

# Generating Structured Metadata via the GeoCODES User Interface using Schema.org and the Project 418 Geoscience Extensions for Indexing by Commercial Search Engines

Sidney Hellman<sup>1</sup>, Stefan Lisowski<sup>2</sup>, Eric Lingerfelt<sup>3</sup>, Maura Allen<sup>4</sup>, and Alexander McNurlan<sup>4</sup>

<sup>1</sup>Instrumental Software Tech Inc

<sup>2</sup>Instrumental Software Technologies, Inc. (ISTI)

<sup>3</sup>University Corporation for Atmospheric Research

<sup>4</sup>Instrumental Software Technologies, Inc.

November 21, 2022

## Abstract

Using web standards including Schema.org and JSON-LD, the GeoCODES project extends Schema.org with Project 418's geoscience specific vocabulary. By embedding properly formatted and populated JSON-LD files in web sites serving geolocated datasets, search engines such as Google and Bing are able to parse and index these data sets and then to provide information concerning these datasets via standard web search tools. Due to the difficult nature of properly formatting and populating these JSON-LD structures, the GeoCODES User Interface was created to guide data providers through the process of describing the data and validating the descriptions against standard vocabularies. The result is user friendly and easily extensible web based, mobile device ready tool for automatically generating JSON-LD metadata for organizations and datasets. This ultimately allows the original data to be found and used by both scientists and the public.

# Generating Structured Metadata via the GeoCODES User Interface using <http://Schema.org> and the Project 418 Geoscience Extensions for Indexing by Commercial Search Engines

**Generating Structured Metadata via the GeoCODES User Interface using <http://Schema.org> and the Project 418 Geoscience Extensions for Indexing by Commercial Search Engines**

Sidney B Hellman, Stefan Lisowski, Eric Lingerfelt, Maura Allen and Alexander McNurlan  
Instrumental Software Technologies, Inc. (ISTI), University Corporation for Atmospheric Research/EarthCube Science Support Office (UCAR/ESSO)

**Easy to use web page helps scientists quickly describe their data in a way that allows Google to present that data in search results**

**Introduction**  
In 2010, the EarthCube Council of Data Centers (CCDC), the Coalition for Publishing Data in the Earth and Space Sciences (CPDSE), and the Registry of Research Data Repositories (re3data) released the Registry Working Group (RWG) recommendations for a common and machine-readable method for sharing information about organizations and their data holdings.  
In 2017, the EarthCube Architecture Refinement Workshop (ARW) was held at UCAR/ESSO, which provided a new way to describe data holdings.

**Make Data FAIR - Findable Accessible Interoperable Reusable**

**The process**  
<https://www.earthcube.org/webapps/geocodes/registration/>

**schema.org**

**Example**

```

{
  "@context": "http://schema.org",
  "@type": "Organization",
  "name": "EarthCube",
  "url": "http://www.earthcube.org",
  "description": "EarthCube is a national research infrastructure for the Earth and Space Sciences (ESS) that provides a common and machine-readable method for sharing information about organizations and their data holdings.",
  "address": "1000 Central Expressway, Suite 100, Boulder, CO 80506",
  "contactPoint": {
    "@type": "ContactPoint",
    "email": "earthcube@ucar.edu",
    "telephone": "303 440 1500"
  },
  "funding": "http://www.earthcube.org/funding/ESSOP/"
}

```

Sidney B Hellman, Stefan Lisowski, Eric Lingerfelt, Maura Allen and Alexander McNurlan

Instrumental Software Technologies, Inc. (ISTI), University Corporation for Atmospheric Research/EarthCube Science Support Office (UCAR/ESSO)



PRESENTED AT:



# EASY TO USE WEB PAGE HELPS SCIENTISTS QUICKLY DESCRIBE THEIR DATA IN A WAY THAT ALLOWS GOOGLE TO PRESENT THAT DATA IN SEARCH RESULTS

## Introduction

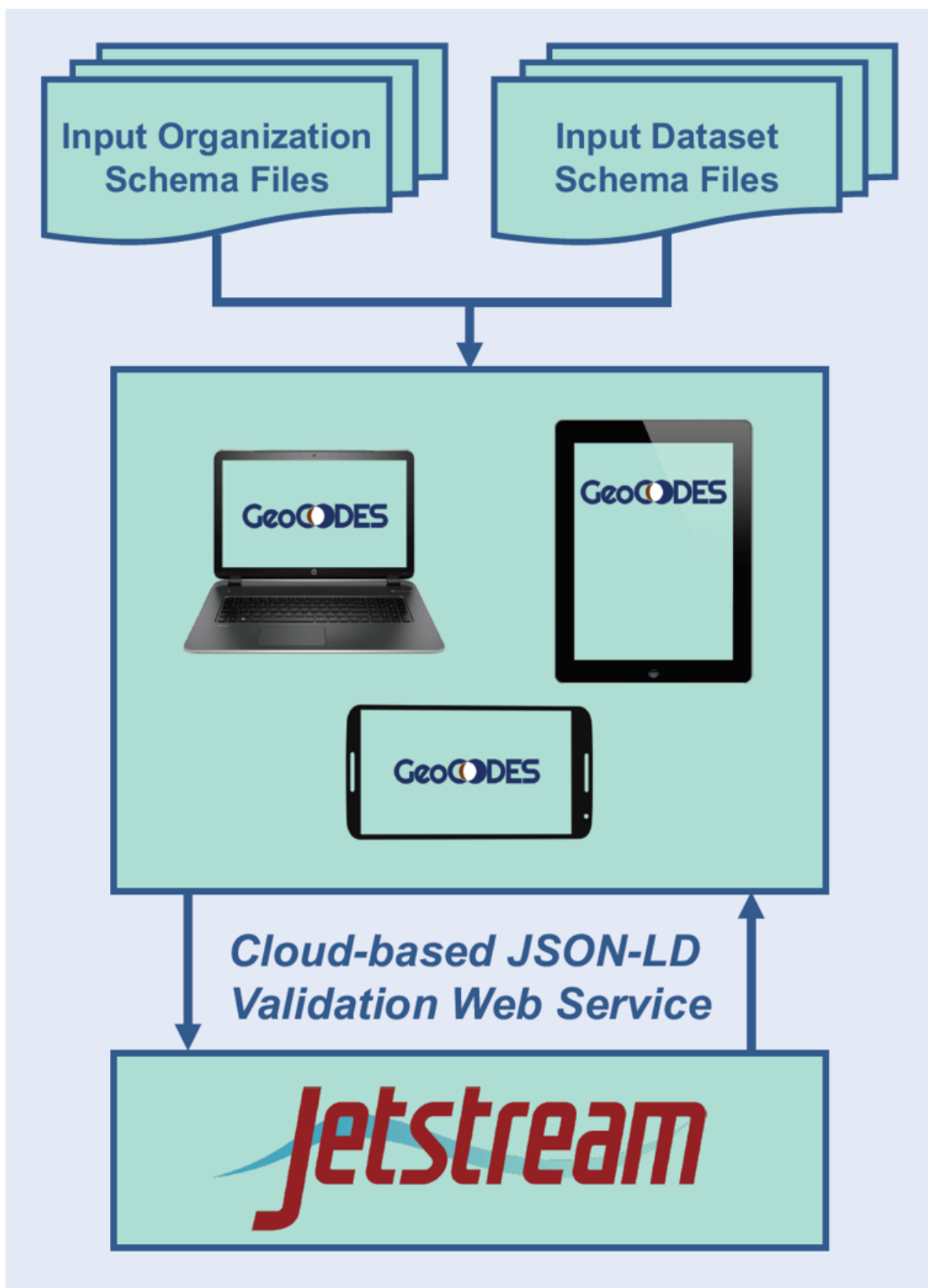
In 2016, the EarthCube Council of Data Facilities (CDF), the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS), and the Registry of Research Data Repositories (re3data) established the Registry Working Group (RWG) to recommend guidelines for a common and machine-readable method for sharing information about organizations and their data holdings.

In 2017, the EarthCube Architecture Refinement Workshop (ARW) was held at UCAR and its results identified three main areas for alignment with F.A.I.R. principles including resource registration, discovery, and access. The ARW also generated input for technology inventories, interoperability specifications, and a project management plan for moving forward with the recommended guidelines from the CDF RWG.

## Motivation:

In order to assist geoscience researchers and data providers with resource registration, discovery, and access, we have developed and deployed a highly intuitive, responsive web app for generating and validating JSON-LD formatted metadata. Once published on your website, this information is harvested and indexed by GeoCODES and the Google Dataset Search Tool. By entering information in a simple step-by-step process, your organization and datasets can be discovered quickly. Since the web app dynamically creates itself from well-defined schemas, it can be extended easily to other resource types.

MAKE DATA FAIR - FINDABLE ACCESSIBLE INTEROPERABLE  
REUSABLE





## THE PROCESS

<https://www.earthcube.org/webapps/geocodes/registration/>  
(<https://www.earthcube.org/webapps/geocodes/registration/>)

The screenshot shows the 'EarthCube GeoCODES Registered Organization' form. It includes a navigation bar with links: Home, Load JSON, Main, Contact, Info, Parent Organizations, Associations, Funding Agencies, Publishing, and Generate JSON-LD. The form fields are as follows:

- Legal Name\***: Enter the name that is legally registered to your organization. Text: Biological and Chemical Oceanography Data Management Office.
- Common Name\***: Enter the name commonly used to refer to your organization. Text: BCO-DMO.
- URL or DOI URL\***: Enter the organization's URL or DOI URL. Text: <https://www.bco-dmo.org>.
- Description\***: Describe your organization, e.g., its mission, goals, etc. Text: The BCO-DMO resource catalog offers free and open access to publicly funded research products whose field of study are biological and chemical oceanography.
- Field(s) of Study**: Enter your organization's field(s) of study. Two entries are shown: Biological Oceanography and Chemical Oceanography, each with up/down arrows and a minus sign.

Entering your organization's and dataset's metadata is quick and easy using GeoCODES' multi-step, crossplatform, cross-device web app. The app guides you through several categories of information and validates your entries with each step.



Several standard categories of information for datasets include DOIs, authors, keywords, citations, spatial coverage, and measurements. For organizations, these categories include website URLs, contact info, publishing policies, funding agencies, and fields of study.

Once you have entered your metadata for your organization or dataset, click “Validate” to perform a final validation using GeoCODES’ cloud-based validation web service and click “Save” to save your resulting JSON-LD into a file on your local machine or simply copy and paste the result.

GeoCODES

Home Load JSON Main Contact Info Parent Organizations Associations Funding Agencies Publishing Generate JSON-LD

Loaded From: <https://www.earthcube.org/webapps/geocodes/registration/examples/bco-dmo-org.json>

File:

URL:

```
{
  "@context": {
    "bco-dmo": "http://schema.org/",
    "geo": "https://geocodes.org/voc/",
    "dataset": "http://purl.org/net/dataset/"
  },
  "@type": [
    "Service",
    "Organization",
    "WebSite",
    "WebPage"
  ],
  "id": "http://bco-dmo.org/id/affiliation/181",
  "additionalType": "https://geocodes.org/voc/ResearchRepositoryService",
  "legalName": "Biological and Chemical Oceanography Data Management Office",
  "name": "BCO-DMO",
  "url": "https://www.bco-dmo.org",
  "category": [
    "Biological Oceanography",
    "Chemical Oceanography"
  ],
  "description": "The BCO-DMO resource catalog offers free and open access to publicly funded research products whose field of study are biological and chemical oceanography",
  "website": [
    "http://www.bco-dmo.org/repository/74000000012",
    "https://twitter.com/BCODMO",
    "https://www.linkedin.com/company/8378426/"
  ],
}
```

You can also edit existing JSON-LD for your organization and datasets by uploading a local JSON file or entering the URL of an HTML website landing page or a JSON file. GeoCODES will automatically parse your metadata and populate the correct fields. Note that only HTTPS URLs are currently accepted.

# SCHEMA.ORG

schema.org

Custom Search

Q

HomeSchemasDocumentation

## Dataset

Thing > CreativeWork > Dataset

A body of structured information describing some topic(s) of interest.

[more...]

Property	Expected Type	Description
Properties from Dataset		
distribution	DataDownload	A downloadable form of this dataset, at a specific location, in a specific format.
includedInDataCatalog	DataCatalog	A data catalog which contains this dataset. Supersedes catalog, includedDataCatalog. Inverse property: dataset.
issn	Text	The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication.




```




{
  "@type": "Dataset",
  "isAccessibleForFree": true,
  "@context": {
    "@vocab": "http://schema.org/",
    "re3data": "http://example.org/re3data/0.1/",
    "earthcollab": "https://library.ucar.edu/earthcollab/schema#",
    "geolink": "http://schema.geolink.org/1.0/base/main#",
    "geolink-vocab": "http://schema.geolink.org/1.0/voc/local#",
    "vivo": "http://vivoweb.org/ontology/core#",
    "dcat": "http://www.w3.org/ns/dcat#",
    "dbpedia": "http://dbpedia.org/resource/",
    "geo-upper": "http://www.geoscienceontology.org/geo-upper#"
  },
  "@id": "https://www.bco-dmo.org/dataset/685783",
  "name": "BCO-DMO",
  "identifier": [
    "http://lod.bco-dmo.org/id/dataset/685783",
    {
      "@type": "PropertyValue",
      "additionalType": [
        "http://schema.geolink.org/1.0/base/main#Identifier",
        "http://purl.org/spar/datacite/Identifier"
      ],
      "@id": "https://doi.org/10.1575/1912/bco-dmo.685952",
      "propertyID": "http://purl.org/spar/datacite/doi",
      "value": "10.1575/1912/bco-dmo.685952",
      "url": "https://doi.org/10.1575/1912/bco-dmo.685952"
    }
  ],
  "url": "https://www.bco-dmo.org/dataset/685783",
  "additionalType": [
    "http://schema.geolink.org/1.0/base/main#Dataset",
    "http://vivoweb.org/ontology/core#Dataset"
  ],
  "alternateName": "Carbon flux",
  "description": "MODIFIED",
  "datePublished": "2017-03-27",
  "keywords": "benthic pelagic coupling, Dissolved Organic Carbon, particulate",
  "creator": [
    {
      "@type": "Role",

```



Google Dataset Search

BCO-DMO



Updated Date

Download Format

Usage Rights

Free

100+ results found

N

Biological, chemical, physical, biogeochemical, ecological,...

data.nodc.noaa.gov  
catalog.data.gov  
+1 more

Updated Dec 25, 2013

B

North American marine species distribution...

www.bco-dmo.org

Published Mar 21, 2019

B

Nutrients, targeted proteomics, and pigments...

www.bco-dmo.org

Published May 24, 2018

B

Trace elements in suspended particles from GO-Flo bottles

Biological, chemical, physical, biogeochemical, ecological, environmental and other data collected from around the world during historical and contemporary periods of biological and chemical oceanographic exploration and research managed and submitted by the Biological and Chemical Oceanography Data Management Office (BCO-DMO)

gov.noaa.nodc:BCO-DMO

Explore at data.nodc.noaa.gov

Explore at catalog.data.gov

Explore at AmeriGEOSS Community Platfo...

Dataset created Dec 16, 2010

Dataset published Dec 16, 2010

Dataset updated Dec 25, 2013

## EXAMPLE

```
{
  "@context": {
    "@vocab": "http://schema.org/",
    "gdx": "https://geodex.org/voc/",
    "datacite": "http://purl.org/spar/datacite/",
    "earthcollab": "https://library.ucar.edu/earthcollab/schema#",
    "geolink": "http://schema.geolink.org/1.0/base/main#",
    "geolink-vocab": "http://schema.geolink.org/1.0/voc/local#",
    "vivo": "http://vivoweb.org/ontology/core#"
  },
  "@id": "https://www.bco-dmo.org/dataset/685783",
  "name": "BCO-DMO",
  "isAccessibleForFree": true,
  "identifier": [
    "http://lod.bco-dmo.org/id/dataset/685783",
    {
      "@type": "PropertyValue",
      "additionalType": [
        "http://schema.geolink.org/1.0/base/main#Identifier",
        "http://purl.org/spar/datacite/Identifier"
      ],
      "@id": "https://doi.org/10.1575/1912/bco-dmo.685952",
      "propertyID": "http://purl.org/spar/datacite/doi",
      "value": "10.1575/1912/bco-dmo.685952",
```

```
    "url": "https://doi.org/10.1575/1912/bco-dmo.685952"
  }
],
"url": "https://www.bco-dmo.org/dataset/685783",
"@type": "Dataset",
"additionalType": [
  "http://schema.geolink.org/1.0/base/main#Dataset",
  "http://vivoweb.org/ontology/core#Dataset"
],
"alternateName": "Carbon flux",
"description": "MODIFIED",
"datePublished": "2017-03-27",
"keywords": "benthic pelagic coupling, Dissolved Organic Carbon, particulate organic carbon, Porifera, benthic suspension feeding, biota, oceans",
"creator": [
  {
    "@type": "Role",
    "@id": "http://www.example.com/MODIFIED",
    "additionalType": "http://schema.geolink.org/1.0/base/main#Participant",
    "roleName": "Principal Investigator",
    "creator": {
      "@type": "Person",
      "additionalType": "http://schema.geolink.org/1.0/base/main#Person",
      "@id": "https://www.bco-dmo.org/person/676145",
      "name": "Christopher Finelli",
      "url": "https://www.bco-dmo.org/person/676145",
      "identifier": {
```

```
"@type": "PropertyValue",

"additionalType": [

  "http://schema.geolink.org/1.0/base/main#Identifier",

  "http://purl.org/spar/datacite/Identifier"

],

"propertyID": "http://purl.org/spar/datacite/orcid",

"value": "0000-0002-4034-5201",

"url": "https://orcid.org/0000-0002-4034-5201"

}

},

"url": "http://www.example.com/MODIFIED"

},

{

  "@type": "Role",

  "@id": "http://www.example.com/MODIFIED",

  "additionalType": "http://schema.geolink.org/1.0/base/main#Participant",

  "roleName": "Co-Principal Investigator",

  "creator": {

    "@type": "Person",

    "additionalType": "http://schema.geolink.org/1.0/base/main#Person",

    "@id": "https://www.bco-dmo.org/person/685690",

    "name": "Steven McMurray",

    "url": "https://www.bco-dmo.org/person/685690",

    "identifier": {

      "@type": "PropertyValue",

      "additionalType": [
```



```
    "http://schema.geolink.org/1.0/base/main#Identifier",  
    "http://purl.org/spar/datacite/Identifier"  
  ],  
  "propertyID": "http://purl.org/spar/datacite/orcid",  
  "value": "0000-0002-1187-5916",  
  "url": "https://orcid.org/0000-0002-1187-5916"  
}  
, ...etc...
```

Sorry but time is up!

## CV

Sidney B Hellman<sup>1</sup>, Stefan Lisowski<sup>2</sup>, Eric Lingerfelt<sup>3</sup>, Maura Allen<sup>4</sup> and Alexander McNurlan<sup>4</sup>, (1)Instrumental Software Tech Inc, Saratoga Springs, NY, United States, (2)Instrumental Software Technologies, Inc. (ISTI), New Paltz, NY, United States, (3)University Corporation for Atmospheric Research, Boulder, CO, United States, (4)Instrumental Software Technologies, Inc., Saratoga Springs, NY, United States

## ABSTRACT

Using web standards including Schema.org and JSON-LD, the GeoCODES project extends Schema.org with Project 418's geoscience specific vocabulary. By embedding properly formatted and populated JSON-LD files in web sites serving geolocated datasets, search engines such as Google and Bing are able to parse and index these data sets and then to provide information concerning these datasets via standard web search tools. Due to the difficult nature of properly formatting and populating these JSON-LD structures, the GeoCODES User Interface was created to guide data providers through the process of describing the data and validating the descriptions against standard vocabularies. The result is user friendly and easily extensible web based, mobile device ready tool for automatically generating JSON-LD metadata for organizations and datasets. This ultimately allows the original data to be found and used by both scientists and the public.

## SWITCH TEMPLATE

