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\(^1\), Stefan Lisowski\(^2\), Eric Lingerfelt\(^3\), Maura Allen\(^4\), and Alexander McNurlan\(^4\)

\(^1\)Instrumental Software Tech Inc
\(^2\)Instrumental Software Technologies, Inc. (ISTI)
\(^3\)University Corporation for Atmospheric Research
\(^4\)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

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 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-bystep 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/)

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.
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

## Dataset

**Thing > CreativeWork > Dataset**

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

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>distribution</code></td>
<td><code>DataDownload</code></td>
<td>A downloadable form of this dataset, at a specific location, in a specific format.</td>
</tr>
<tr>
<td><code>includedInDataCatalog</code></td>
<td><code>DataCatalog</code></td>
<td>A data catalog which contains this dataset. Supersedes catalog, includedDataCatalog.</td>
</tr>
<tr>
<td><code>issn</code></td>
<td><code>Text</code></td>
<td>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.</td>
</tr>
</tbody>
</table>
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",
        }
    ]
}
"@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

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.
<table>
<thead>
<tr>
<th>SWITCH TEMPLATE</th>
</tr>
</thead>
</table>

```markdown

```