox: (list of float) Array of coordinates defining a bounding box of the form:
                [x_0,y_0, x_1,y_1] A valid epsg parameter must be provided. If the EPSG code is
                not provided the 4326 is assumed.

        dz:
            lat: (float) Latitude in the specified projection.

            lon: (float) Latitude in the specified projection.

            tolerance: (float) Geographic (EPSG:4326) distance in meters specifying the
                diameter within features will intersect. Default is 1000 meters.

        epsg: (integer) EPSG projection coordinates code.

        start_time: (float) Number of seconds before or after January 1, 1970 00:00:00 GMT.
                Start of creation time range, see end_time below.

        end_time: (float) Number of seconds before or after January 1, 1970 00:00:00 GMT.
                End of creation time range, see start_time above.

    Searchable resources ({resource} Variable):

    Datasets:

        GET /apps/{app_id}/search/datasets.json

    Dataset_categories:

        Distinct dataset categories listed as: '{theme} - {subtheme} - {groupname}'

        GET /apps/{app_id}/search/dataset_categories.json

    Geo-lookups:

        Additional reference vector layers like counties, GNIS, quads, etc.
        The layer's features must be present in the application's geolookup table.

        GET /apps/{app_id}/search/geolookups.json?layer={layer_name}

        Additional parameters:

            what: Layer name specified by string value. Example:

                GET /apps/rgis/search/geoLookups.json?layer=nm_gmisaquery=mesa

            categoryr: (string) Dataset Category name to filter results.


Formats available for download:

    GET /apps/{app_id}/datasets/{dataset_id}.{format}

        dataset_id: (integer) Any valid dataset id.

        format: (string)

            Vector Datasets:

                zip (original source zipped),

                shp: ESRI shapefile

                gml: Geographic Markup Language

                kml: Keyhole Markup Language

                json: Geographic Javascript Object Notation

                csv: Comma separated values (no geometries included)

                xls: Excel sheet (no geometries included)

            Raster Datasets: ecw, aid, jpg, etc.

            Other: zip or external url redirection.

Features by dataset for vector based datasets:

    /apps/{app_id}/datasets/{dataset_id}/features.{format}

        dataset_id: (integer) Any valid dataset id.

        format: (string)

            json: GeoJSON

            kml: Keyhole Markup Language

            gml: Geographic Markup Language

            csv: Comma separated values (no geometries included)

        Parameters:
            gid: (integer) relative id of the feature for
                the given dataset.

            bbox: Bounding box (see above definition).

            time: Number of seconds after January 1, 1970 00:00:00 GMT. Single timestamp.

            start_time: (float) Number of seconds before or after January 1, 1970 00:00:00 GMT.
                        Start of creation time range, see end_time below.

            end_time: (float) Number of seconds before or after January 1, 1970 00:00:00 GMT.
                        End of creation time range, see start_time above.

* Features extracted from any dataset or list of datasets:

    /apps/{app_id}/features.{format}

        Same as features by datasets with the following additional

        Parameters:
            dataset_ids: comma delimited string of integers representing valid datasets.
                    If this is not provided the search will be global, on all vector
                    datasets. If limit parameter is provided then the count will limit the
                    specified value per each dataset.

            strip_collection_markup: (any)  Strips all root tags in XML based formats
                    and key values for GeoJSON together with the commas separating features.
                    Each line is a feature in its native format.
                    Useful for processing large data streams.

* This request requires sufficient storage capacity and may take a long period of time to complete.
Make sure you limit the number of results in case you are using a web or GIS desktop client to avoid
memory overflow problems.

Metadata

    /apps/{app_id}/datasets/{dataset_id}/metadata/{dataset_id}.{format}
        format: (string) XML, txt, htm

Javascript Client Mapper:

    /apps/{app_id}/datasets/{dataset_id}/mapper

Services:

    OGC WMS capabilities:

        /apps/{app_id}/datasets/{dataset_id}/services/ogc/wms?VERSION=1.1.1&SERVICE=WMS&REQUEST=GetCapabilities

    OGC WFS capabilities: (if dataset's taxonomy is Vector)

        /apps/{app_id}/datasets/{dataset_id}/services/ogc/wfs?VERSION=1.0.0&SERVICE=WFS&REQUEST=GetCapabilities

    OGC WCS capabilities: (if dataset's taxonomy is Raster)

        /apps/{app_id}/datasets/{dataset_id}/services/ogc/wcs?SERVICE=WCS&REQUEST=GetCapabilities

External Mapping Services:

    Google Maps:

        GET http://maps.google.com/maps?q=http://gstore.unm.edu/apps/{app_id}/datasets/{dataset_id}.kml

    Bing Maps (limited to 2MB files only):

        GET http://www.bing.com/maps/?mapurl=http://gstore.unm.edu/apps/{app_id}/datasets/{dataset_id}.kml


Please send your comments and report problems to:
    Renzo Sanchez-Silva
    renzo@edac.unm.edu
    Analyst Programmer and Research Scientist
    GSTORE Author and Maintainer
    University of New Mexico
```

All content Copyright © 2010 by the Earth Data Analysis Center
Send all comments to webmaster.

The New Mexico EPSCoR Program is funded in part by the National Science Foundation award #0814449 and the State of New Mexico. The NM RGIS program is funded by the State of New Mexico