



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Google Dataset Search & CMR

WGISS-47

Doug Newman

EED-2 Data Use Architect

douglas.j.newman@nasa.gov

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.
This document does not contain technology or Technical Data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

Agenda

- **Concept**
- **Implementation**
- **Results**
- **Tooling**
- **Conclusions**
- **Future**

CONCEPT

What was at stake

Traditionally it has been difficult to get your collection landing pages near the top of a Google search.

First exposure

‘Facilitating Dataset Discovery using new developments within schema.org’ @ Summer ESIP 2013, Peter Fox

Adding schema.org ‘Dataset’ markup to your collection landing pages *might* solve the problem

IMPLEMENTATION

The players

CMR

The Common Metadata Repository (CMR) is a high-performance, high-quality, continuously evolving metadata system that catalogs Earth Science data and associated service metadata records. These metadata records are registered, modified, discovered, and accessed through programmatic interfaces leveraging standard protocols and APIs.

<https://cmr.earthdata.nasa.gov/search/>

Schema.org semantic markup

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. Over 10 million sites use Schema.org to markup their web pages and email messages. Many applications from Google, Microsoft, Pinterest, Yandex and others already use these vocabularies to power rich, extensible experiences.

<https://schema.org/>

In collaboration

CMR collection landing pages, rendered in HTML can provide markup conforming to the schema.org standard to convey semantic information to commercial search engines and, in theory, allow users to discover and acquire NASA's earth science data in an intuitive and efficient manner.

Strategy

1. Comprehensive site map of cmr.earthdata.nasa.gov
2. RDFa schema.org tags embedded into each collection's html landing page.

Schema.org mappings

CMR concept	Schema.org dataset concept
Entry Title	Name
Short Name	Alternate Name
Version ID	Version
Description	Description
Spatial extent of dataset	Spatial Coverage
Temporal extent of dataset	Temporal Coverage
GCMD Science Keywords for dataset	Keywords
DOI	DOI
Author/DAAC/POC	Citation
Creation date	Date Created
Last update date	Date Modified
Provider	Provider
Online Access URLs	URL of type distribution
Other URLs	URL

Example markup (1 of 3)

```
<div itemscope itemtype="http://schema.org/Dataset">
  <meta itemprop="name" content="MODIS/Terra Calibrated Radiances 5-Min L1B Swath 250m V006"/>
  <meta itemprop="alternateName" content="MOD02QKM_6"/>
  <meta itemprop="version" content="6"/>
  ...
</div>

<span itemprop='description'>The MODIS/Terra Calibrated Radiances 5-Min L1B Swath 250m...</span>

<meta itemprop="spatialCoverage">
  <div vocab="http://schema.org/" typeof="Place">
    <div property="geo" typeof="GeoShape">
      <meta property="box" content="90.0 -180.0 90.0 180.0" />
    </div>
  </div>
</meta>

<time itemprop="temporalCoverage" datetime="2000-02-24T00:00:00.000Z/..">2000-02-24 to present</time>

<meta itemprop="keywords" content="EARTH SCIENCE,SPECTRAL/ENGINEERING,INFRARED WAVELENGTHS"/>
```

Example markup (2 of 3)

```
<li itemprop="identifier" itemscope="" itemType="http://schema.org/PropertyValue">
  <h5>DOI</h5>
  <p>
    <meta itemprop="propertyID" content="DOI">
    <span itemprop="value">10.5067/MODIS/MOD02QKM.006</span>
  </p>
</li>

<li itemprop="citation" itemType="http://schema.org/CreativeWork" itemscope>
  <h5 itemprop="headline">MODIS/Terra Calibrated Radiances 5-Min L1B Swath 250m V006</h5>
  <span itemprop="author">MCST Team</span>
  <span itemprop="publisher">L1 and Atmosphere Archive and Distribution System (LAADS)</span>
  <a itemprop="url" href="http://example.com">https://dx.doi.org/10.5067/MODIS/MOD02QKM.006</a>
</li>

<time itemprop='dateCreated' datetime='2012-11-05T00:00:00.000Z'>2012-11-05T00:00:00.000Z</time>
<time itemprop='dateModified' datetime='2017-12-28T00:00:00.000Z'>2017-12-28T00:00:00.000Z</time>

<h5 itemprop='provider'>GSFC</h5>
```

Example markup (3 of 3)

```
<a href="https://modaps.nascom.nasa.gov/services/about/product_descriptions_terra.html" itemprop="url">
  https://modaps.nascom.nasa.gov/services/about/product_descriptions_terra.html
</a>
<a href="https://ladsweb.modaps.eosdis.nasa.gov" itemprop="distribution" itemscope="itemscope" itemtype="http://schema.org/DataDownload">
  <meta itemprop="contentUrl" content="https://ladsweb.modaps.eosdis.nasa.gov"/>
  https://ladsweb.modaps.eosdis.nasa.gov
</a>
```

RESULTS

The long wait

We waited 5 years.

We continued to report back to ESIP and made modifications based on the evolution of schema.org

But no tangible results from searches of Google.

Finally...

In 2018, NASA, along with other agencies, were asked to help with the development of a bespoke Google search engine for science data.

Google Dataset Search Beta

Search for Datasets



Try [boston education data](#) or [weather site:noaa.gov](#)

[Learn more](#) about including your datasets in Dataset Search.

Image source: <https://toolbox.google.com/datasetsearch>

An example from CMR

Google Dataset Search

MOD021KM

About



Feedback

15 results found

E TERRA MODIS MODIS/Terra Level 1B calibrated relocated...
www.europeandataportal.eu

D MODIS/Terra Calibrated Radiances 5-Min L1B Swath...
catalog.data.gov
Updated Mar 23, 2019

N MODIS/Terra Calibrated Radiances 5-Min L1B Swath 1k...
cmr.earthdata.nasa.gov
Updated Sep 27, 2018

D MODIS/Terra Level 1B Subsampled Calibrated...
catalog.data.gov
Updated Mar 23, 2019

N MODIS/Terra Calibrated Radiances 5-Min L1B Swath 1k...
data.nasa.gov
Updated Jan 30, 2019

N MODIS/Terra Calibrated Radiances 5-Min L1B Swath 1k...

MODIS/Terra Calibrated Radiances 5-Min L1B Swath 1km V006 MOD021KM_6

cmr.earthdata.nasa.gov

DOI link

<https://doi.org/10.5067/MODIS/MOD021KM.006>

Dataset created Nov 5, 2012

Dataset updated Sep 27, 2018

Dataset provided by

NASA/GSFC/SED/ESD/HBSL/BISB/MODAPS
NASA/GSFC/SED/ESD/HBSL/BSB/MCST
NASA/GSFC/SED/ESD/HBSL/BISB/LAADS

Time period covered

Feb 24, 2000 - Present

Description

The MODIS/Terra Calibrated Radiances 5-Min L1B Swath 1km (MOD021KM) contains calibrated and geolocated at-aperture radiances for 36 discrete bands located in the 0.4 to 14.4 micron region of the electromagnetic spectrum. These data are generated from the MODIS Level 1A scans of raw radiance which during processing are converted to geophysical units of $W/(m^2 \text{ um sr})$. In addition, the Earth Bi-directional Reflectance Distribution Function (BRDF) may be determined for the solar reflective bands (1-19, 26) through knowledge of the solar irradiance (e.g., determined from MODIS solar diffuser data, and from the target illumination geometry). Additional data are provided including quality flags, error estimates and calibration data. Visible, shortwave infrared, and near infrared measurements are only made during the daytime, while radiances for the thermal infrared region (bands 20-25, 27-36) are measured continuously. The Shortname for this product is MOD021KM and is stored in the Earth Observing System Hierarchical Data Format (HDF-EOS). A typical file size will be approximately 115 MB. Environmental information derived from MODIS L1B measurements will offer a comprehensive and unprecedented look at terrestrial, atmospheric, and ocean phenomenology for a wide and diverse community of users throughout the world. See the MODIS Characterization Support Team webpage for more C6 product information at: <https://mcst.gsfc.nasa.gov/l1b/product-information> or visit Science Team homepage at: <https://modis.gsfc.nasa.gov/data/dataproduct/>

Testing our markup

Schema org element	Search query	Comments
Alternate Name	MOD021KM	
Name	MODIS/Terra Calibrated Radiances 5-Min L1B Swath 1km	
Version	MOD021KM 6.1	
Temporal	MOD021KM 1999	No results
Temporal	MOD021KM 2018	Results
DOI	10.5067/MODIS/MOD021KM.061S	Single result
Science Keywords	VISIBLE RADIANCE MOD021KM	Results
Science Keywords	ultraviolet radiance MOD021KM	No results

TOOLING

Google Dataset Guidelines

Google Search

HOME GUIDES REFERENCE TOOLS HELP

Overview

Structured data

Article

Breadcrumb

Book

Carousel

Corporate contact

Course

Dataset

Employer Aggregate Rating

Event

Fact Check

Job Posting

Local Business

Logo

Media

Occupation

Product

Recipe

Review

Sitelinks searchbox

Social profile

Software App

Speakable

Subscription and paywalled content

Top Places List

Video

RSS feeds

Crawling and indexing

Indexing API

Required properties

description	Text
	A short summary describing a dataset.
name	Text
	A descriptive name of a dataset. For example, "Snow depth in Northern Hemisphere".

Recommended properties

citation	Text or CreativeWork
	A citation for a publication that describes the dataset. For example, "J.Smith 'How I created an awesome dataset', Journal of Data Science, 1966".
identifier	URL, Text, or PropertyValue
	An identifier for the dataset, such as a DOI.
keywords	Text
	Keywords summarizing the dataset.
license	URL, Text
	A license under which the dataset is distributed.
sameAs	URL
	A link to a page that provides more information about the same dataset, usually in a different repository.
spatialCoverage	Text, Place
	You can provide a single point that describes the spatial aspect of the dataset. Only include this property if the dataset has a spatial dimension. For example, a single point where all the measurements were collected, or the coordinates of a bounding box for an area.

Points

```
"spatialCoverage": {  
  "@type": "Place",  
  "geo": {  
    "@type": "GeoCoordinates",  
    "latitude": 42,  
    "longitude": -71  
  }  
}
```

Contents

- Our approach to dataset discovery
- Examples
- Guidelines
 - Sitemap best practices
 - Source and provenance best practices
- Known Errors and Warnings
- Structured data type definitions
 - Dataset**
 - DataCatalog
 - DataDownload
 - Tabular datasets
- Help and tools

Image source: <https://developers.google.com/search/docs/data-types/dataset>

Structured Data Testing Tool

Google Structured Data Testing Tool



https://cmr.earthdata.nasa.gov/search/concepts/C203234490-LAADS

NEW TEST

```
1 <!DOCTYPE html>
2 <!--[if lt IE 7]> <html class="no-js lt-ie9 lt-ie8 lt-ie7" lang="en"> <![endif-->
3 <!--[if IE 7]> <html class="no-js lt-ie9 lt-ie8" lang="en"> <![endif-->
4 <!--[if IE 8]> <html class="no-js lt-ie9" lang="en"> <![endif-->
5 <!--[if gt IE 8]><!--> <html class="no-js" lang="en"> <!--<![endif-->
6 <head>
7 <meta charset="utf-8">
8 <meta http-equiv="X-UA-Compatible" content="IE=edge">
9 <title>MODIS/Terra Calibrated Radiances 5-Min L1B Swath 500m V006</title>
10 <meta name="description" content="">
11
12 <!-- Open Sans Font - Google Font -->
13 <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Open+Sans:400,700">
14 <!-- Font Awesome: Use this link for development since icons do not show up properly on local dev -->
15 <!-- TODO huh? fix this ^ -->
16 <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.css">
17 <!-- Font Awesome: Local file linked below for production -->
18 <!--<link rel="stylesheet" href="/dist/stylesheets/font-awesome.min.css"-->
19 <!--link rel="stylesheet" href="/dist/stylesheets/mmt.css"-->
20 <link rel="stylesheet" media="all" href="/search/stylesheets/application.css" />
21 <script src="/search/javascripts/application.js"></script>
22 </head>
23 <body class="">
24 <main class="internal record" role="main">
25 <header>
26
27 <div class="row content">
28 <div class="collection-basics">
29
30 <h2 >MOD02HKM_6</h2>
31
32 <p class="subtitle">MODIS/Terra Calibrated Radiances 5-Min L1B Swath 500m V006</p>
33 </div>
34
35 <div class="collection-details">
36 <!-- Only display version if one exists. -->
37
38 <span class="eui-badge version">Version 6</span>
39
40 <!-- Only display data language if one exists. -->
41
42 <span class="eui-badge language">eng</span>
43
44 <!-- Only display CollectionDataType if it is NRT. -->
45
46 </div>
47
48 </div>
```

dateCreated	https://modis.gsfc.nasa.gov/data/dataproduct/
dateCreated	2012-11-05T00:00:00+00:00
url	https://modaps.nascom.nasa.gov/services/about/product_descriptions_terra.html
sameAs	https://mst.gsfc.nasa.gov/l1b/product-information
keywords	EARTH SCIENCE,SPECTRAL/ENGINEERING,INFRARED WAVELENGTHS,INFRARED RADIANCE,REFLECTED INFRARED,VISIBLE WAVELENGTHS,VISIBLE RADIANCE,IMAGERY/BASE MAPS/EARTH COVER
dateCreated	2012-07-02T00:00:00+00:00
dateModified	2018-09-27T00:00:00+00:00
temporalCoverage	2000-02-24T00:00:00.000Z/
potentialAction	
@type	SearchAction
target	
@type	EntryPoint
uriTemplate	https://search.earthdata.nasa.gov/search/granules?p=C203234490-LAADS&q={query}
distribution	
@type	DataDownload
contentUrl	https://ladsweb.modaps.eosdis.nasa.gov
distribution	
@type	DataDownload
contentUrl	https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/6/MOD02HKM/
provider	
@type	Thing
name	NASA/GSFC/SED/ESD/HBSL/BSB/MCST
provider	
@type	Thing
name	NASA/GSFC/SED/ESD/HBSL/BISB/MODAPS
provider	
@type	Thing
name	NASA/GSFC/SED/ESD/HBSL/BISB/LAADS
citation	

Image source: <https://search.google.com/structured-data/testing-tool>

CONCLUSIONS

Google's Dataset Search Tool provides an excellent entry point for earth data science user data discovery.

Rainfall example

Google Dataset Search

rainfall in south dakota

About

Feedback

9 results found

 General Investigation of Lake Andes, South Dakota: Lake-...
catalog.data.gov
data.doi.gov
Updated Feb 22, 2019

 EnviroAtlas - Potential Evapotranspiration 1950 - 209...
catalog.data.gov
data.wu.ac.at
Updated Feb 8, 2018

 Data from simulations of ecological and hydrologic...
data.doi.gov
data.wu.ac.at
Updated Nov 6, 2018

 General Investigation of Lake Andes, South Dakota: Lake-L...
data.world
Updated Sep 25, 2018

 USACE CWMS - Mississippi River Watershed MVP
www.hydroshare.org



Data from simulations of ecological and hydrologic response to climate change scenarios at Wind Cave National Park, South Dakota, 1901-2050

 

Dataset updated Nov 6, 2018

Dataset provided by
[United States Geological Survey](#)

Description

This data release contains data discussed in its larger work citation. "ClimateComparisonData.csv" contains summary metrics of six climate projections used as climate input for quantitative simulations of hydrologic and ecological responses to climate change at Wind Cave National Park (WCNP) and the same summary metrics for 38 other climate projections available at the time that these simulations were done. "HydroData.csv" contains mean annual streamflow of a stream in WCNP and mean annual hydraulic head of a subterranean lake in Wind Cave as simulated by the rainfall-response aquifer and watershed flow (RRAWFLOW) model for two climate projections in the climate dataset. The remaining files contain aboveground live forest carbon, frequency of high-fire-danger days, and annual grass production as simulated by the dynamic vegetation model MC1 parameterized for WCNP for combinations of four climate projections in the climate dataset with a variety of management alternatives.

FUTURE

Upstream content providers

CMR provides metadata to both nasa.data.gov and data.gov

Google also ranks their search results higher than that of CMR.

CMR have begun an effort to increase the semantic content of nasa.data.gov and, by extension, data.gov to resolve this problem.

Search to Dataset Search

There is no concrete link between Google Search and Google Dataset Search.

It should emulate Google Jobs, for example.

Search Action usage

When searching
with an Amtrak.com

One could
between
Earthdata

```
{
  "@context": "http://schema.org",
  "@type": "ServiceChannel",
  "url": "https://search.earthdata.nasa.gov",
  "providesService": {
    "@type": "Service",
    "name": "Earthdata Search",
    "url": "https://search.earthdata.nasa.gov",
    "potentialAction": {
      "@type": "SearchAction",
      "target": "https://search.earthdata.nasa.gov/search?q={collection}&qt={temporal}&sb={box}",
      "query-input": [
        {
          "@type": "PropertyValueSpecification",
          "valueRequired": false,
          "valueName": "temporal",
          "defaultValue": {
            "@id": "schema:datasetTimeInterval",
            "@type": "Property"
          }
        },
        {
          "@type": "PropertyValueSpecification",
          "valueRequired": false,
          "valueName": "box",
          "defaultValue": {
            "@type": "Place",
            "geo": {
              "@type": "GeoShape",
              "box": "-90.0000 180.0000 90.0000 -180.0000"
            }
          }
        },
        {
          "@type": "PropertyValueSpecification",
          "valueRequired": false,
          "valueName": "query"
        }
      ]
    }
  }
}
```

provided

relationship
d

QUESTIONS

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.

Raytheon

*in partnership
with*

