

---

# **The Ocean InfoHub Project and the development of the ODIS-architecture**

**IODEPO**

**Jul 19, 2021**



# CONTENTS

<b>I</b>	<b>Introduction</b>	<b>5</b>
<b>1</b>	<b>Structured Data on the Web</b>	<b>7</b>
1.1	About . . . . .	7
1.2	Web architecture approach . . . . .	7
1.3	Terminology . . . . .	8
1.4	Intellectual Merit . . . . .	8
1.5	Broader Impacts . . . . .	8
<b>2</b>	<b>Personas</b>	<b>9</b>
2.1	About . . . . .	9
2.2	Persona: Publisher . . . . .	10
2.3	Persona: Aggregator . . . . .	10
2.4	Persona: User . . . . .	10
<b>3</b>	<b>Publisher</b>	<b>11</b>
3.1	About . . . . .	11
3.2	Basics . . . . .	13
3.3	Full Workflow . . . . .	15
3.4	Existing support in software . . . . .	16
<b>4</b>	<b>JSON-LD Foundation</b>	<b>17</b>
4.1	Introduction . . . . .	17
4.2	The Context . . . . .	17
4.3	Graph . . . . .	18
<b>II</b>	<b>Profiles</b>	<b>21</b>
<b>5</b>	<b>Thematic Patterns</b>	<b>23</b>
5.1	Introduction . . . . .	23
5.2	Thematic Profiles . . . . .	23
5.3	Example Connections . . . . .	24
<b>6</b>	<b>Experts and Institutions</b>	<b>27</b>
6.1	About . . . . .	27
6.2	Example: Person Graph . . . . .	27
6.3	Example: Institution Graph . . . . .	29
6.4	References . . . . .	30
<b>7</b>	<b>Documents</b>	<b>31</b>
7.1	About . . . . .	31

7.2	Creative works (documents)	31
<b>8</b>	<b>Spatial Maps</b>	<b>35</b>
8.1	About	35
<b>9</b>	<b>Projects</b>	<b>39</b>
9.1	About	39
9.2	Research Project	39
<b>10</b>	<b>Training</b>	<b>43</b>
10.1	About	43
10.2	Basic Course style	43
10.3	Simple Course	44
10.4	References	45
<b>11</b>	<b>Vessels</b>	<b>47</b>
11.1	About	47
11.2	References	48
<b>12</b>	<b>Spatial Geometry</b>	<b>49</b>
12.1	About	49
12.2	Simple GeoSPARQL WKT	49
12.3	Classic Schema.org	50
12.4	Option review, SOS Issue 105	50
12.5	References	51
<b>13</b>	<b>Services</b>	<b>53</b>
13.1	About	53
13.2	References	54
<b>14</b>	<b>Defined Terms</b>	<b>55</b>
14.1	About	55
14.2	Organization in List	56
14.3	References	57
<b>15</b>	<b>Languages</b>	<b>59</b>
15.1	About	59
<b>16</b>	<b>Linking to documents and resources</b>	<b>61</b>
16.1	Organization link options	61
16.2	SDG Linkage	63
16.3	Creative work link options	64
16.4	Refs	65
<b>17</b>	<b>Identifier</b>	<b>67</b>
17.1	About	67
<b>III</b>	<b>Aggregation</b>	<b>69</b>
<b>18</b>	<b>Aggregator</b>	<b>71</b>
18.1	Intoduction	71
18.2	ODIS Catalog as Index Source	73
<b>19</b>	<b>Indexing with Gleaner</b>	<b>75</b>
19.1	Gleaner (app)	75

<b>20 Indexing Services</b>	<b>83</b>
<b>21 Data Services</b>	<b>87</b>
21.1 Gleaner Data Services (DS) . . . . .	87
<b>22 Interfaces</b>	<b>95</b>
22.1 About . . . . .	95
22.2 Gleaner Web UI (WUI) . . . . .	95
<b>23 Graph First Approach</b>	<b>97</b>
23.1 About . . . . .	97
23.2 Graph Only . . . . .	97
23.3 Item Catalogue Page . . . . .	98
23.4 Publishing and referencing . . . . .	98
23.5 Testing . . . . .	98
<b>24 Prov</b>	<b>101</b>
24.1 About . . . . .	101
24.2 Gleaner Prov . . . . .	101
24.3 Nano Prov . . . . .	103
24.4 Refs . . . . .	105
<b>25 Alternatives</b>	<b>107</b>
25.1 Options . . . . .	107
 <b>IV Tooling</b>	 <b>109</b>
<b>26 Tooling</b>	<b>111</b>
26.1 About . . . . .	111
26.2 OpenRefine . . . . .	111
26.3 OIH Notebooks . . . . .	117
 <b>V Validation</b>	 <b>177</b>
<b>27 Validation</b>	<b>179</b>
27.1 About . . . . .	179
27.2 Implementation . . . . .	179
 <b>VI Interfaces</b>	 <b>181</b>
<b>28 Users</b>	<b>183</b>
28.1 About . . . . .	183
<b>29 Querying SPARQL</b>	<b>185</b>
29.1 About . . . . .	185
 <b>VII Appendix</b>	 <b>187</b>
<b>30 Appendix</b>	<b>189</b>
30.1 About . . . . .	189
30.2 Known Issues . . . . .	189
30.3 References . . . . .	190

30.4 Registries . . . . . 193

30.5 Controlled Vocabularies . . . . . 194

## Introduction

Organizations are increasingly exposing data and resources on the Web. A popular approach to this is using web architecture to expose structured data on the web using the [schema.org](http://schema.org) vocabulary. Doing this makes resources discoverable by a range of organizations leveraging this architecture to build indexes. These include major commercial indexes, large domain focused groups and community focused services.

The Ocean Data and Information System (ODIS) will provide a [schema.org](http://schema.org) based interoperability layer and supporting technology to allow existing and emerging ocean data and information systems, from any stakeholder, to interoperate with one another. This will enable and accelerate more effective development and dissemination of digital technology and sharing of ocean data, information, and knowledge. As such, ODIS will not be a new portal or centralised system, but will provide a collaborative solution to interlink distributed systems for common goals. Together with global project partners and partners in the three regions, a process of co-design will enable a number of global and regional nodes to test the proof of concept for the ODIS.

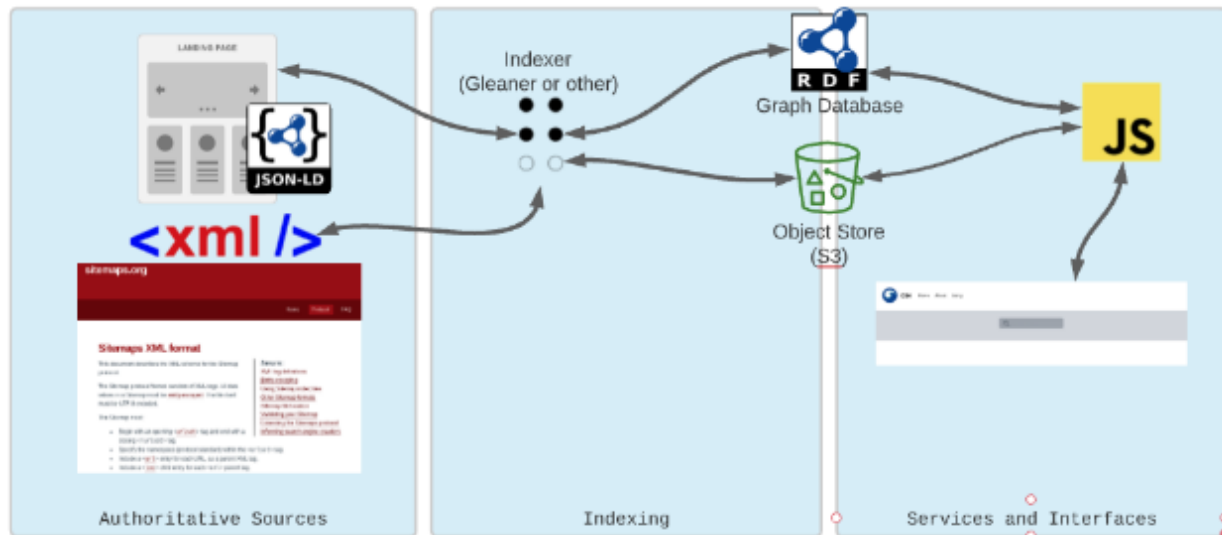
The ODIS-architecture development is being supported by the Ocean InfoHub Project, and it has been tested initially on IOC and partner databases. However, the system and standards are open for any institution or initiative that is interested in accessing the global data ecosystem to adopt and implement.

## Guidance for the implementation of the ODIS-architecture

OIH is providing guidance on the various stages of such an architecture including authoring, publishing, indexing and interfaces.

The basics of this approach can be described as:

- Providers publish HTML pages for a resource. This may be a publication, course description, research instrument or other. The core themes for OIH are described in the Authoring section below.
- A HTML page then has a small JSON based snippet added to the HTML. This is described in the Including JSON-LD in your resource page in the Publishing resource below.
- If you wish a resource to be included in the OIH index, then you need to include it in a sitemap file. This is a small XML document that lists links to the resources you wish to be part of the index. This approach is shown in the sitemap.xml section of the Publishing resource.
- Once the above is done the publishing phase is over. At this point, OIH or other groups can now access and index your resources. OIH is using some existing software to index and generate the graph and expose a simple reference interface to them. This software is open and available and others are free to implement the approach with other software. Links to other software are at the repository.
- The OIH index/graph and a simple interface is current at a development site and in a later phase of OIH a production interface will be developed.



Additionally, software to aid in validating and checking the resources is under development and will be available at the repository. This will aid providers in expressing the information needed to address interfaces and services of interest to the community.

The result is a sustainable architecture to address discovery and access to various resources published by the community and a shared graph of these resources. That shared graph can be used by all members to link and discover across groups.

### Key links to the OIH GitHub repository

Interested groups can review material addressing these stages at the OIH GitHub repository. Links and descriptions of these stages are described below.

#### *Authoring Thematic Patterns*

The ODIS OIH is working across five major thematic areas; Experts and Institutions, Documents, Projects, Training, Vessels. Examples of these thematic concepts are being hosted and developed with input from the community. Additionally, methods for validation and simple tooling for authoring and testing are hosted at this repository. Alongside these five thematic topics guidance on connecting services and spatial context on resources.

#### *Publishing*

Guidance on implementing the web architecture approach is also available. This includes approaches on leveraging robots.txt and sitemaps.xml file for expressing hosted resources to the net.

#### *Indexing*

The architecture approach is open and standards based. As such, many organizations will be able to leverage the authoring and publishing approaches above to index a providers resources. OIH will be providing reference implementations of



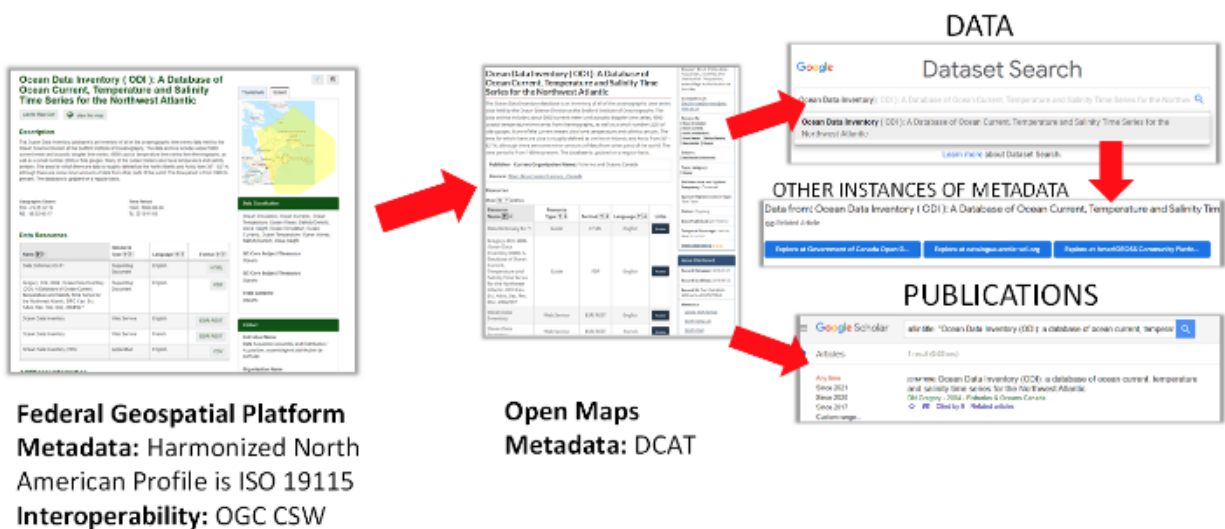
software that can generate the index.

## Interfaces and Services

During the development of the OIH a basic reference implementation for an interface has been generated. This is a development site meant to test and exercise the above elements. It serves to demonstrate how others could also implement this approach and how future interfaces could be developed.

An example of the value of implementing a lightweight can be seen with the Government of Canada:

- The Federal Geospatial Platform is a intra-governmental data catalogue implementing the Harmonized North American Profile of ISO 19115 (HNAP), with content exposed externally via OGC CSW (Catalogue Services Web).
- This content is harvested by the public facing Open Maps platform, which includes a catalogue component that is fed in part by the Federal Geospatial Platform. DCAT-based metadata is derived from the original ISO 19115 based metadata. As this markup is recognized by web crawlers such as those hosted by Google, content is harvested and is subsequently visible through Google Dataset Search. Furthermore, the cited publication for the data is also link via a complementary link to Google Scholar.



## Info-graphic

The following is a simple overview info-graphic of the Ocean Info Hub activity flow.

## CONTENTS



## **Part I**

# **Introduction**



## STRUCTURED DATA ON THE WEB

### 1.1 About

Structured data on the web is a way to provide semantics and linked data in an approachable manner. This approach expresses concepts in JSON-LD which is a JavaScript notation popular among developers which easily expresses concepts (terms) and links to related resources (things). This structured data on the web approach has been popularized by the large commercial search providers like Google, Bing, Yandex and others via [schema.org](http://schema.org). As described at [schema.org](http://schema.org): “[Schema.org](http://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.”

The popularity of leveraging the [schema.org](http://schema.org) approach in the earth sciences can be attributed to both this ease of developer adoption and also to its foundational use of web architecture. A web architecture foundation aids adoption by the operations side as well as the developer side. It also takes advantage of the scale and resilience of the web.

The broad nature of [schema.org](http://schema.org) even scopes to the concepts of Datasets. It is the existence of [schema.org/Dataset](http://schema.org/Dataset) that was a focus of several EarthCube projects (Project 418, Project 419 and the Resource Registry) from which spun up the ESIP Science on Schema work.

Additionally, Google leveraged [schema.org/Dataset](http://schema.org/Dataset) to develop and populate the Google Data Set Search and provides guidance to developers to facilitate this.

### 1.2 Web architecture approach

OIH is focused on leveraging the web architecture as the foundation for this approach. There are several key reasons for this vs approaches like OAI-PMH or others.

A key point is that in the processes of establishing a web presence, a standard step for groups, they have already begun to build the infrastructure needed for structured data on the web. Setting up special servers or establishing and maintaining special APIs to support harvesting is not required.

Also, a large collection of tooling already exists around JSON that is directly usable in JSON-LD. That scale extends to the use of [schema.org](http://schema.org) patterns which have become common in the commercial web. Allowing us to bring those same patterns and the tooling to the science community.

Additionally, this approach keeps the metadata and its representation a product of the data providers. The actor in the life cycle most aware of needed edits, new records or other events. That same record then serves multiple consumers able to generate various value add products. This benefits the provider by facilitating multiple and varied discovery vectors for their holdings.

Another key factor is the web native and semantic nature of this representation of metadata. Traditional metadata, such as ISO, by itself does not express a web referenceable instance of concepts. In doing this, structured data on the web allow connections to be made and discovered by people and machines across many holdings. This aids in both serendipitous discovery and can also be leveraged to aid discovery via semantic relations.

### 1.3 Terminology

A CSV file is a text file containing spreadsheet information following a data model that is encoded using a convention of rows and commas defining columns.

A JSON-LD file is a text file containing graph information following the RDF data model that is encoded using a convention based on JSON syntax.

JSON-LD is a way to serialize RDF that uses JSON notation. It is really no different than RDF/XML, turtle, n-triples, etc. There are several ways to represent the RDF data model in text files (and some emerging binary ones like CBOR and parquet patterns).

[Schema.org](https://schema.org) is a vocabulary for describing things similar to DCAT, FOAF, Dublin Core. It does this by using RDF as the underlying data model to represent this “ontology”.

The confusion comes from the collision of outcomes. JSON-LD came about, partly, to allow the use of the RDF data model by a broader audience. This is done by leveraging a more popular notation for the data model, JSON, in the form of JSON-LD. [Schema.org](https://schema.org) also wanted to advance the use of structured [meta]data by making it easier to use and connecting structured data to web pages. At the start, there were three approaches; RDFa, microformats and JSON-LD, to putting [schema.org](https://schema.org) in web pages. However, the JSON-LD approach to incorporating this structured data has grown in popularity far beyond the others. As the popularity of both JSON-LD and [schema.org](https://schema.org) grew, they often got conflated with each other.

The term “structured data on the web” is perhaps a more neutral way to discuss the use of vocabularies encoding in JSON-LD used in web pages. However, the phrase “[schema.org](https://schema.org)” is starting to become the term for “structured data on the web using JSON-LD as a serialization”. Even in cases where you combine other vocabularies such as DCAT with JSON-LD with no [schema.org](https://schema.org) involved, it seems the way to convey this is to say: “We will use the [schema.org](https://schema.org) ‘pattern’ with DCAT”.

It is arguably not the best or most accurate communications strategy. It can conflate data models, serialization and vocabularies. However, it is concise and ubiquitous and not likely to change.

### 1.4 Intellectual Merit

OIH leverages structured data on the web patterns in the form of [Schema.org](https://schema.org) and JSON-LD encoding. This means that much of what is done to address OIH implementation by providers also is available both to existing commercial indexing approaches as well as emerging community practices

Additionally, both the publishing and indexing approaches are based on several web architecture patterns. Meaning that existing organization skills are leveraged and staff experience is enhanced. This helps to address both the sustainability of the OIH connection and the efficiency of organizational operation.

### 1.5 Broader Impacts

By leveraging existing technology and approaches a larger community is enabled to engage and make more samples discoverable and usable.

The nature of structured data on the web also provides the ability to apply semantic context to samples. This means richer discovery and information about samples, the past uses and potential future uses is more readily available.

Simplified architecture also means easier development of tools and interfaces to present the data. Allowing the presentation of samples and their information in a manner aligned with a given community’s needs. A simplified architecture aids sustainability from both a technical and financial perspective.

## PERSONAS

### 2.1 About

The OIH system can be viewed as involving three personas. These are briefly presented here.



Publisher

A key persona whose activities are covered in detail in *Publishing patterns for OIH*



Aggregator

Leverages web architecture to retrieve structured data on the web and generate usable indexes.



User

The end user of the publishing and aggregation activities. May leverage the web for discovery or tools such as Jupyter for analytics and visualization.

## 2.2 Persona: Publisher

In OIH the Publisher is the one authoring the JSON-LD documents and publishing them to the web. Details on this persona can be found in the [Publisher](#) section. Additionally, this persona would be leveraging this encoding described in the [JSON-LD Foundation](#) section and the profiles described in the [Thematic Patterns](#).

## 2.3 Persona: Aggregator

In OIH the Aggregator is a person or organization who is indexing resources on the web using the structured data on the web patterns described in this documentation. Details on the approach used by OIH and potential alternatives can be found in the [Aggregator](#) section.

## 2.4 Persona: User

The user is the individual or community who wished to leverage the indexes generated as a result of the publishing and aggregation activities. The user may be using the developed knowledge graph or some web interface built on top of the knowledge graph or other index.

User tools may be web sites or scientific notebooks. Some examples of these user experiences are described in the [User](#) section.



## 3.1 About

This page describes the publishing process for structured data on the web approach OIH will use.

Note many software packages you are using might already implement this approach. See the section: *Existing support in software* at the bottom of this document.

### See also:

We also recommend reviewing the document: [Schema.org for Research Data Managers: A Primer](#)

### 3.1.1 Architecture Implementation

The Ocean Info Hub (OIH) will leverage structured data on the web and web architecture patterns to expose metadata about resources of interest to the community. The primary tasks include:

- Authoring JSON-LD documents (<https://json-ld.org/>) aligned with ODIS OIH guidance to express the structured metadata for a resource. This step will require experience with using the existing metadata resources within an organization. So any necessary skills needed to access or query existing facility data systems will be needed to assemble the information to populate the JSON-LD data graph. The JSON-LD documents need to be generated using the tools/languages at the previous reference or through other means.
- Within the system architecture of the site, a JSON-LD document needs to be placed into the HTML DOM as a SCRIPT tag within the HEAD tag of each published resource. The SCRIPT tag pattern is:

```
<script type="application/ld+json">JSON_LD content</script>
```

- Additionally these resources that are marked up with these tags and JSON-LD documents should be expressed in an XML sitemap file. This should follow the guidance at <https://www.sitemaps.org/>. It should also include a lastmod node element as described at <https://www.sitemaps.org/protocol.html> which should indicate the date the resource metadata was last updated and published to the web.
- The process of aligning the JSON-LD is iterative at this stage as the OIH profile is evolved. To aid this we can leverage existing validation tools including JSONSchema, W3C SPARQL and more to communicate structure changes. These tools exist and need only be implemented using knowledge of command line environments. The results will then indicate revisions needed in the JSON-LD. OIH will provide the necessary templates for the tools to use against the authored JSON-LD documents.

Information on the sources, standards and vocabularies to be used can be found at: <https://github.com/iodepo/odis-arch/tree/schema-dev/docs>

### 3.1.2 Including JSON-LD in your resource page

To provide detailed and semantically described details on a resource, OIH uses a [JSON-LD](#) snippet or *data graph*. This small document provides details on the resource. It can also express any explicate connections to other resources an author may wish to express. The semantic nature of the document also means that connections may later be discovered through graph queries.

Pages will need a JSON-LD data graph placed in it via a typed script tag/element in the document head element like the following.

```
<script type="application/ld+json"></script>
```

An example data graph can be seen below. However, check the various thematic sections for more examples for a given thematic area.

```
{
  "@context": {
    "@vocab": "https://schema.org/",
    "endDate": {
      "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
    },
    "startDate": {
      "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
    }
  },
  "@id": "https://foo.org/url/to/metadata/representation",
  "@type": "Course",
  "description": "In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox...",
  "hasCourseInstance": {
    "@type": "CourseInstance",
    "courseMode": [
      "MOOC",
      "online"
    ],
    "endDate": "2019-03-21",
    "startDate": "2019-02-15"
  }
}
```

This example is from the [training and courses](#) thematic section. To view all the types being developed reference the [Thematic section](#).

These JSON-LD documents leverage [schema.org](#) as the primary vocabulary. The examples in the thematic section provide examples for the various type.

### JSON-LD Tools and References

A key resource for JSON-LD can be found at [JSON-LD](#). There is also an interactive *playground* hosted there. The [JSON-LD Playground](#) is useful when testing or exploring approaches for JSON-LD data graphs. It will catch basic errors of syntax and use. Note, it will not catch semantic issues such as using properties on types that are out of range. Tools like the [Structured Data Testing Tool](#) are better at that. Also the documents and validation material created here OIH will also allow for that sort of testing and feedback.

Providers may also wish to provide content negotiation for type `application/ld+json` for these resources. Some indexers, like Gleaner, will attempt to negotiate for the specific serialization and this will likely lighten the load on the servers going forward.

## Validation With SHACL or ShEx

To help facilitate the interconnection of resource, some application focused validation will be developed. Note, this validation does not limit what can be in the graphs. Rather, it simply provides insight on to how well a given graph can be leveraged for a specific application. For this project, the application will be the OIH search portal.

Some initial development work for this can be found in the *[validation directory](#)*

## Validation Tools and References

- [SHACL playground](#)
- [Schemarama](#)
- [Schimatos.org](#)
  - [demo](#)
- [Comparing ShEx and SHACL](#)

## Validation Leveraging JSON Schema

We have been exploring the potential to use JSON Schema combined with various on-line JSON editors (JSON Schema driven) to provide a potential approach to a more visual editing workflow. The workflow presented here is very ad hoc but exposes a potential route a group might take to develop a usable tool. Such a tool might, for example, leverage the Electron app dev environment to evolve this approach in a more dedicated tool/manner.

Use a JSON-LD document ([Example](#)) one could load this into something like the [JSONschema.net](#) tool.

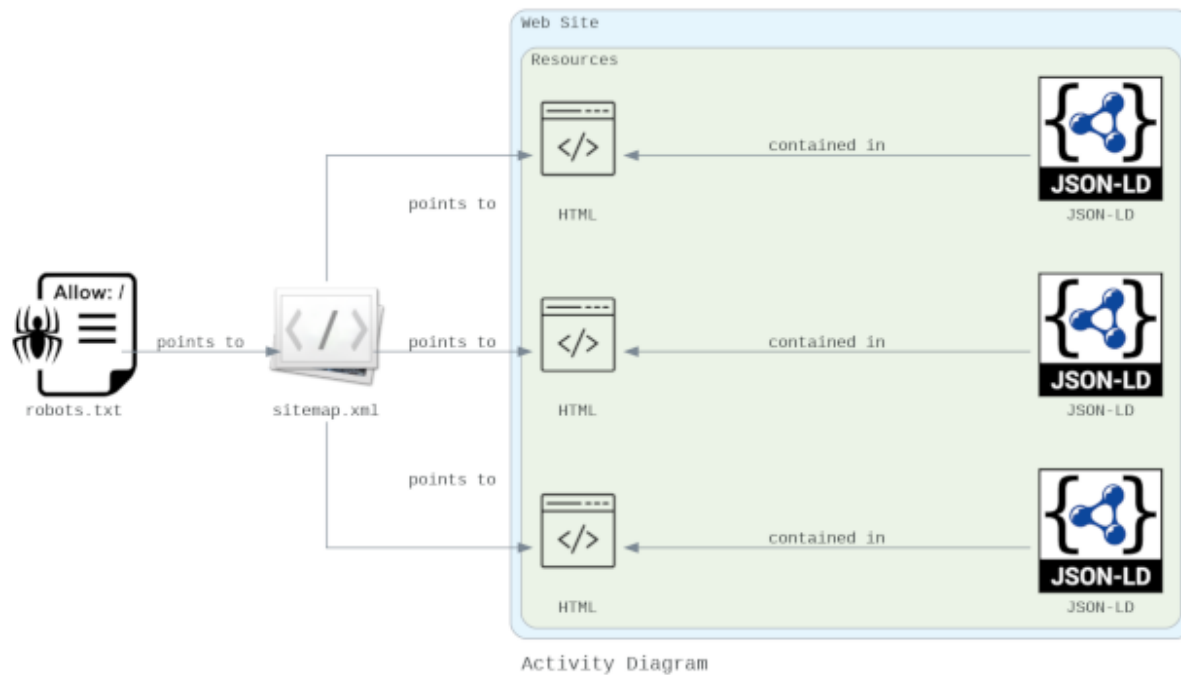
The results of the above can then be loaded into the online JSON-Editor at <https://json-editor.github.io/json-editor/>. (Ref: <https://github.com/json-editor/json-editor>)

The results of this then can be load into <https://json-ld.org/playground/> to validate that we have well formed JSON-LD.

Though this workflow is rather crude and manual it exposes a route to a defined workflow based around established schema that leverages other tools and software libraries to generate a workable tool.

## 3.2 Basics

The basic activity can be seen in the following diagram:



### 3.2.1 Elements in detail

#### robots.txt

OPTIONAL: Providers may decide to generate or modify their robots.txt file to provide guidance to the aggregators. The plan is to use the Gleaner software ([gleaner.io](https://gleaner.io)) as well as some Python based notebooks and a few other approaches in this test.

Gleaner uses an agent string of EarthCube\_DataBot/1.0 and this can be used a robots.txt file to specify alternative sitemaps and guidance. This also allows a provider to provide guidance to Google and other potential indexers both for allow and disallow directives.

```

Sitemap: http://samples.earth/sitemap.xml

User-agent: *
Crawl-delay: 4
Allow: /

User-agent: Googlebot
Disallow: /id

User-agent: EarthCube_DataBot/1.0
Allow: /
Sitemap: https://example.org/sitemap.xml

```

## sitemap.xml

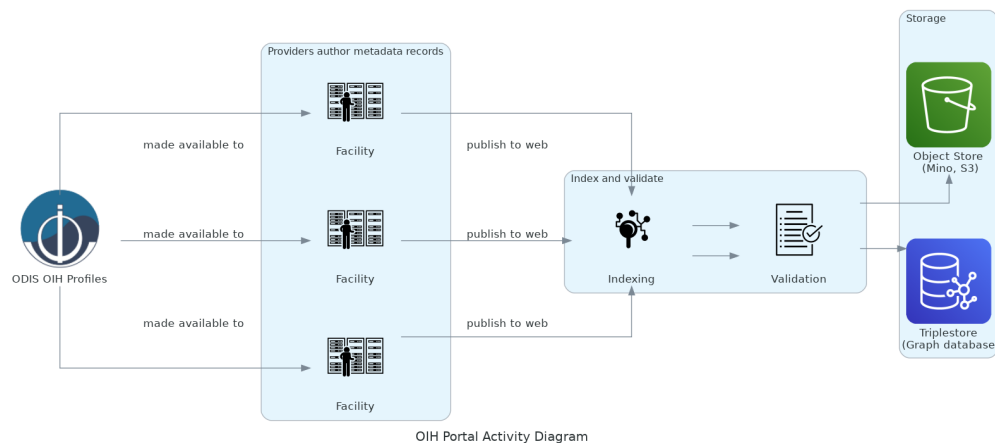
Providers will need to expose a set of resource landing pages using a sitemap.xml file. As noted above, providers can expose a sitemap file to just the target agent to avoid indexing test pages by commercial providers. You may wish to do this during testing or for other reasons. Otherwise, a sitemap.xml file exposed in general from somewhere in your site is perfectly fine.

Information on the sitemap structure can be found at [sitemaps.org](http://sitemaps.org).

It is encouraged to use the sitemap lastmod node to provide guidance to indexers on page updates. Additionally indexers may test ways to evaluate additions and removals from the sitemap URL set to manage new or removed resources.

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/site0.9">
  <sitemap>
    <loc>http://samples.earth/sitemap_websites_sampleseaxml</loc>
    <lastmod>2004-10-01T18:23:17+00:00</lastmod>
  </sitemap>
  <sitemap>
    <loc>http://samples.easitemap_docclouds_igsndatagraphs.xml</loc>
    <lastmod>2005-01-01</lastmod>
  </sitemap>
</sitemapindex>
```

## 3.3 Full Workflow



The architecture defines a workflow for objects seen in the above diagram.

The documents flow from; authoring, publishing and indexing to storage for the objects and the resulting graph. These resources are then ready for use in search and other functions.

Moving left to right we can review the image.

1. Providers are engaged in the process of developing the OIH example documents. These provide a *profile* to follow to represent the semantic metadata. Note, these are not limiters, simply guidance on minimum and recommend elements to address the functional goals of the OIH portal.
2. Providers use these documents to generate the JSON-LD data graphs. These can be either static documents or generated and placed in pages dynamically with Javascript or server side templates. These are the existing web pages for the resources, not enhanced with the semantic metadata snippets in the HTML source.
3. These are published to the web and referenced in the sitemap.xml document that is also made available. At this point this material is available to anyone who may wish to index it and provide discovery for these resources.
4. OIH Portal will then index and validate these resources on a recurring bases to maintain a current index. This index will include both the JSON-LD objects and the graph they form. This graph can be used for search, connections and other value add services for the community. The graph is also directly available to the community for them to use in support of services they may wish to provide.

### 3.4 Existing support in software

Many content management approaches and packages may already have support for the structured data on the web pattern.

A collection of starting points for their support follows.

- [WordPress](#)
- [Drupal](#)
- [CKAN](#)
- [DSpace](#)
- [ERDDAP](#) (native support)
- [OPeNDAP](#) (native support)
- [GeoNode](#)
  - [schema.org](#) issue ref

## JSON-LD FOUNDATION

### 4.1 Introduction

This document provide a very brief introduction to the JSON-LD serialization format. The [JSON-LD](#) website has some detailed material and videos in their [documentation section](#).

The material here is just a brief introduction. For this page we will be using a simplified version of a CreativeWork document. All the types used by OIH are defined by [Schema.org](#) types. In this case it is [CreativeWork](#).

At the [Schema.org](#) site you will find extensive details on what the various types mean and the full range of their properties. For OIH we are defining only a few of these properties as of interest in the [Thematic section](#). You are free to use additional properties to describe your resources. It will not cause any issues, however, the OIH interfaces may not leverage them. However, if you feel others would, or you use them yourself, it's encouraged to add them.

We will use the following simple JSON-LD document to show the various features of the format.

```
1 {  
2   "@context": {  
3     "@vocab": "https://schema.org/"  
4   },  
5   "@type": "CreativeWork",  
6   "@id": "https://example.org/id/XYZ",  
7   "name": "Name or title of the document",  
8   "description": "Description of the creative work to aid in searching",  
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf"  
10 }
```

<graphviz.dot.Digraph at 0x7f7435db9070>

### 4.2 The Context

The context is where the terms used in a document are connected to definitions and identifiers for them. If you wish to dive into the details of the context check out the [W3 JSON-LD 1.1 Recommendations Context section](#).

The context part of this document is highlighted below.

---

**Note:** This section will be the same for all the documents described in OIH documentation with the exception of the spatial patterns.

---

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "CreativeWork",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the creative work to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf"
10 }
```

As noted, for the spatial patterns we add in the OGC context to all us to use terms from that vocabulary. Below we can see the addition of the geosparql context in line 4 and the use of the vocabulary, using the defined geosparql: prefix in lines 9, 11 and 15.

If we wanted to use other vocabularies like DCAT or FOAF, we would add them to the context with a prefix and then the vocabulary namespace. We could then use terms from that vocabulary in our document following the same prefix:term pattern.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/",
4     "geosparql": "http://www.opengis.net/ont/geosparql#"
5   },
6   "@id": "https://example.org/id/XYZ",
7   "@type": "Dataset",
8   "name": "Data set name",
9   "geosparql:hasGeometry": {
10     "@type": "http://www.opengis.net/ont/sf#Point",
11     "geosparql:asWKT": {
12       "@type": "http://www.opengis.net/ont/geosparql#wktLiteral",
13       "@value": "POINT(-76 -18)"
14     },
15     "geosparql:crs": {
16       "@id": "http://www.opengis.net/def/crs/OGC/1.3/CRS84"
17     }
18   }
19 }
20 }
```

### 4.3 Graph

The next section we will discuss is the graph part of the document. This is where the properties and values of our resource are described. First though, let's visit a couple special properties in our document.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "CreativeWork",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the creative work to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf"
10 }
```



### 4.3.1 Node identifiers (@id)

```

1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "CreativeWork",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the creative work to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf"
10 }

```

The first special property is the @id property. This is the identifier for the top level node in the graph and is typically the identifier for the record.

**Note:** It should be noted this is not the ID for the object being described but rather the record itself. If you are describing a dataset with a DOI, for example, the @id is not that DOI. Rather it is the ID, potentially the URL, for the metadata record about that dataset. Your dataset ID would be included in the metadata record using the identifier property.

It's good practice to ensure all your records have an @id property. If there is no value then the resource is identified by what is known as a blank node. Such identifiers do not allow use in a Linked Open Data approach and are generally not recommended.

The @id should be the URL for the metadata record itself. Not the HTML page the record is in. However, these might be the same if you use content negotiation to select between HTML and JSON-LD representations of the record.

### 4.3.2 Type identifiers (@type)

```

1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "CreativeWork",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the creative work to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf"
10 }

```

The next property to focus on is the @type property. This describes the type of record we are describing.

**Note:** In [Schema.org](https://schema.org/) and in most vocabularies, types will be named with a capital letter. Properties on these types will be all lower case. So, CreativeWork, as a type, starts with a capital C. Then, name, as a property on the CreativeWork type, starts with a lower case n.

For OIH these types for the various thematic profiles are defined in the documentation for the types.

### 4.3.3 Other properties

At this point we can look at the other properties for our type.

```
1 {  
2   "@context": {  
3     "@vocab": "https://schema.org/"  
4   },  
5   "@type": "CreativeWork",  
6   "@id": "https://example.org/id/XYZ",  
7   "name": "Name or title of the document",  
8   "description": "Description of the creative work to aid in searching",  
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf"  
10 }
```

As noted, we are using [Schema.org](https://schema.org/) type for OIH. In this case, as mentioned, this is type [CreativeWork](https://schema.org/CreativeWork). So any of the properties seen at the [Schema.org](https://schema.org/) site can be used. The key properties of value to the OIH implementation can then be found, for this type, in the *Documents thematic type*.

# **Part II**

# **Profiles**



## THEMATIC PATTERNS

### 5.1 Introduction

These thematic patterns are managed by OIH and the community to add in the discovery and use of ocean related resources. The patterns are simple examples of [Schema.org](https://schema.org) types, with a focus on the properties and type relations of value to the Ocean InfoHub and the community it engages.

These “profiles” provide both a starting point for new users and a catalysis for discussion and extension with the community.

### 5.2 Thematic Profiles

These profiles represent reference implementation of [schema.org](https://schema.org) Types related to the identified ODIS thematic areas. They provide a set of minimal elements and notes on more detailed elements.

These are not final and will evolve with community input. As this process moves forward we will implement versioning the profiles to provide stable implementations providers can reliably leverage in their workflows.

#### 5.2.1 Core Profiles

Six key categories of interest:

1. *Experts and Institutions*
2. *Documents*
3. *Spatial Maps*
4. *Projects*
5. *Training*
6. *Vessels*

### 5.2.2 Supporting Profiles

In support of these five thematic types above, these cross cutting types and properties were selected for attention. They represent some key patterns people may wish to leverage when describing their resources.

1. *Spatial Geometry*
2. *Services*
3. *Term Lists*
4. *Languages*
5. *Linking to Principles*
6. *Identifier Patterns*

**See also:**

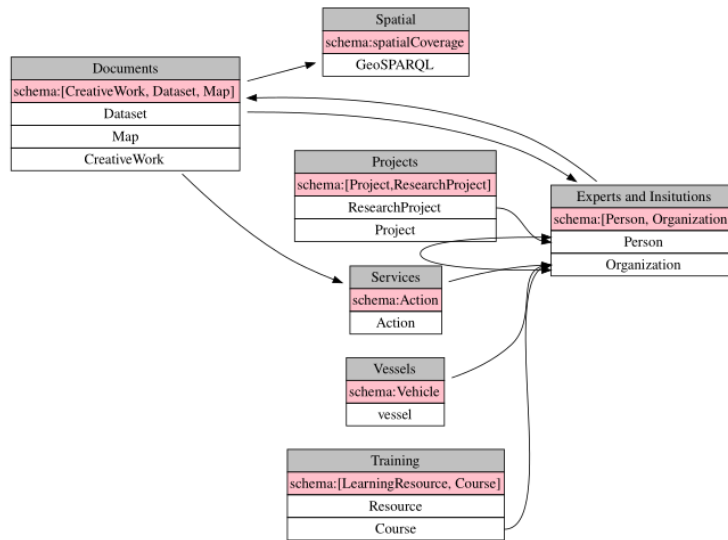
For OIH the focus is on generic documents which can scope reports, data and other resources. In those cases where the resources being described are of type Dataset you may wish to review patterns developed for GeoScience Datasets by the ESIP [Science on Schema](#) community.

**See also:**

For OIH the focus is on generic documents which can scope reports, data and other resources. In those cases where the resources being described are life sciences resources such as datasets, software, and training materials. we recommend following patterns developed by [Bioschemas](#).

## 5.3 Example Connections

The following figure provides an overview of potential connections between the various thematic types in OIH. These are not *all* the relations but simply some to demonstrate the concept as well as give some guidance on the connection of value in query results.







## EXPERTS AND INSTITUTIONS

### 6.1 About

Expert: A person who has a deep understanding of a particular subject area.

Institution: A group of people working together to provide a particular service.

### 6.2 Example: Person Graph

The following graph present a basic record we might present about a person. We can break down some of the key attributes of this record.

OIH is using [schema.org/Person](https://schema.org/Person) for this type. Any of the properties of Person seen there are valid to use in such a record.

At the core though there are a few key items OIH is looking for in a Person record.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@id": "https://example.org/id/x",
6   "@type": "Person",
7   "name": "Jane Doe",
8   "jobTitle": "Professor",
9   "telephone": "(425) 123-4567",
10  "url": "http://www.janedoe.com",
11  "knowsAbout": [
12    {
13      "@type": "Text",
14      "description": "Invasive species in brackish water"
15    },
16    {
17      "@type": "URL",
18      "url": "https://www.wikidata.org/wiki/Q183368"
19    },
20    {
21      "@id": "https://example.org/id/course/x",
22      "@type": "Course",
23      "description": "In this course ...",
24      "url": "URL to the course"
25    }
26  ],
```

(continues on next page)

(continued from previous page)

```

27  "identifier": {
28    "@id": "https://orcid.org/0000-0002-2257-9127",
29    "@type": "PropertyValue",
30    "propertyID": "https://registry.identifiers.org/registry/orcid",
31    "url": "https://orcid.org/0000-0002-2257-9127",
32    "description": "Optional description of this record..."
33  }
34  }

```

<graphviz.dot.Digraph at 0x7fdc74223f70>

## 6.2.1 Details: Authoritative Reference

For each profile there are a few key elements we need to know about. One key element is what the authoritative reference or canonical identifier is for a resource.

```

{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/x",
  "@type": "Person",
  "identifier": {
    "@id": "https://orcid.org/0000-0002-2257-9127",
    "@type": "PropertyValue",
    "description": "Optional description of this record...",
    "propertyID": "https://registry.identifiers.org/registry/orcid",
    "url": "https://orcid.org/0000-0002-2257-9127"
  }
}

```

<graphviz.dot.Digraph at 0x7fdc65dd9a60>

## 6.2.2 Details: Knows About

Knows about provide connections to resources a person is connected with.

```

{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/x",
  "@type": "Person",
  "knowsAbout": [
    {
      "@type": "Text",
      "description": "Invasive species in brackish water"
    },
    {
      "@type": "URL",
      "url": "https://www.wikidata.org/wiki/Q183368"
    }
  ],
}

```

(continues on next page)

(continued from previous page)

```
{
  "@id": "https://example.org/id/course/x",
  "@type": "Course",
  "description": "In this course ...",
  "url": "URL to the course"
}
```

<graphviz.dot.Digraph at 0x7fdc65de1c40>

## 6.3 Example: Institution Graph

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@id": "https://example.org/id/org/x",
6   "@type": "Organization",
7   "address": {
8     "@type": "PostalAddress",
9     "addressLocality": "Paris, France",
10    "postalCode": "F-75002",
11    "streetAddress": "38 avenue de l'Opera"
12  },
13   "email": "secretariat(at)example.org",
14   "name": "Organization X",
15   "description": "Description of org ...",
16   "telephone": "( 33 1) 42 68 53 00",
17   "member": [
18     {
19       "@type": "Organization",
20       "name": "Organization A",
21       "description": "Org A is a potential parent organization of Org X"
22     }
23   ],
24   "identifier": {
25     "@id": "https://grid.ac/institutes/grid.475727.4",
26     "@type": "PropertyValue",
27     "description": "UN Department of Economic and Social Affairs Sustainable
28     ↪Development",
29     "propertyID": "https://registry.identifiers.org/registry/grid",
30     "url": "https://grid.ac/institutes/grid.475727.4"
31   }
32 }
```

<graphviz.dot.Digraph at 0x7fdc65de0730>

### 6.3.1 Details: Authoritative Reference

For each profile there are a few key elements we need to know about. One key element is what the authoritative reference or canonical identifier is for a resource.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/org/x",
  "@type": "Organization",
  "identifier": {
    "@id": "https://grid.ac/institutes/grid.475727.4",
    "@type": "PropertyValue",
    "description": "UN Department of Economic and Social Affairs Sustainable ↵
↵Development",
    "propertyID": "https://registry.identifiers.org/registry/grid",
    "url": "https://grid.ac/institutes/grid.475727.4"
  }
}
```

```
<graphviz.dot.Digraph at 0x7fdc65de00a0>
```

## 6.4 References

- [schema:Person](#)
- [scheme:Organization](#)
- [Science on Schema Repository](#)
- <https://oceanexpert.org/>
  - [Example page expert](#)
  - [Example page institution](#)
  - [Ocean Expert: refernce: Adam Leadbetter](#)

## DOCUMENTS

### 7.1 About

Documents: These include datasets, reports or other documents

**See also:**

For OIH the focus is on generic documents which can scope reports, data and other resources. In those cases where the resources being described are of type Dataset you may wish to review patterns developed for GeoScience Datasets by the [ESIP Science on Schema](#) community.

### 7.2 Creative works (documents)

Documents will include maps, reports, guidance and other creative works. Due to this OIH will focus on a generic example of [schema.org/CreativeWork](#) and then provide examples for more focused creative work examples.

[Load in JSON-LD Playground](#)

[Load in Structured Data Testing Tool](#)

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "CreativeWork",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the creative work to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/report.pdf",
10  "contributor": {
11    "@type": "Organization",
12    "@id": "http://www.foo.org/orgID",
13    "legalName": "Some Institute"
14  },
15  "author": {
16    "@id": "https://www.sample-data-repository.org/person/51317",
17    "@type": "Person",
18    "name": "Dr Uta Passow",
19    "givenName": "Uta",
20    "familyName": "Passow",
21    "url": "https://www.sample-data-repository.org/person/51317"
22  },
23  "identifier":
```

(continues on next page)

(continued from previous page)

```

24     {
25         "@id": "https://doi.org/10.5066/F7VX0DMQ",
26         "@type": "PropertyValue",
27         "propertyID": "https://registry.identifiers.org/registry/doi",
28         "value": "doi:10.5066/F7VX0DMQ",
29         "url": "https://doi.org/10.5066/F7VX0DMQ"
30     },
31     "keywords":
32     {
33         "@type": "DefinedTerm",
34         "inDefinedTermSet": {
35             "@type": "DefinedTermSet",
36             "name": "Name of the set",
37             "description": "Description of the set",
38             "url": "url for the set"
39         },
40         "termCode": "A code that identifies this DefinedTerm within a
41         ↪DefinedTermSet"
42     }

```

```
<graphviz.dot.Digraph at 0x7ff53c712f70>
```

## 7.2.1 Details: Authoritative Reference

For each profile there are a few key elements we need to know about. One key element is what the authoritative reference or canonical identifier is for a resource.

```

{
    "@context": {
        "@vocab": "https://schema.org/"
    },
    "@id": "https://example.org/id/XYZ",
    "@type": "CreativeWork",
    "identifier": {
        "@id": "https://doi.org/10.5066/F7VX0DMQ",
        "@type": "PropertyValue",
        "propertyID": "https://registry.identifiers.org/registry/doi",
        "url": "https://doi.org/10.5066/F7VX0DMQ",
        "value": "doi:10.5066/F7VX0DMQ"
    }
}

```

```
<graphviz.dot.Digraph at 0x7ff52eacdb20>
```

## 7.2.2 Frame on author type Person

Our JSON-LD documents are graphs that can use framing to subset. In this case we can look closer at the author property which points to a type Person.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/XYZ",
  "@type": "CreativeWork",
  "author": {
    "@id": "https://www.sample-data-repository.org/person/51317",
    "@type": "Person",
    "familyName": "Passow",
    "givenName": "Uta",
    "name": "Dr Uta Passow",
    "url": "https://www.sample-data-repository.org/person/51317"
  }
}
```

```
<graphviz.dot.Digraph at 0x7ff52ead3c40>
```

## 7.2.3 References

- For dataset we can use [SOS Dataset](#)
- OBPS group is using JericoS3 API (ref: <https://www.jerico-ri.eu/>)
  - Traditional knowledge points here
  - sounds like they use dspace
- For other document these are likely going to be some [schema:CreativeWork](#) with there being many subtypes we can explore. See also here Adam Leadbetter's work at [Ocean best practices](#)
  - This is a great start and perhaps helps to highlight why SHACL shapes are useful
  - <https://irishmarineinstitute.github.io/erddap-lint/>
  - <https://github.com/earthcubearchitecture-project418/p419dcatservices/blob/master/CHORDS/DataFeed.jsonld> \*EMODnet (Coner Delaney)
  - ERDAP also
  - Are we talking links from [schema.org](#) that link to OGC and ERDAP services
  - Are these methods?
  - Sounds like may link to external metadata for interop they have developed in the community
- NOAA connected as well
  - Interested in OGC assets
  - ERDAP data platform





## **SPATIAL MAPS**

### **8.1 About**

Maps: A map represented by a static file or document

A map in this context would be a static file or document of some sort. Map services like those described by an OGC Catalogue Service or other GIS service would be described as a service.

---

**Note:** In the current context, [schema.org Map](https://schema.org/Map) typically references maps a document. Here we are likely to reference a KML, Shapefile or GeoPackage. We may wish to then indicate the type of document it is through a mimetype via encoding.

---

The [schema.org](https://schema.org/Map) type Map only offers one special property beyond the parent CreativeWork. That is a [mapType](https://schema.org/MapType) which is an enumeration of types that do not apply to OIH use cases. However, the use of the Map typing itself may aid in narrowing search requests later to a specific creative work.

[Load in JSON-LD Playground](#)

[Load in Structured Data Testing Tool](#)

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "Map",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the map to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/map.pdf",
10  "identifier": {
11    "@id": "https://doi.org/10.5066/F7VX0DMQ",
12    "@type": "PropertyValue",
13    "propertyID": "https://registry.identifiers.org/registry/doi",
14    "value": "doi:10.5066/F7VX0DMQ",
15    "url": "https://doi.org/10.5066/F7VX0DMQ"
16  },
17  "keywords": {
18    "@type": "DefinedTerm",
19    "inDefinedTermSet": {
20      "@type": "DefinedTermSet",
21      "name": "Name of the set",
22      "description": "Description of the set",
23      "url": "url for the set"
```

(continues on next page)

(continued from previous page)

```
24     },
25     "termCode": "A code that identifies this DefinedTerm within a DefinedTermSet"
26   }
27 }
```

```
<graphviz.dot.Digraph at 0x7f22f036fa90>
```

### 8.1.1 Details: Authoritative Reference

For each profile there are a few key elements we need to know about. One key element is what the authoritative reference or canonical identifier is for a resource.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/XYZ",
  "@type": "Map",
  "identifier": {
    "@id": "https://doi.org/10.5066/F7VX0DMQ",
    "@type": "PropertyValue",
    "propertyID": "https://registry.identifiers.org/registry/doi",
    "url": "https://doi.org/10.5066/F7VX0DMQ",
    "value": "doi:10.5066/F7VX0DMQ"
  }
}
```

```
<graphviz.dot.Digraph at 0x7f22e8ed1b20>
```

### 8.1.2 References

- For dataset we can use [SOS Dataset](#)
- OBPS group is using JericoS3 API (ref: <https://www.jerico-ri.eu/>)
  - Traditional knowledge points here
  - sounds like they use dspace
- For other document these are likely going to be some [schema:CretiveWork](#) with there being many subtypes we can explore. See also here Adam Leadbetter's work at [Ocean best practices](#)
  - This is a great start and perhaps helps to highlight why SHACL shapes are useful
  - <https://irishmarineinstitute.github.io/erddap-lint/>
  - <https://github.com/earthcubearchitecture-project418/p419dcatservices/blob/master/CHORDS/DataFeed.jsonld> \*EMODnet (Coner Delaney)
  - ERDAP also
  - Are we talking links from [schema.org](#) that link to OGC and ERDAP services
  - Are these methods?
  - Sounds like may link to external metadata for interop they have developed in the community

- NOAA connected as well
  - Interested in OGC assets
  - ERDAP data platform



## PROJECTS

### 9.1 About

**Project:** An enterprise (potentially individual but typically collaborative), planned to achieve a particular aim. Use properties from `Organization`, `subOrganization`/`parentOrganization` to indicate project sub-structures.

### 9.2 Research Project

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "ResearchProject",
6   "@id": "https://example.org/id/resproj/X",
7   "legalName": "Example Data Repository",
8   "name": "ExDaRepo",
9   "url": "https://www.example-data-repository.org",
10  "description": "Repo description ... ",
11  "logo": {
12    "@type": "ImageObject",
13    "url": "https://www.example-data-repository.org/logo.jpg"
14  },
15  "identifier": {
16    "@type": "PropertyValue",
17    "name": "Re3data DOI: 10.17616/R37P4C",
18    "propertyID": "https://registry.identifiers.org/registry/doi",
19    "value": "doi:10.17616/R37P4C",
20    "url": "https://doi.org/10.17616/R37P4C"
21  },
22  "contactPoint": {
23    "@id": "https://www.example-data-repository.org/about-us",
24    "@type": "ContactPoint",
25    "name": "Support",
26    "email": "info@example-data-repository.org",
27    "url": "https://www.example-data-repository.org/about-us",
28    "contactType": "customer support"
29  },
30  "funder": {
31    "@type": "FundingAgency",
32    "@id": "https://dx.doi.org/10.13039/100000001",
33    "legalName": "National Science Foundation",
```

(continues on next page)

(continued from previous page)

```

34     "alternateName": "NSF",
35     "url": "https://www.nsf.gov/"
36 },
37 "ethicsPolicy": {
38     "@type": "CreativeWork",
39     "@id": "https://example.org/id/XYZ",
40     "name": "Name or title of the document",
41     "description": "Description of the creative work ",
42     "url": "https://www.foo.org/creativework/ethicsPolicy.pdf"
43 },
44 "diversityPolicy": {
45     "@type": "CreativeWork",
46     "@id": "https://example.org/id/ABC",
47     "name": "Name or title of the document",
48     "description": "Description of the creative work",
49     "url": "https://www.foo.org/creativework/diversityPolicy.pdf"
50 },
51 "areaServed": [
52     {
53         "@type": "Place",
54         "geo": {
55             "@type": "GeoCoordinates",
56             "latitude": 39.3280,
57             "longitude": 120.1633
58         },
59         "description": "Description of the area served"
60     },
61     {
62         "@type": "Text",
63         "description": "Textual description of area served"
64     },
65     {
66         "@type": "AdministrativeArea",
67         "geo": {
68             "@type": "GeoCoordinates",
69             "latitude": 39.3280,
70             "longitude": 120.1633
71         },
72         "description": "Needs to be subset of Place, Review Place"
73     }
74 ],
75 "parentOrganization": {
76     "@type": "Organization",
77     "@id": "http://www.someinstitute.edu",
78     "legalName": "Some Institute",
79     "name": "SI",
80     "url": "http://www.someinstitute.edu",
81     "address": {
82         "@type": "PostalAddress",
83         "streetAddress": "234 Main St.",
84         "addressLocality": "Anytown",
85         "addressRegion": "ST",
86         "postalCode": "12345",
87         "addressCountry": "USA"
88     }
89 }
90 }

```

```
<graphviz.dot.Digraph at 0x7f28e80237f0>
```

## 9.2.1 Details: Authoritative Reference

For each profile there are a few key elements we need to know about. One key element is what the authoritative reference or canonical identifier is for a resource.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/resproj/X",
  "@type": "ResearchProject",
  "identifier": {
    "@type": "PropertyValue",
    "name": "Re3data DOI: 10.17616/R37P4C",
    "propertyID": "https://registry.identifiers.org/registry/doi",
    "url": "https://doi.org/10.17616/R37P4C",
    "value": "doi:10.17616/R37P4C"
  }
}
```

```
<graphviz.dot.Digraph at 0x7f28d9d63a90>
```

## 9.2.2 References

- <https://schema.org/Project>





## 10.1 About

From <https://schema.org/Course>

Course: A description of an educational course which may be offered as distinct instances at which take place at different times or take place at different locations, or be offered through different media or modes of study. An educational course is a sequence of one or more educational events and/or creative works which aims to build knowledge, competence or ability of learners.

## 10.2 Basic Course style

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/",
4     "endDate": {
5       "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
6     },
7     "startDate": {
8       "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
9     }
10  },
11  "@id": "https://example.org/id/course/1",
12  "@type": "Course",
13  "description": "In this course you will get an introduction to the main tools and
14  ideas in the data scientist's toolbox...",
15  "hasCourseInstance": [
16    {
17      "@type": "CourseInstance",
18      "courseMode": [
19        "MOOC1",
20        "online"
21      ],
22      "endDate": "2019-03-21",
23      "startDate": "2019-02-15",
24      "attendee": {
25        "@type": "Person",
26        "name": "Jane Doe",
27        "jobTitle": "Professor",
28        "telephone": "(425) 123-4567",
29        "url": "http://www.janedoe.com",
```

(continues on next page)

(continued from previous page)

```

29         "identifier": {
30             "@id": "ID_value_string",
31             "@type": "PropertyValue",
32             "propertyID": "This can be text or URL for an ID like ORCID",
33             "url": "https://foo.org/linkToPropertyIDPage",
34             "description": "Optional description of the ID"
35         }
36     },
37 },
38 {
39     "@type": "CourseInstance",
40     "courseMode": [
41         "MOOC2",
42         "online"
43     ],
44     "endDate": "2019-05-21",
45     "startDate": "2019-04-15"
46 }
47 ]
48 }
```

```
<graphviz.dot.Digraph at 0x7efc1d2d6190>
```

## 10.3 Simple Course

```

1 {
2     "@context": {
3         "@vocab": "https://schema.org/"
4     },
5     "@id": "https://example.org/id/course/2",
6     "@type": "Course",
7     "courseCode": "F300",
8     "name": "Physics",
9     "provider": {
10         "@type": "CollegeOrUniversity",
11         "name": "University of Bristol",
12         "url": {
13             "@id": "/provider/324/university-of-bristol"
14         }
15     }
16 }
```

```
<graphviz.dot.Digraph at 0x7efc1c250730>
```

## 10.4 References

- RDA Education and Training on handling of research data IG
- DC Tabular Application Profiles (DC TAP) - Primer
- <https://www.w3.org/TR/xmlschema11-2/>
  - Use YYYY-MM-DDThh:mm:ss or YYYY-MM-DD
- <http://www.marinetraining.eu/>
  - Example page
- <https://oceanexpert.org/>
- Example page
- OCTO
- <https://oceansummerschools.iode.org/>
- <https://www.openchannels.org/upcoming-events-list>
- <https://catalogue.odis.org/search/type=16>
- <https://clmeplus.org/>



## VESSELS

### 11.1 About

Exploring using [schema.org](https://schema.org) to describe research vessels. Note that [schema.org](https://schema.org) is a very broad vocabulary and as such specific concepts like research vessel is not well aligned.

One approach we can use, however, is to leverage [schema.org](https://schema.org) as a discovery layer and link more directly to detailed institutional metadata records.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@id": "https://example.org/id/X",
6   "@type": "Vehicle",
7   "name": "JOIDES Resolution",
8   "identifier": {
9     "@id": "https://example.org/id/vessel/X",
10    "@type": "PropertyValue",
11    "propertyID": "https://en.wikipedia.org/wiki/IMO_number",
12    "url": "https://example.org/id/vessel/X",
13    "description": "Vessel ID "
14  },
15  "additionalProperty": {
16    "@id": "ID_value_string",
17    "@type": "PropertyValue",
18    "propertyID": "https://en.wikipedia.org/wiki/IMO_number",
19    "url": "https://foo.org/linkToPropertyIDPage",
20    "description": "Any additional properties for the vessel"
21  },
22  "subjectOf": {
23    "@type": "DataDownload",
24    "name": "external-metadata.xml",
25    "description": "Metadata describing the vessel",
26    "encodingFormat": [
27      "application/xml",
28      "https://foo.org/ship01"
29    ],
30    "dateModified": "2019-06-12T14:44:15Z"
31  }
32 }
```

<graphviz.dot.Digraph at 0x7fdbe4240490>

### 11.1.1 Details: Authoritative Reference

For each profile there are a few key elements we need to know about. One key element is what the authoritative reference or canonical identifier is for a resource.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/X",
  "@type": "Vehicle",
  "identifier": {
    "@id": "https://example.org/id/vessel/X",
    "@type": "PropertyValue",
    "description": "Vessel ID ",
    "propertyID": "https://en.wikipedia.org/wiki/IMO_number",
    "url": "https://example.org/id/vessel/X"
  }
}
```

```
<graphviz.dot.Digraph at 0x7fdbd65adaf0>
```

## 11.2 References

- [ICES](#)
- [POGO](#)
- [EurOcean](#)
- [https://vocab.nerc.ac.uk/search\\_nvs/C17/](https://vocab.nerc.ac.uk/search_nvs/C17/)
- [SeaDataNet](#)
- [Marine Facilities Planner](#)
- [EuroFleets](#)
- Identifiers to use include NOCD Code, Call Sign, ICES Shipcode, MMSI Code, IMO Code

## SPATIAL GEOMETRY

### 12.1 About

The primary OIH guidance will be to use the OGC [GeoSPARQL](https://www.opengis.net/ont/geosparql) vocabulary. The [schema.org](https://schema.org) spatial types and properties are not well defined and difficult at times to reliably translate to geometries.

### 12.2 Simple GeoSPARQL WKT

This is a simple example of how to embed a WKT string via GeoSPARQL into a record. Well Know Text (WKT) is a OGC standard referenced at: <https://www.opengis.net/standards/wkt-crs>. A more accessible example can be found at Wikipedia: [Well-known text representation of geometry](https://en.wikipedia.org/wiki/Well-known_text_representation_of_geometry).

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/",
4     "geosparql": "http://www.opengis.net/ont/geosparql#"
5   },
6   "@id": "https://example.org/id/XYZ",
7   "@type": "Dataset",
8   "name": "Data set name",
9   "geosparql:hasGeometry": {
10     "@type": "http://www.opengis.net/ont/sf#Point",
11     "geosparql:asWKT": {
12       "@type": "http://www.opengis.net/ont/geosparql#wktLiteral",
13       "@value": "POINT(-76 -18)"
14     },
15     "geosparql:crs": {
16       "@id": "http://www.opengis.net/def/crs/OGC/1.3/CRS84"
17     }
18   }
19 }
```

```
<graphviz.dot.Digraph at 0x7f54442f6760>
```

## 12.3 Classic Schema.org

Is is a simple example of the existing Schema.org pattern for a lat long value. This pattern is of little use other than perhaps to Google. There is the pending GeospatialGeometry which is a type Intangible (and not Place referenced by spatialCoverage). This will be a subtype of GeoShape. There are inconsistencies with Schema.org guidance for textual geometry representation and that of Well Known Text (WKT).

```

1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@id": "https://example.org/id/XYZ",
6   "@type": "Dataset",
7   "name": "Data set name",
8   "spatialCoverage": {
9     "@type": "Place",
10    "geo": {
11      "@type": "GeoCoordinates",
12      "latitude": 39.3280,
13      "longitude": 120.1633
14    }
15  }
16 }

```

```
<graphviz.dot.Digraph at 0x7f5444324490>
```

## 12.4 Option review, SOS Issue 105

From the referenced SOS issue 105:

```

1 {
2   "@context": {
3     "@version": 1.1,
4     "geoblob": {
5       "@id": "http://example.com/vocab/json",
6       "@type": "@json"
7     },
8     "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
9     "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
10    "xsd": "http://www.w3.org/2001/XMLSchema#",
11    "description": "http://igsn.org/core/v1/description",
12    "geosparql": "http://www.opengis.net/ont/geosparql#",
13    "schema": "https://schema.org/"
14  },
15  "@id": "https://samples.earth/id/do/bqs2dn2u6s73o70jdup0",
16  "@type": "http://igsn.org/core/v1/Sample",
17  "description": "A fake ID for testing",
18  "schema:subjectOf": [
19    {
20      "schema:url": "https://samples.earth/id/do/bqs2dn2u6s73o70jdup0.geojson",
21      "@type": "schema:DigitalDocument",
22      "schema:format": [
23        "application/vnd.geo+json"
24      ],

```

(continues on next page)



(continued from previous page)

```

25         "schema:conformsTo": "https://igsn.org/schema/spatial.schema.json"
26     },
27 ],
28     "geosparql:hasGeometry": {
29         "@id": "_:N98e75cacc29f40deb555eb583cb162dc",
30         "@type": "http://www.opengis.net/ont/sf#Point",
31         "geosparql:asWKT": {
32             "@type": "http://www.opengis.net/ont/geosparql#wktLiteral",
33             "@value": "POINT(-76 -18)"
34         },
35         "geosparql:crs": {
36             "@id": "http://www.opengis.net/def/crs/OGC/1.3/CRS84"
37         }
38     },
39     "geoblob": {
40         "type": "GeometryCollection",
41         "geometries": [{
42             "type": "Point",
43             "coordinates": [-76, -18]
44         }]
45     },
46     "schema:spatialCoverage": {
47         "@type": "schema:Place",
48         "schema:geo": {
49             "@type": "schema:GeoCoordinates",
50             "schema:latitude": -18,
51             "schema:longitude": -76
52         }
53     }
54 }

```

```
<graphviz.dot.Digraph at 0x7f54442f6c40>
```

## 12.5 References

- GeoAPI at GitHub
- Science on Schema Issue 105
  - Leverages subjectOf to connect to a Thing / CreativeWork
- <https://www.unsalb.org/>
- <https://www.un.org/geospatial/>
- [schema.org/spatial](http://schema.org/spatial)
- [schema.org/GeospatialGeometry](http://schema.org/GeospatialGeometry)
- SOS pattern follows:
  - spatialCoverage -> Place -> geo -> GeoCoordinates OR GeoShape
- Some groups are using GeoNode
  - [schema.org](http://schema.org) issues
- ICAN & Schema.org

- OGC SELFIE
- Think broad
- Science on Schema [spatial](#) for dataset guidance

## 13.1 About

This section will provide information on the service type. This is not one of the main OIH types. However, we will provide guidance here on describing services using [schema.org](https://schema.org).

It should be noted that this might be a simple link to an OpenAPI or some other descriptor document. Also, [schema.org](https://schema.org) is not rich enough for complex descriptions and itself borrows from the [Hydra](https://www.hydra.cc/) vocabulary. It may be required to leverage Hydra if complex descriptions are needed.

The graph describes a service than can be invoked with:

```
curl --data-binary "@yourfile.jpg" -X POST https://us-central1-top-operand-112611.
➔cloudfunctions.net/function-1
```

This with POST a jpeg to the service and get back a simple text response with some information about the image.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "Action",
6   "@id": "https://us-central1-top-operand-112611.cloudfunctions.net/function-1",
7   "result": {
8     "@type": "DataDownload",
9     "encodingFormat": "text/plain",
10    "description": "a simple text result for the RGB counts"
11  },
12  "target": {
13    "@type": "EntryPoint",
14    "urlTemplate": "https://us-central1-top-operand-112611.cloudfunctions.net/
15 ➔function-1",
16    "httpMethod": "POST",
17    "contentType": ["image/jpeg", "image/png"]
18  },
19  "object": {
20    "@type": "ImageObject",
21    "description": "A JPEG or PNG to analyze the RGB counts"
22  }
23 }
```

```
<graphviz.dot.Digraph at 0x7f66d83a5f70>
```

## 13.2 References

- <https://schema.org/docs/actions.html>
- <https://schema.org/Action>
- <https://www.w3.org/TR/web-share/>
- <https://www.hydra-cg.com/spec/latest/core/>

## DEFINED TERMS

### 14.1 About

During generation of the structured data a provide may wish to either use or publish a set of controlled vocabulary terms or a similar set.

Within [schema.org](https://schema.org) this could be done by leveraging the “DefinedTerm” amd “DefinedTermSet” types.

These types allow us both to define a set of terms and use a set of terms in describing a thing.

Note that DefinedTerm is an intangible and can connect to most types in [Schema.org](https://schema.org). So we can use them in places such as:

- CreativeWork -> keyword
- LearningResource -> teaches
- PropertyValue -> valueReference
- LearningResource -> competencyRequired
- CreativeWork -> learningResourceType

```
1  [
2      {
3          "@context": "https://schema.org/"
4      },
5      {
6          "@type": [
7              "DefinedTermSet",
8              "Book"
9          ],
10         "@id": "http://openjurist.org/dictionary/Ballentine",
11         "name": "Ballentine's Law Dictionary"
12     },
13     {
14         "@type": "DefinedTerm",
15         "@id": "http://openjurist.org/dictionary/Ballentine/term/calendar-year",
16         "name": "calendar year",
17         "description": "The period from January 1st to December 31st, ↵
↵inclusive, of any year.",
18         "inDefinedTermSet": "http://openjurist.org/dictionary/Ballentine"
19     },
20     {
21         "@type": "DefinedTerm",
```

(continues on next page)

(continued from previous page)

```

22         "@id": "http://openjurist.org/dictionary/Ballentine/term/schema",
23         "name": "schema",
24         "description": "A representation of a plan or theory in the form of_
↪an outline or model.",
25         "inDefinedTermSet": "http://openjurist.org/dictionary/Ballentine"
26     }
27 ]

```

## 14.2 Organization in List

```

1  {
2    "@context": {
3      "@vocab": "https://schema.org/"
4    },
5    "@id": "ID_for_this_metadata_record",
6    "@type": "ItemList",
7    "url": "http://multivarki.ru?filters%5Bprice%5D%5BLTE%5D=39600",
8    "numberOfItems": "315",
9    "itemListElement": [
10     {
11       "@type": [
12         "Service",
13         "ResearchProject"
14       ],
15       "legalName": "Sample Data Repository Office",
16       "name": "SDRO",
17       "url": "https://www.sample-data-repository.org",
18       "description": "The Sample Data Repository Service...",
19       "category": [
20         "Biological Oceanography",
21         "Chemical Oceanography"
22       ],
23       "provider": {
24         "@id": "https://www.sample-data-repository.org"
25       },
26       "parentOrganization": {
27         "@type": "Organization",
28         "@id": "http://www.someinstitute.edu",
29         "legalName": "Some Institute",
30         "name": "SI",
31         "url": "http://www.someinstitute.edu",
32         "address": {
33           "@type": "PostalAddress",
34           "streetAddress": "234 Main St.",
35           "addressLocality": "Anytown",
36           "addressRegion": "ST",
37           "postalCode": "12345",
38           "addressCountry": "USA"
39         }
40       }
41     },
42     {
43       "@type": "Product",
44       "name": "Product name"
45     }
46   ]
47 }

```

(continues on next page)

(continued from previous page)

```
45     }  
46   ]  
47 }
```

```
<graphviz.dot.Digraph at 0x7fd8b8616ac0>
```

## 14.3 References

- [schema.org/DefinedTerm](http://schema.org/DefinedTerm)
- [schema.org/DefinedTermSet](http://schema.org/DefinedTermSet)





## LANGUAGES

### 15.1 About

JSON-LD fully support the identification of the language types.

Properties such as label, description, keyword etc can be extended in the context with a container language attribute notation.

This will allow the use of standard language codes (fr, es, en, de, etc) to be used when describing these properties.

```
1 {  
2   "@context": {  
3     "vocab": "http://example.com/vocab/",  
4     "label": {  
5       "@id": "vocab:label",  
6       "@container": "@language"  
7     }  
8   },  
9   "@id": "http://example.com/queen",  
10  "label": {  
11    "en": "The Queen",  
12    "de": [ "Die Königin", "Ihre Majestät" ]  
13  }  
14 }
```

```
<graphviz.dot.Digraph at 0x7fdad8513cd0>
```

In graph space the resulting triples from the above are:

```
<http://example.com/queen> <http://example.com/vocab/label> "Die Königin"@de .  
<http://example.com/queen> <http://example.com/vocab/label> "Ihre Majestät"@de .  
<http://example.com/queen> <http://example.com/vocab/label> "The Queen"@en .
```

with language encoding attributes in place. These can be used in searching and result filters.

Note, this can cause issues in query space since the concept of

```
"The Queen"
```

and

```
"The Queen"@en
```

are different and so care must be taken the creation of the SPARQL queries not to accidentally imposed implicate filters through the use of language types.



## LINKING TO DOCUMENTS AND RESOURCES

Leveraging the ability to link between resources can serve many goals. We may wish to demonstrate connections between people and courses they have taken or organizations they are connected with. We may be wishing to link documents to people or organizations.

This section will review two key thematic profiles and some examples of how to express links from them to other resources. Or goal will be different in cases. The two profiles are type CreativeWork and type Organization.

In the case of CreativeWork we are looking to express connections to the publisher and provider of the creative work. This is mostly to connect these works with the responsible party associated with them.

In the case of Organization our purpose may be to express alignment to various principles and policies. These might provide people with an understanding of the goals of an organization when they are searching for or assessing them.

### 16.1 Organization link options

In the following section we will look at three different options for expressing links between an organization and resources that describe the policy and principles of the subject organization.

First we will see the full data graph. We have highlighted the sections we we will review here. Namely the subjectOF and publishingPrinciples predicates.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@id": "https://example.org/id/org/x",
6   "@type": "Organization",
7   "address": {
8     "@type": "PostalAddress",
9     "addressLocality": "Paris, France",
10    "postalCode": "F-75002",
11    "streetAddress": "38 avenue de l'Opera"
12  },
13  "email": "secretariat(at)example.org",
14  "name": "Organization X",
15  "description": "Description of org ...",
16  "telephone": "( 33 1) 42 68 53 00",
17  "member": [
18    {
19      "@type": "Organization",
20      "name": "Organization A",
21      "description": "Org A is a potential parent organization of Org X"
```

(continues on next page)

(continued from previous page)

```

22     }
23   ],
24   "identifier": {
25     "@id": "https://grid.ac/institutes/grid.475727.4",
26     "@type": "PropertyValue",
27     "description": "UN Department of Economic and Social Affairs Sustainable_
↪Development",
28     "propertyID": "https://registry.identifiers.org/registry/grid",
29     "url": "https://grid.ac/institutes/grid.475727.4"
30   },
31   "subjectOf": {
32     "@type": "CreativeWork",
33     "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
34     "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
↪SDGIO_00020173",
35     "name": "UNSD SDG indicator code:C140c01",
36     "description": "Number of countries making progress ... the oceans and their_
↪resources"
37   },
38   "publishingPrinciples": {
39     "@type": "CreativeWork",
40     "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
41     "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
↪SDGIO_00020173",
42     "name": "FAIR data principles",
43     "description": "Number of countries making progress ... the oceans and their_
↪resources"
44   }
45 }

```

```
<graphviz.dot.Digraph at 0x7f27844c5af0>
```

## 16.1.1 subjectOf

Lets take a look at subjectOf.

```

{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/org/x",
  "@type": "Organization",
  "subjectOf": {
    "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
    "@type": "CreativeWork",
    "description": "Number of countries making progress ... the oceans and their_
↪resources",
    "name": [
      "FAIR data principles",
      "UNSD SDG indicator code:C140c01"
    ],
    "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
↪SDGIO_00020173"
  }
}

```

```
<graphviz.dot.Digraph at 0x7f2776035a90>
```

## 16.1.2 publishingPrinciples

Lets take a look at publishingPrinciples.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/org/x",
  "@type": "Organization",
  "publishingPrinciples": {
    "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
    "@type": "CreativeWork",
    "description": "Number of countries making progress ... the oceans and their_
↪resources",
    "name": [
      "FAIR data principles",
      "UNSD SDG indicator code:C140c01"
    ],
    "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
↪SDGIO_00020173"
  }
}
```

```
<graphviz.dot.Digraph at 0x7f277603ed00>
```

## 16.2 SDG Linkage

The following provides an example of how Sustainable Development Goals (SDGs) could be linked to a [Schema.org](https://schema.org) defined type using `subjectOf`.

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "CreativeWork",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the resource to aid in searching",
9   "distribution": {
10    "@type": "DataDownload",
11    "contentUrl": "https://www.sample-data-repository.org/dataset/472032.tsv",
12    "encodingFormat": "text/tab-separated-values"
13  },
14  "subjectOf": {
15    "@type": "CreativeWork",
16    "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
17    "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/SDGIO_
↪00020173",
18    "name": "UNSD SDG indicator code:C140c01",
```

(continues on next page)

(continued from previous page)

```

19   "description": "Number of countries making progress ... the oceans and their
    ↪resources"
20   },
21   "maintainer" : {
22     "@type" : "Organization",
23     "@id": "https://ror.org/050bms902",
24     "description": "UN Department of Economic and Social Affairs Sustainable
    ↪Development"
25   }
26 }

```

```
<graphviz.dot.Digraph at 0x7f277643e070>
```

## 16.3 Creative work link options

```

1  {
2    "@context": {
3      "@vocab": "https://schema.org/"
4    },
5    "@id": "https://example.org/id/XYZ",
6    "@type": "CreativeWork",
7    "identifier": {
8      "@id": "https://doi.org/10.5066/F7VX0DMQ",
9      "@type": "PropertyValue",
10     "propertyID": "https://registry.identifiers.org/registry/doi",
11     "url": "https://doi.org/10.5066/F7VX0DMQ",
12     "value": "doi:10.5066/F7VX0DMQ"
13   },
14   "subjectOf": {
15     "@type": "CreativeWork",
16     "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
17     "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
    ↪SDGIO_00020173",
18     "name": "UNSD SDG indicator code:C140c01",
19     "description": "Number of countries making progress ... the oceans and their
    ↪resources"
20   },
21   "publishingPrinciples": {
22     "@type": "CreativeWork",
23     "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
24     "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
    ↪SDGIO_00020173",
25     "name": "FAIR data principles",
26     "description": "Number of countries making progress ... the oceans and their
    ↪resources"
27   },
28   "citation": {
29     "@type": "CreativeWork",
30     "@id": "http://purl.unep.org/sdg/SDGIO_00020173",
31     "url": "http://www.ontobee.org/ontology/SDGIO?iri=http://purl.unep.org/sdg/
    ↪SDGIO_00020173",
32     "name": "UNSD SDG indicator code:C140c01",
33     "description": "Number of countries making progress ... the oceans and their
    ↪resources"

```

(continues on next page)

(continued from previous page)

```
34  },
35  "provider": {
36    "@id": "https://www.sample-data-repository.org",
37    "@type": "Organization",
38    "legalName": "Sample Data Repository Office",
39    "name": "SDRO",
40    "sameAs": "http://www.re3data.org/repository/r3dxxxxxxx",
41    "url": "https://www.sample-data-repository.org"
42  },
43  "publisher": {
44    "@id": "https://www.sample-data-repository.org"
45  }
46 }
```

```
<graphviz.dot.Digraph at 0x7f277643ff10>
```

## 16.4 Refs

- SDGs
- SDG targets
- SDG indicators





## IDENTIFIER

### 17.1 About

```
1 {
2   "@context": {
3     "@vocab": "https://schema.org/"
4   },
5   "@type": "Map",
6   "@id": "https://example.org/id/XYZ",
7   "name": "Name or title of the document",
8   "description": "Description of the map to aid in searching",
9   "url": "https://www.sample-data-repository.org/creativework/map.pdf",
10  "identifier": {
11    "@id": "https://doi.org/10.5066/F7VX0DMQ",
12    "@type": "PropertyValue",
13    "propertyID": "https://registry.identifiers.org/registry/doi",
14    "value": "doi:10.5066/F7VX0DMQ",
15    "url": "https://doi.org/10.5066/F7VX0DMQ"
16  },
17  "keywords": {
18    "@type": "DefinedTerm",
19    "inDefinedTermSet": {
20      "@type": "DefinedTermSet",
21      "name": "Name of the set",
22      "description": "Description of the set",
23      "url": "url for the set"
24    },
25    "termCode": "A code that identifies this DefinedTerm within a DefinedTermSet"
26  }
27 }
```

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/XYZ",
  "@type": "Map",
  "identifier": {
    "@id": "https://doi.org/10.5066/F7VX0DMQ",
    "@type": "PropertyValue",
    "propertyID": "https://registry.identifiers.org/registry/doi",
    "url": "https://doi.org/10.5066/F7VX0DMQ",
    "value": "doi:10.5066/F7VX0DMQ"
```

(continues on next page)

(continued from previous page)

```
}  
}
```

```
<graphviz.dot.Digraph at 0x7f0949adba00>
```

# **Part III**

## **Aggregation**



## AGGREGATOR

### 18.1 Introduction

This section introduces the the OIH approach to indexing. Currently, OIH is using the [Gleaner](#) software to do the indexing and leverages the Gleaner IO [gleaner-compose](#) Docker Compose files for the server side architecture. For more information on Docker Compose files visit the [Overview of Docker Compose](#). The gleaner-compose repository holds Docker compose files that can set up various environments that Gleaner needs.

The figure below gives a quick overview of the various compose options for setting up the supporting architecture for Gleaner. A fully configured system where all the indexing and data services are running and exposing services to the net, a total of five containers are run. In many case you may run fewer than this.

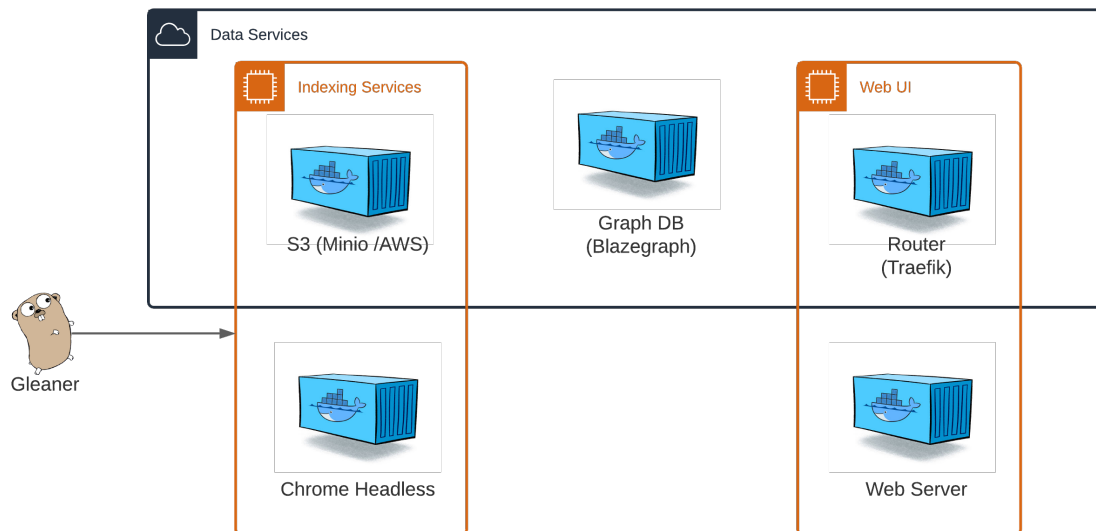


Fig. 1: The various compose options for Gleaner

### 18.1.1 Container overview

- S3 (Minio / AWS): This is the only container that is required in all cases to run. Gleaner needs an S3 compatible object store. By default we use the [Minio](#) object store
- Chrome Headless: In cases where providers place the JSON-LD documents into the pages with Javascript, we need to render the page before reading and accessing the DOM. This is done using Chrome Headless
- Graph data base: Gleaner extracts JSON-LD documents from resources. These JSON-LD documents are representations of the RDF data mode. To queries on them at scale, it easiest to load the triples into a compatible graph database. Sometimes we call this a triplestore. For OIH we use the [Blazgraph triplestore](#).
- Router: If we wish to deploy this setup onto the net, we will route to route all the services through a single domain. To do this network routing we use [Traefik](#). This router is not required for local use and alternative routers like [Caddy](#) or [nginx](#) are also valid options.
- Web Server: If you wish to serve a web UI for the index, then you can also leverage this setup to serve that. Again, this is optional and your web site may be hosted elsewhere and simply call to the index in compliance with CORS settings. There is an example web server that leverages the object store available in this setup.

### 18.1.2 Gleaner

As mentioned Gleaner is a single binary app (ie, one file). It can be run on Linux, Mac OS X or Windows. It does not need to be run on the same machine as the supporting services as it can connect to them over the network. So, for example, they could be hosted in commercial cloud services or on remote servers.

You can download and compile the code from the previously mentioned github repository or the [releases page](#).

A single configuration file provides the settings Gleaner needs. Additionally, a local copy of the current [schema.org](#) context file should be downloaded and available to the app. This file is needed for many operations and access it over the net is slow and often rate limited depending on the source. You can download the file at the [Schema.org for Developers](#) page.

This setup show in the above figure is the typical setup for Gleaner and is detailed in the [Quick Start](#) section.

### 18.1.3 Indexing Services

This is the basic indexing service requirements. At a minimum we need the object store and the Chrome headless containers scoped in the *Indexing Services* box above. More details on this set can be found in [Indexing Services](#).

### 18.1.4 Data Services

A more expanded set of services is defined in the [Data Services](#) section. This section discussion a setup more designed to address a server setup tht will support indexing and also present the resulting indexes to the broader internet.

### 18.1.5 Web UI

As mentioned, if you wish to serve a web UI for the index, then you can leverage this setup to serve that. Again, this is optional and your web site can be hosted elsewhere and simply call to the index in compliance with CORS settings. Some tailes on this can be found in the [Interfaces](#) section.

### 18.1.6 Alternatives

Note, the Gleaner ecosystem is not a requirement. OIH follows the structured data on the web and data on the web best practices patterns. Being web architecture based, there are many open source tools and scripting solutions you might use. You may wish to explore the [Alternative Approaches](#) section for more on this.

What follows is a bit more detail on the setup used by Gleaner. Experienced users will see where they can swap out elements for their own preference. Like a different triplestore, or wish to leverage a commercial object store? Simply modify the architecture to do so.

## 18.2 ODIS Catalog as Index Source

Before we discuss indexing source a key question is what source will be indexed. OIH is not a web crawl in that it doesn't move from source to source based on the content of those sources.

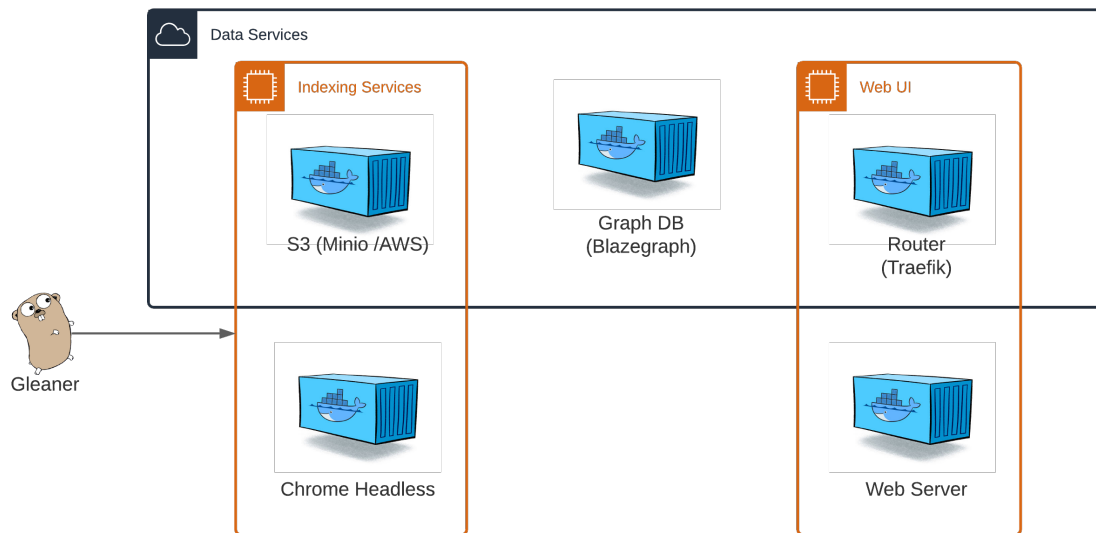
Rather, the OIH index is based on a list of sources selected ahead of time. At this time that set of sources is based on those partners engaged in the development phase of OIH. As the work moves to a more routine operation the sources will come from the [ODIS Catalog](#).

The ODIS Catalog will then act as a curated source of domains for inclusion in the Ocean InfoHub. This will provide a level of curation and vetting of sources and ensure sources are aware of the technical requirements for inclusion in the OIH index.





## INDEXING WITH GLEANER



### 19.1 Gleaner (app)

The Gleaner applications performs the retrieval and loading of JSON-LD documents from the web following structured data on the web patterns. Gleaner is available for Linux, Mac OS X and Windows.

While Gleaner is a stand alone app, it needs to interact with an object store to support data storage and other operations. These dependencies are met within the Gleaner Indexing Services or Data Service Docker compose files.

**Warning:** This documentation is in development. The primary testing environments are Linux and other UNIX based platforms such as Mac OS X. If you are on Windows, there may be some issues. If you can use a Linux subsystem on Windows, you may experience better results. We will test with Windows eventually and update documentation as needed.

### 19.1.1 Quick Start steps

This quick start guide is focused on setting up and testing Gleaner in a local environment. It is similar to how you might run Gleaner in a production environment but lacks the routing and other features likely desired for such a situation.

---

**Note:** This documentation assumes a basic understanding of Docker and experience with basic Docker activities like starting and stopping containers. It also assumes an understanding of using a command line interface and editing configuration files in the YAML format.

---

---

#### Command

From this point down, the documentation will attempt to put all commands you should issue in this admonition style box.

---

In the end, this is the table of applications and config files you will need. In this guide we will go through downloading, setting them up and running Gleaner to index documents from the web.

Table 1: Required Applications and Config Files

Gleaner	Docker	Minio Client
config.yaml	<a href="#">setenv.sh</a>	<a href="#">load2blaze.sh</a>
schemaorg-current-https.jsonld	gleaner-DS-NoRouter.yml	

#### Grab Gleaner and the support files we need

We will need to get the Gleaner binary for your platform and also the Gleaner configuration file template. To do this, visit the [Gleaner Releases page](#) and pick the release *Ocean InfoHubdev rc1*. Under the *Assets* drop down you should see the files we need. Get:

- Gleaner for your platform
- Gleaner config template: `template_v2.0.yaml`
- Gleaner indexing service compose file: `gleaner-IS.yml`
- Helper environment setup script: [setenvIS.sh](#)

For this demonstration, we will be running on linux, so this would look something like:

---

#### Command

```
curl -L -O https://github.com/earthcubearchitecture-project418/gleaner/releases/
↳download/2.0.25/gleaner
curl -L -O https://github.com/earthcubearchitecture-project418/gleaner/releases/
↳download/2.0.25/gleaner-IS.yml
curl -L -O https://github.com/earthcubearchitecture-project418/gleaner/releases/
↳download/2.0.25/setenvIS.sh
curl -L -O https://github.com/earthcubearchitecture-project418/gleaner/releases/
↳download/2.0.25/template_v2.0.yaml
```

---

**Note:** You can download these with any tool you wish or through the browser. Above we downloaded used the command line curl tool. For GitHub, be sure to add the `-L` to inform curl to follow redirects to the object to download.

---

---

### Command

You may need to change the permission on your gleaner file to ensure it can be run. On Linux this would look something like the following.

```
chmod 755 gleaner
```

---

We then need to visit [Schema.org](https://schema.org/) for Developers to pull down the appropriate JSON-LD context. For this work we will want to pull down the *schemaorg-current-https* in JSON-LD format. It also should work to do something similar to the following:

---

### Command

```
curl -O https://schema.org/version/latest/schemaorg-current-https.jsonld
```

---

## About the compose file(s)

The above steps have collected the resources for the indexer. We now want to set up the services that Gleaner will use to perform the indexing. To do that we use Docker or an appropriate run time alternative like Podman or others. For this example, we will assume you are using the Docker client.

As noted, a basic understanding of Docker and the ability to issue Docker cli commands to start and stop containers is required. If you are new to Docker, we recommend you visit and read: [Get Started with Docker](#).

We need to select the type of services we wish to run. The various versions of these Docker compose file can be found in the [Gleaner-compose deployment directory](#).

Why pick one over the other?

Choose Gleaner IS if you simply wish to retrieve the JSON-LD into a data warehouse to use in your own workflows

Choose Gleaner DS if you wish to build out a graph and want to use the default contains used by Gleaner.

---

**Note:** We won't look at this file in detail here since there will hopefully be no required edits. You can see the file in detail in the Index Services section.

---

## Edit environment variables setup script

We have Docker and the appropriate compose file. The compose files require a set of environment variables to be populated to provide the local hosts information needed to run. You can set these yourself or use or reference the [setenv.sh](#) file in the Gleaner-compose repository in the [Gleaner-compose deployment directory](#). You may also need to visit information about permissions at [Post-installation steps for Linux](#) if you are having permission issues.

Let's take a look at the script.

```
1 #!/bin/bash
2
3 # Object store keys
4 export MINIO_ACCESS_KEY=worldsbestaccesskey
5 export MINIO_SECRET_KEY=worldsbestsecretkey
```

(continues on next page)

(continued from previous page)

```
6
7 # local data volumes
8 export GLEANER_BASE=/tmp/gleaner/
9 mkdir --parents ${GLEANER_BASE}
10 export GLEANER_OBJECTS=${GLEANER_BASE}/datavol/s3
11 export GLEANER_GRAPH=${GLEANER_BASE}/datavol/graph
12
```

You may wish to edit file to work better with your environment. By default it will attempt to use localhost to resolve with and host local runtime data in a /tmp/gleaner directory.

### Spin up the containers

Load our environment variables to the shell:

#### Command

```
source setenv.sh
```

Then start the containers:

#### Command

```
docker-compose -f gleaner-IS.yml up -d
```

If all has gone well, you should be able to see your running containers with

#### Command

```
docker ps
```

and see results similar to:

CONTAINER ID	STATUS	IMAGE	PORTS	COMMAND	NAMES	CREATED
c4b7097f5e06	↪seconds ago	nawer/blazegraph	Up 7 seconds	0.0.0.0:9999->9999/tcp	"docker-entrypoint.s..." test_triplestore_1	8
ca08c24963a0	↪seconds ago	minio/minio:latest	Up 7 seconds	0.0.0.0:9000->9000/tcp	"/usr/bin/docker-ent..." test_s3system_1	8
24274eba0d34	↪seconds ago	chromedp/headless-shell:latest	Up 7 seconds	0.0.0.0:9222->9222/tcp	"/headless-shell/hea..." test_headless_1	8

## Edit Gleaner config file

We have all the files we need and we have our support services running. The next and final step is to edit our Gleaner configuration file. This will let Gleaner know the location of the support services, the JSON-LD context file and the locations of the resources we wish to index.

Let's take a look at the configuration file and then break down each section.

```

1  ---
2  minio:
3      address: 0.0.0.0
4      port: 9000
5      accessKey: worldsbestaccesskey
6      secretKey: worldsbestsecretkey
7      ssl: false
8      bucket: gleaner
9  gleaner:
10     runid: oih # this will be the bucket the output is placed in...
11     summon: true # do we want to visit the web sites and pull down the files
12     mill: true
13  context:
14     cache: true
15  contextmaps:
16  - prefix: "https://schema.org/"
17     file: "../jsonldcontext.json" # wget http://schema.org/docs/jsonldcontext.jsonld
18  - prefix: "http://schema.org/"
19     file: "../jsonldcontext.json" # wget http://schema.org/docs/jsonldcontext.jsonld
20  summoner:
21     after: "" # "21 May 20 10:00 UTC"
22     mode: full # full || diff: If diff compare what we have currently in gleaner to
23     ↪ sitemap, get only new, delete missing
24     threads: 1
25     delay: 0 # milliseconds (1000 = 1 second) to delay between calls (will FORCE
26     ↪ threads to 1)
27     headless: http://0.0.0.0:9222 # URL for headless see docs/headless
28  millers:
29     graph: true
30     shacl: false
31  sitegraphs:
32  - name: aquadocs
33     url: https://oih.aquadocs.org/aquadocs.json
34     headless: false
35     pid: https://www.re3data.org/repository/aquadocs
36     properName: AquaDocs
37     domain: https://aquadocs.org
38  sources:
39  - name: marinetraining
40     url: https://www.marinetraining.eu/sitemap.xml
41     headless: false
42     pid: https://www.re3data.org/repository/marinetraining
43     properName: Marine Training EU
44     domain: https://marinetraining.eu/
45  - name: oceanexperts
46     url: https://oceanexpert.org/assets/sitemaps/sitemapTraining.xml
47     headless: false
48     pid: https://www.re3data.org/repository/oceanexpert
49     properName: OceanExpert UNESCO/IOC Project Office for IODE
50     domain: https://oceanexpert.org/

```

(continues on next page)

(continued from previous page)

```
49 - name: obis
50   url: https://obis.org/sitemap/sitemap_datasets.xml
51   headless: false
52   pid: https://www.re3data.org/repository/obis
53   properName: Ocean Biodiversity Information System
54   domain: https://obis.org
```

### Run gleaner

For this example we are going to run Gleaner directly. In a deployed instance you may run Gleaner via a script or cron style service. We will document that elsewhere.

We can do a quick test of the setup.

---

#### Command

```
./gleaner -cfg template_v2.0 -setup
```

---

For now, we are ready to run Gleaner. Try:

---

#### Command

```
./gleaner -cfg template_v2.0
```

---

---

**Note:** Leave the suffix like .yaml off the name of the config file. The config system can also read json and other formats. So simply leave the suffix off and let the config code inspect the contents.

---

### 19.1.2 Load results to a graph and test

You have set up the server environment and Gleaner and done your run. Things look good but you don't have a graph you can work with yet. You need to load the JSON-LD into the triplestore in order to start playing.

#### Minio Object store

To view the object store you could use your browser and point it on the default minio port at 9000. S typically something like localhost:9000.

If you wish to continue to use the command line you can use the Minio client at [Minio Client Quickstart guide](#).

Once you have it installed and working, you can write an entry for our object store with:

---

#### Command

```
./mc alias set minio http://0.0.0.0:9000 worldsbetaccesskey worldsbetsecretkey
```

---

## Load Triplestore

We now want to load these objects, which are JSON-LD files holding RDF based graph data, into a graph database. We use the term, triplestore, to define a graph database designed to work with the RDF data model and provide SPARQL query support over that graph data.

- Simple script loading
- Nabu
- Try out a simple SPARQL query





## INDEXING SERVICES

Gleaner can not run alone and relies on a couple of Open Container Initiative (OCI) containers to support it. For this document, we will assume you are using Docker but this will work with Podman or other OCI compliant orchestration environments. These Gleaner Indexing Services are necessary to use Gleaner. The exception to this would be if you are using a 3rd party objects store like AWS S3 or Wasabi.

- **Object Store** An S3 compliant object store supporting S3 APIs including S3Select. For open source this is best satisfied with the Minio Object Store. For commercial cloud AWS S3 or hosted Ceph services will work.
- **Headless Chrome** (technically optional) This is only needed where you expect the sources you index to use Javascript to include the JSON-LD in the pages. If you know your sources do not use this publishing pattern and rather include the JSON-LD in the static page, then you don't need this container running.

IS represent the minimum required services to support Gleaner. With IS you have an object store in the form of [Minio](#) and a headless chrome server in the form of [chromedp/headless-shell](#).

As shown in the figure below, and support the basic harvesting of resources with Gleaner and loading the JSON-LD objects into Minio.

It does not result in these objects ending up in a graph / triplestore. You would use this option if you intend to work on the JSON-LD objects yourself. Perhaps loading them into a alternative graphdb like Janus or working on them with python tooling.

Gleaner Indexing Services (IS) Environment Variables The Docker Compose file used to launch the Gleaner IS has a set of configurable elements that can be set and passed to the orchestration system with environment variables.

These can be set manually or through the command line. A simple script to set the variables could look like:

```
#!/bin/bash

# domains
export GLEANER_ADMIN_DOMAIN=admin.local.dev
export GLEANER_OSS_DOMAIN=oss.local.dev
export GLEANER_GRAPH_DOMAIN=graph.local.dev
export GLEANER_WEB_DOMAIN=web.local.dev
export GLEANER_WEB2_DOMAIN=web2.local.dev

# Object store keys
export MINIO_ACCESS_KEY=worldsbestaccesskey
export MINIO_SECRET_KEY=worldsbestsecretkey

# local data volumes
export GLEANER_BASE=/tmp/gleaner/
export GLEANER_TRAEFIK=${GLEANER_BASE}/config
export GLEANER_OBJECTS=${GLEANER_BASE}/datavol/s3
export GLEANER_GRAPH=${GLEANER_BASE}/datavol/graph
```

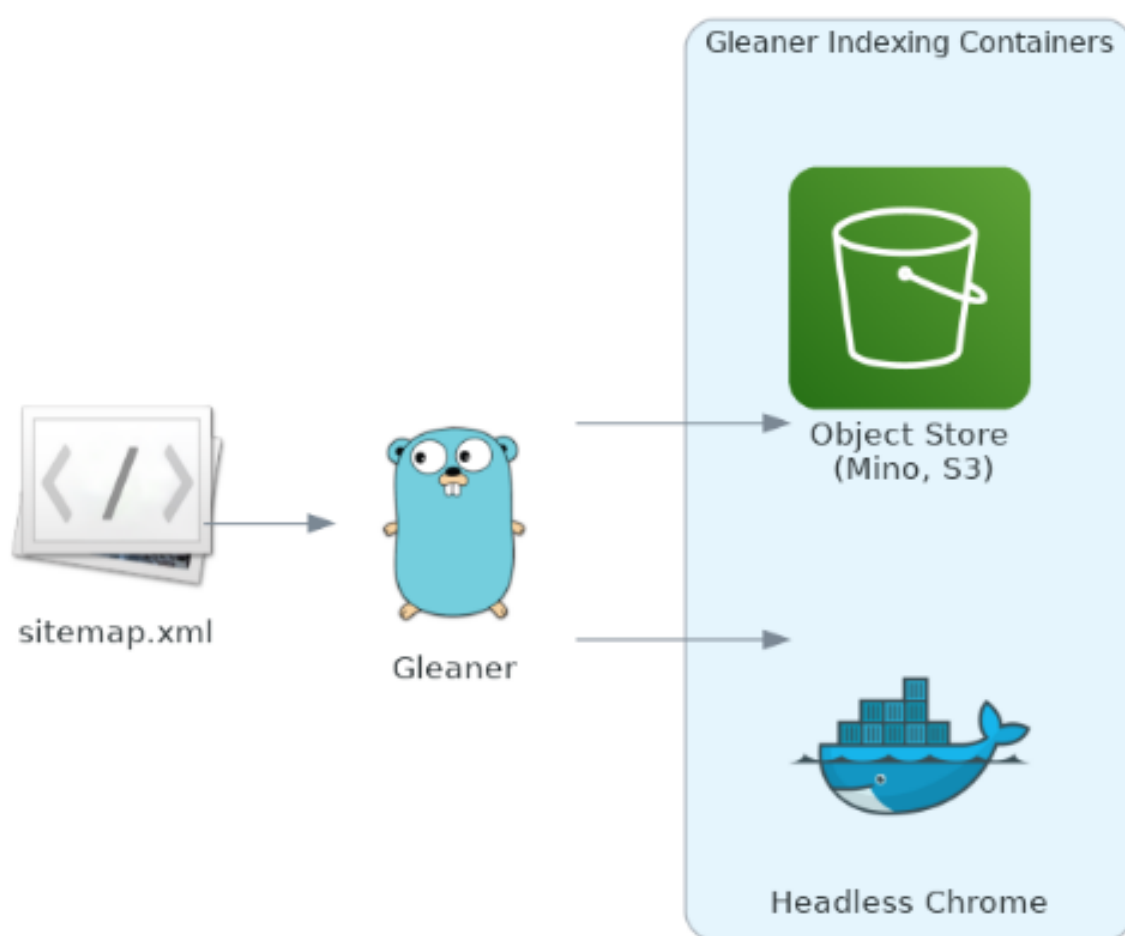


Fig. 1: Basic Gleaner Indexing Service Activity Workflow

The actual services can be deployed via a Docker Compose file (also works with Podman). An example of that file and details about it follow.

– Break down the compose files here link to them



## DATA SERVICES

The typical functional goal of this work is the development and use of a Graph that can be accessed via a triplestore (Graph Database). To do that we need a set of additional containers to support this and expose these services on the web through a single domain with https support.

- Object Store An S3 compliant object store supporting S3 APIs including S3Select. For open source this is best satisfied with the Minio Object Store. For commercial cloud AWS S3 or hosted Ceph services will work.
- Graph Database
- Web Router (technically optional)

### 21.1 Gleaner Data Services (DS)

If you wish to work with a triplestore and wish to use the default app used by OIH you can use the compose file that sets up the Gleaner Data Services environment.

This adds the Blazegraph triplestore to the configuration along with the object store.

The details of the OIH data services are found in the *Data Services* section.

Typically, a user would wish to run the full Gleaner DS stack which supports both the indexing process and the serving of the resulting data warehouse and graph database capacity.

Combined, these would then look like the following where the indexing and data services shared a common object store.

#### 21.1.1 Object store pattern

Within in the object store the following digital object pattern is used. This is based on the work of the RDA Digital Fabric working group.

At this point the graph and data warehouse (object store) can be exposed to the net for use by clients such as jupyter notebooks or direct client calls to the S3 object APIs and SPARQL endpoint.

Gleaner Data Services (DS) Environment Variables The Docker Compose file used to launch the Gleaner DS has a set of configurable elements that can be set and passed to the orchestration system with environment variables.

These can be set manually or through the command line. A simple script to set the variables could look like:

– Environment Var settings script

The actual services can be deployed via a Docker Compose file (also works with Podman). An example of that file and details about it follow.

Let's take a look at this.

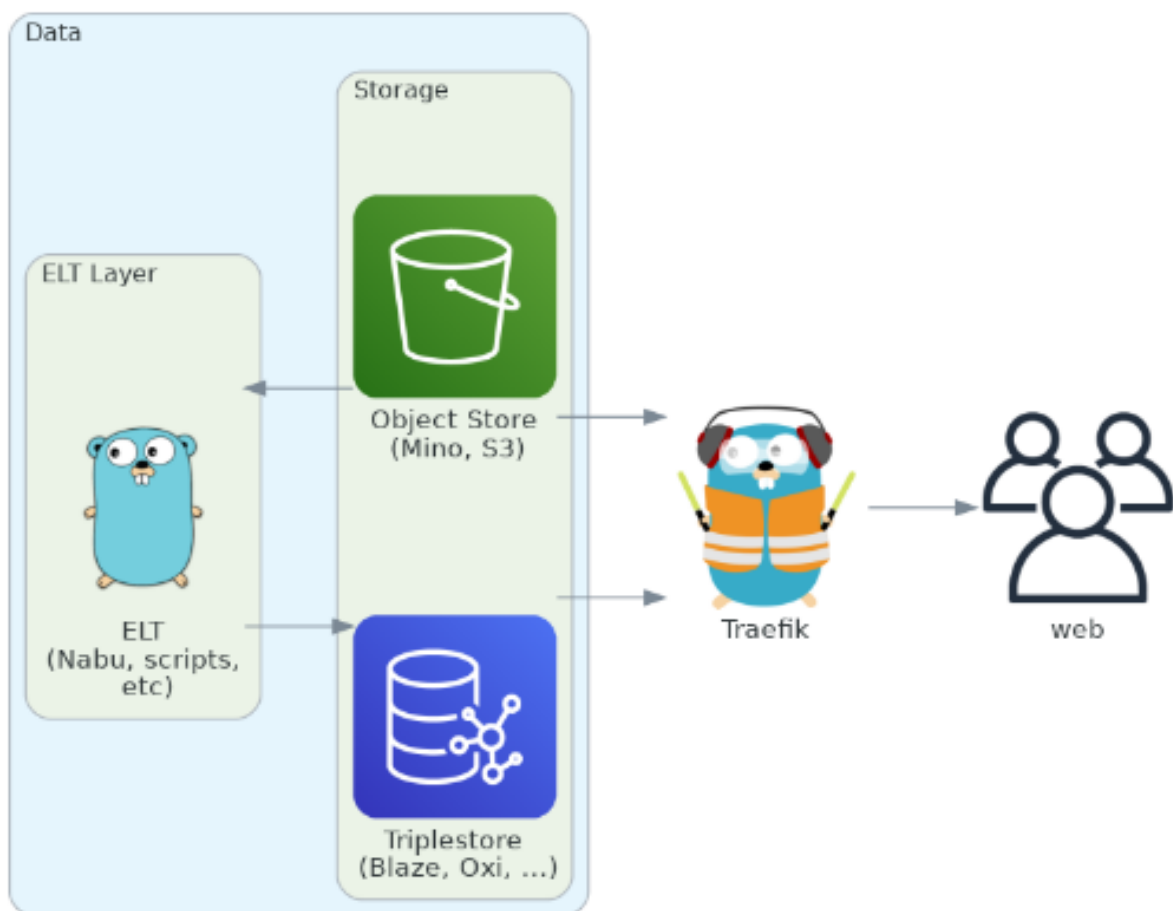


Fig. 1: Gleaner Data Service Activity Workflow

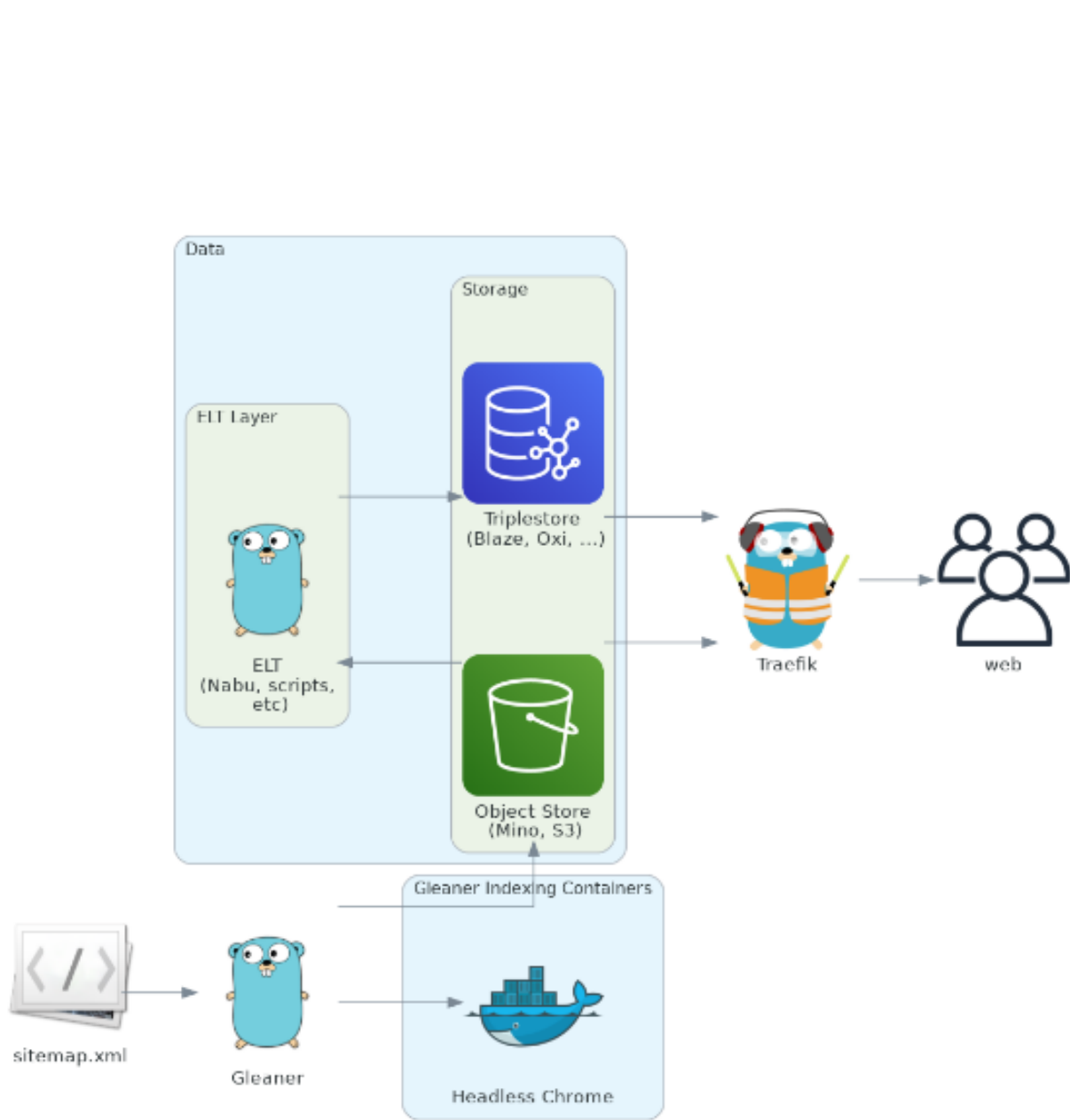


Fig. 2: Gleaner Indexing and Data Service Combined

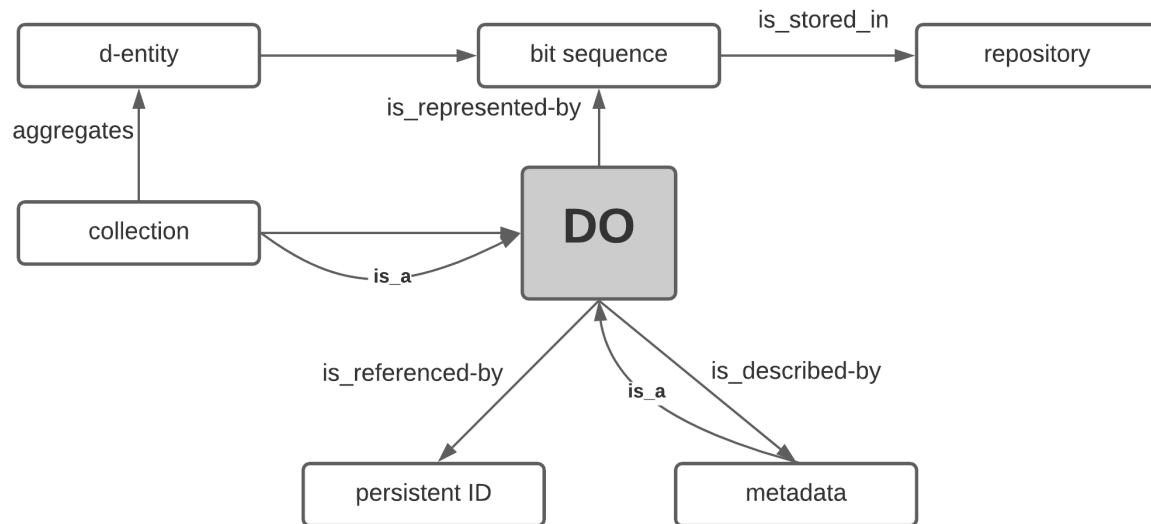


Figure 1: Research Data Alliance Digital Object Cloud Pattern

Fig. 3: Gleaner Digital Object Pattern

```

1  #!/bin/bash
2
3  # domains
4  export GLEANER_ADMIN_DOMAIN=admin.local.dev
5  export GLEANER_OSS_DOMAIN=oss.local.dev
6  export GLEANER_GRAPH_DOMAIN=graph.local.dev
7  export GLEANER_WEB_DOMAIN=web.local.dev
8  export GLEANER_WEB2_DOMAIN=web2.local.dev
9
10 # Object store keys
11 export MINIO_ACCESS_KEY=worldsbestaccesskey
12 export MINIO_SECRET_KEY=worldsbestsecretkey
13
14 # local data volumes
15 export GLEANER_BASE=/tmp/gleaner/
16 export GLEANER_TRAEFIK=${GLEANER_BASE}/config
17 export GLEANER_OBJECTS=${GLEANER_BASE}/datavol/s3
18 export GLEANER_GRAPH=${GLEANER_BASE}/datavol/graph
19

```

– Break down the compose file here

```

1  version: '3'
2
3  # ${GLEANER_ADMIN_DOMAIN}
4  # ${GLEANER_OSS_DOMAIN}
5  # ${GLEANER_GRAPH_DOMAIN}
6  # ${GLEANER_WEB_DOMAIN}
7  # ${GLEANER_WEB2_DOMAIN}

```

(continues on next page)



(continued from previous page)

```

8  # ${MINIO_ACCESS_KEY}
9  # ${MINIO_SECRET_KEY}
10 #
11 # ${GLEANER_TRAEFIK}
12 # ${GLEANER_OBJECTS}
13 # ${GLEANER_GRAPH}
14
15 services:
16   triplestore:
17     image: nower/blazegraph
18     environment:
19       JAVA_XMS: 2g
20       JAVA_XMX: 8g
21       JAVA_OPTS: -Xmx6g -Xms2g --XX:+UseG1GC
22     ports:
23       - 9999:9999
24     labels:
25       - "traefik.enable=true"
26       - "traefik.http.routers.triplestore.entrypoints=http"
27       - "traefik.http.routers.triplestore.rule=Host(`${GLEANER_GRAPH_DOMAIN}`)"
28       - "traefik.http.middlewares.triplestore-https-redirect.redirectscheme.
↪scheme=https"
29       - "traefik.http.routers.triplestore.middlewares=triplestore-https-redirect"
30       - "traefik.http.routers.triplestore-secure.entrypoints=https"
31       - "traefik.http.routers.triplestore-secure.rule=Host(`${GLEANER_GRAPH_DOMAIN}`)"
32       - "traefik.http.routers.triplestore-secure.tls=true"
33       - "traefik.http.routers.triplestore-secure.tls.certresolver=http"
34       - "traefik.http.routers.triplestore-secure.service=triplestore"
35       - "traefik.http.middlewares.triplestore-secure.headers.
↪accesscontrolallowmethods=GET,OPTIONS,PUT,POST"
36       - "traefik.http.middlewares.triplestore-secure.headers.
↪accesscontrolalloworigin=*"
37       - "traefik.http.middlewares.triplestore-secure.headers.accesscontrolmaxage=200"
38       - "traefik.http.middlewares.triplestore-secure.headers.addvaryheader=true"
39       - "traefik.http.middlewares.triplestore-secure.headers.
↪accesscontrolallowcredentials=true"
40       - "traefik.http.middlewares.triplestore-secure.headers.
↪accesscontrolallowheaders=Authorization,Origin,Content-Type,Accept"
41       - "traefik.http.middlewares.triplestore-secure.headers.customresponseheaders.
↪Access-Control-Allow-Headers=Authorization,Origin,Content-Type,Accept"
42       - "traefik.http.routers.triplestore-secure.middlewares=triplestore-secure@docker
↪"
43       - "traefik.http.services.triplestore.loadbalancer.server.port=9999"
44       - "traefik.docker.network=traefik_default"
45     volumes:
46       - ${GLEANER_GRAPH}:/var/lib/blazegraph
47     networks:
48       - traefik_default
49
50   s3system:
51     image: minio/minio:latest
52     ports:
53       - 9000:9000
54     labels:
55       - "traefik.enable=true"
56       - "traefik.http.routers.s3system.entrypoints=http"
57       - "traefik.http.routers.s3system.rule=Host(`${GLEANER_OSS_DOMAIN}`)"

```

(continues on next page)

(continued from previous page)

```

58     - "traefik.http.middlewares.s3system-https-redirect.redirectscheme.scheme=https"
59     - "traefik.http.routers.s3system.middlewares=s3system-https-redirect"
60     - "traefik.http.routers.s3system-secure.entrypoints=https"
61     - "traefik.http.routers.s3system-secure.rule=Host(`${GLEANER_OSS_DOMAIN}`)"
62     - "traefik.http.routers.s3system-secure.tls=true"
63     - "traefik.http.routers.s3system-secure.tls.certresolver=http"
64     - "traefik.http.routers.s3system-secure.service=s3system"
65     - "traefik.http.services.s3system.loadbalancer.server.port=9000"
66     - "traefik.docker.network=traefik_default"
67 volumes:
68     - ${GLEANER_OBJECTS}:/data
69 environment:
70     - MINIO_ACCESS_KEY=${MINIO_ACCESS_KEY}
71     - MINIO_SECRET_KEY=${MINIO_SECRET_KEY}
72 networks:
73     - traefik_default
74 command: ["server", "/data"]
75
76 features:
77     image: fils/grow-general:latest
78     ports:
79         - 8080:8080
80     environment:
81         - S3ADDRESS=s3system:9000
82         - S3BUCKET=sites
83         - S3PREFIX=domain
84         - DOMAIN=https://${GLEANER_WEB_DOMAIN}/
85         - S3KEY=${MINIO_ACCESS_KEY}
86         - S3SECRET=${MINIO_SECRET_KEY}
87     labels:
88         - "traefik.enable=true"
89         - "traefik.http.routers.features.entrypoints=http"
90         - "traefik.http.routers.features.rule=Host(`${GLEANER_WEB_DOMAIN}`,`${GLEANER_
↪WEB2_DOMAIN}`)"
91         - "traefik.http.middlewares.features-https-redirect.redirectscheme.scheme=https"
92         - "traefik.http.routers.features.middlewares=features-https-redirect"
93         - "traefik.http.routers.features-secure.entrypoints=https"
94         - "traefik.http.routers.features-secure.rule=Host(`${GLEANER_WEB_DOMAIN}`,`${
↪GLEANER_WEB2_DOMAIN}`)"
95         - "traefik.http.routers.features-secure.tls=true"
96         - "traefik.http.routers.features-secure.tls.certresolver=http"
97         - "traefik.http.routers.features-secure.service=features"
98         - "traefik.http.services.features.loadbalancer.server.port=8080"
99         - "traefik.docker.network=traefik_default"
100        - "traefik.http.middlewares.features.headers.accesscontrolallowmethods=GET,
↪OPTIONS,PUT,POST"
101        - "traefik.http.middlewares.features.headers.accesscontrolalloworigin=*"
102        - "traefik.http.middlewares.features.headers.accesscontrolmaxage=100"
103        - "traefik.http.middlewares.features.headers.addvaryheader=true"
104        - "traefik.http.middlewares.features-secure.headers.accesscontrolallowheaders=*"
105        - "traefik.http.middlewares.features-secure.headers.customresponseheaders.
↪Access-Control-Allow-Headers=*"
106    networks:
107        - traefik_default
108
109 networks:
110     traefik_default:

```

(continues on next page)

(continued from previous page)

111

NOTE: DS also needs the object -> graph sync (via Nabu) NOTE: Should also add in (here or to the side) the ELT local Data Lake to Data Warehouse path (ala CSDCO VaultWalker)

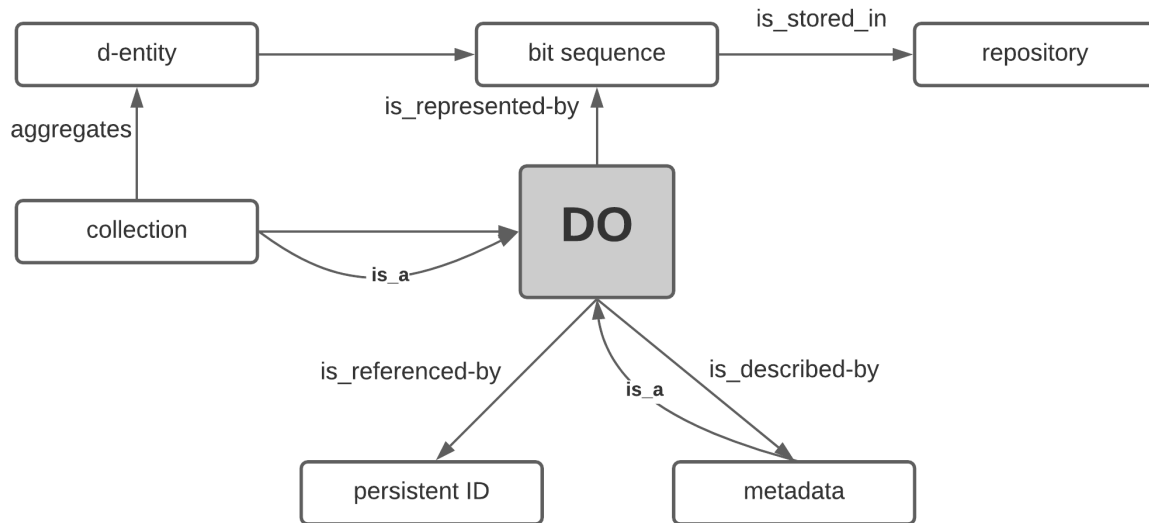


Figure 1: Research Data Alliance Digital Object Cloud Pattern



## INTERFACES

### 22.1 About

In the end the goal is to provide use of the generated index. There are several possible used for an index.

- Web UI such as the reference client at [oceans.collaborium.io](https://oceans.collaborium.io)
  - A variation on this is the development of web components that can be easily included in domain sites to perform operations on the OIH index
- graph access via SPARQL
- access to the graph and objects via workflows like Jupyter notebooks

### 22.2 Gleaner Web UI (WUI)

The user of the index may take several forms. A user may be a software developer creating a web based interface to the generated index. It may also be an end user accessing the index (indexes) through notebooks or special clients.

Those wishing to run a web site can augment the compose files to run their preferred web server, object server (to serve files from the object store) or software such as node or others to support their deployment pattern.

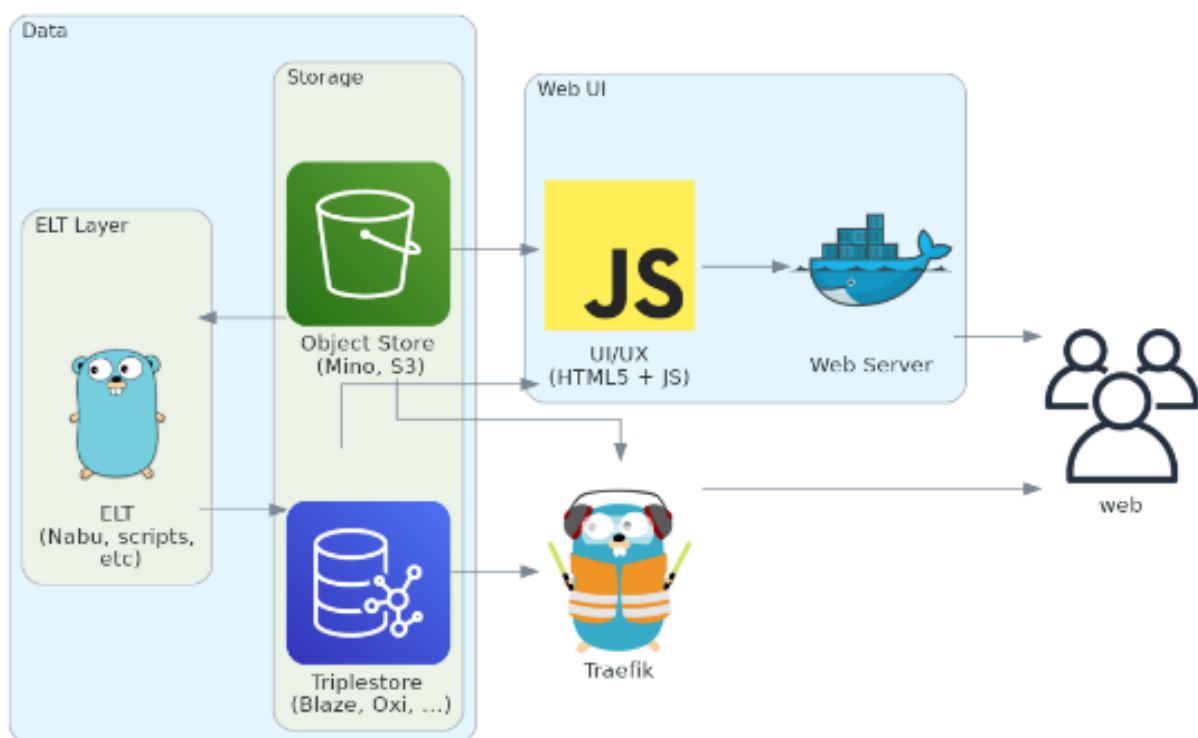


Fig. 1: Gleaner Optional Web UI

## GRAPH FIRST APPROACH

### 23.1 About

During the early adopters meetings and in discussion with others an alternative publication pattern came up. This is the pattern where it is not possible to update the web resources with the metadata content. This may be due to access or technical issues. Regardless, what was possible was to generate the metadata in bulk locally and make the resulting document available.

This approach is not ideal since it is a non-standard pattern and makes the data and information more obscure to other users. However, it is one the OIH architecture can adapt to and is preferable to the option of excluding those partners in this activity.

As such, we are making some changes to allow for this pattern. This means documenting the published graph structure based on the existing thematic patterns and some updates in the indexing workflow to obtain and integrate these graphs into the OIH graph.

**Warning:** Anti-pattern: Using the approach here is not in alignment with Google guidance nor with W3C patterns for structured data on the web.

It is documented here for edge cases where this is the minimum viable approach. The hope is it could act as a gateway to a more standards aligned implementation later.

### 23.2 Graph Only

There are cases where it is only possible to generate the graph based on the metadata. Access to the HTML pages is either difficult or the process of inserting the data into the pages is not supportable.

For this case the goal is to create a simple graph in JSON-LD. To do this we need a collection approach that is valid for a range of Things.

For this it is proposed to use ItemList which can be used on a list of type Thing, ie anything type in the [Schema.org](https://schema.org/) vocabulary.

This would define a ListItem with item of any type. Below is an example for a CreativeWork (map) and a Course. Once you are in a “item” any of the details from the other thematic type descriptions can be used.

```
{
  "@context": "https://schema.org/",
  "@type": ["ItemList", "CreativeWork"],
  "name": "Resource collection for site X",
  "author": "Creator of the list",
```

(continues on next page)

(continued from previous page)

```

"itemListOrder": "https://schema.org/ItemListUnordered",
"numberOfItems": 2,
"itemListElement": [
  {
    "@type": "ListItem",
    "item": {
      "@id": "ID_for_this_metadata_record1",
      "@type": "Map",
      "@id": "https://example.org/id/XYZ",
      "name": "Name or title of the document",
      "description": "Description of the map to aid in searching",
      "url": "https://www.sample-data-repository.org/creativework/map.pdf"
    }
  },
  {
    "@type": "ListItem",
    "item": {
      "@id": "ID_for_this_metadata_record2",
      "@type": "Course",
      "courseCode": "F300",
      "name": "Physics",
      "provider": {
        "@type": "CollegeOrUniversity",
        "name": "University of Bristol",
        "url": {
          "@id": "/provider/324/university-of-bristol"
        }
      }
    }
  }
]
}

```

In the case of schema:Dataset one might use schema:DataCatalogue for the following approach. However, since OIH is addressing a wide range of types a more generic collection of Things or CreativeWorks approach is needed.

## 23.3 Item Catalogue Page

It's not hard to generate a simple HTML page based on the structured metadata file. This doesn't alter the content of the graph, just builds an automated HTML page around it.

## 23.4 Publishing and referencing

## 23.5 Testing

Since we are now dealing with a graph that is pulled as a complete entity there are a few thoughts.

1. How do ensure a connection between a record in the list and a resolvable URL? Do we need to:
  1. ensure each record has a IRI it is subject of
  2. in the case where IRI is or can be URL, do a validation of at least a 200 on it



2. How do we publish this?
  1. entry in robots.txt (might be able due to reasons above?)
  2. published and provided to OIH
3. Need guidance on format and structure



## 24.1 About

This is the start of some discussion on issues around prov tracking in OIH. This may take two paths. One would be the prov tracking indexers might do and the other prov that providers would encode to provide specific prov the community requests.

## 24.2 Gleaner Prov

The Gleaner application generates a prov graph of the activity of accessing and indexing provider resources. The main goal of this prov is to connect an indexed URL to the digital object stored in the object store. This digital object should be the JSON-LD data graph presented by the provider.

By contrast, the authoritative reference in the various profiles will connect the the data graph ID, or in the absence of that the data graph URL or the referenced resources URL by gleaner, to another reference. This may be an organization ID or a PID of the connected resource.

```
1 {
2   "@context": {
3     "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
4     "prov": "http://www.w3.org/ns/prov#",
5     "rdfs": "http://www.w3.org/2000/01/rdf-schema#"
6   },
7   "@graph": [
8     {
9       "@id": "https://www.re3data.org/repository/obis",
10      "@type": "prov:Organization",
11      "rdf:name": "Ocean Biodiversity Information System",
12      "rdfs:seeAlso": "https://obis.org"
13    },
14    {
15      "@id": "https://obis.org/dataset/9381239f-3d64-48b4-80c9-b9ebb674edc2",
16      "@type": "prov:Entity",
17      "prov:wasAttributedTo": {
18        "@id": "https://www.re3data.org/repository/obis"
19      },
20      "prov:value": "https://obis.org/dataset/9381239f-3d64-48b4-80c9-
21      ↪b9ebb674edc2"
22    },
23    {
24      "@id": "https://gleaner.io/id/collection/
25      ↪7c1eaa1aaed95861330109026c42e57a31ecae55",
```

(continues on next page)

(continued from previous page)

```

24         "@type": "prov:Collection",
25         "prov:hadMember": {
26             "@id": "https://obis.org/dataset/9381239f-3d64-48b4-80c9-b9ebb674edc2"
27         }
28     },
29     {
30         "@id": "urn:gleaner:milled:obis:7c1eaa1aaed95861330109026c42e57a31ecae55",
31         "@type": "prov:Entity",
32         "prov:value": "7c1eaa1aaed95861330109026c42e57a31ecae55.jsonld"
33     },
34     {
35         "@id": "https://gleaner.io/id/run/7c1eaa1aaed95861330109026c42e57a31ecae55
36     ↪",
37         "@type": "prov:Activity",
38         "prov:endedAtTime": {
39             "@value": "2021-04-20",
40             "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
41         },
42         "prov:generated": {
43             "@id":
44 ↪"urn:gleaner:milled:obis:7c1eaa1aaed95861330109026c42e57a31ecae55"
45         },
46         "prov:used": {
47             "@id": "https://gleaner.io/id/collection/
48 ↪7c1eaa1aaed95861330109026c42e57a31ecae55"
49         }
50     }
51 ]
52 }

```

```

{
    "@context": {
        "@vocab": "https://schema.org/",
        "prov": "http://www.w3.org/ns/prov#"
    },
    "@id": "https://gleaner.io/id/run/7c1eaa1aaed95861330109026c42e57a31ecae55",
    "@type": "prov:Activity",
    "prov:endedAtTime": {
        "@type": "http://www.w3.org/2001/XMLSchema#dateTime",
        "@value": "2021-04-20"
    },
    "prov:generated": {
        "@id": "urn:gleaner:milled:obis:7c1eaa1aaed95861330109026c42e57a31ecae55",
        "@type": "prov:Entity",
        "prov:value": "7c1eaa1aaed95861330109026c42e57a31ecae55.jsonld"
    },
    "prov:used": {
        "@id": "https://gleaner.io/id/collection/
61 ↪7c1eaa1aaed95861330109026c42e57a31ecae55",
        "@type": "prov:Collection",
        "prov:hadMember": {
            "@id": "https://obis.org/dataset/9381239f-3d64-48b4-80c9-b9ebb674edc2",
            "@type": "prov:Entity",
            "prov:value": "https://obis.org/dataset/9381239f-3d64-48b4-80c9-
62 ↪b9ebb674edc2",
            "prov:wasAttributedTo": {

```

(continues on next page)

(continued from previous page)

```

        "@id": "https://www.re3data.org/repository/obis",
        "@type": "prov:Organization",
        "http://www.w3.org/1999/02/22-rdf-syntax-ns#name": "Ocean_
        Biodiversity Information System",
        "http://www.w3.org/2000/01/rdf-schema#seeAlso": "https://obis.org"
    }
}
}
}

```

## 24.3 Nano Prov

This is a basic nanoprov example. Note, this is a draft and the ID connections and examples have not been made yet.

```

1 {
2   "@context": {
3     "gleaner": "https://voc.gleaner.io/id/",
4     "np": "http://www.nanopub.org/nschema#",
5     "prov": "http://www.w3.org/ns/prov#",
6     "xsd": "http://www.w3.org/2001/XMLSchema#"
7   },
8   "@set": [
9     {
10      "@id": "gleaner:nanopub/XID",
11      "@type": "np:NanoPublication",
12      "np:hasAssertion": {
13        "@id": "gleaner:nanopub/XID#assertion"
14      },
15      "np:hasProvenance": {
16        "@id": "gleaner:nanopub/XID#provenance"
17      },
18      "np:hasPublicationInfo": {
19        "@id": "gleaner:nanopub/XID#pubInfo"
20      }
21    },
22    {
23      "@id": "gleaner:nanopub/XID#assertion",
24      "@graph": {
25        "@id": "DataSetURI",
26        "@type": "schema:Dataset",
27        "description": "This is where you would put corrections or annotations
28      ",
29        "identifier": [
30          {
31            "@type": "schema:PropertyValue",
32            "name": "GraphSHA",
33            "description": "A SHA256 sha stamp on the harvested data_
34      graph from a URL",
35            "value": "{SHA256 HASH HERE}"
36          },
37          {
38            "@type": "schema:PropertyValue",
39            "name": "ProviderID",
40            "description": "The id provided with the data graph by the_
41      provider",

```

(continues on next page)

(continued from previous page)

```

39         "value": "{{re3 or URL noted in config}}"
40     },
41     {
42         "@type": "schema:PropertyValue",
43         "name": "URL",
44         "description": "The URL harvested by gleaner",
45         "value": "{{The URL the JSON-LD came from}}"
46     }
47 ]
48 }
49 },
50 {
51     "@id": "gleaner:nanopub/XID#provenance",
52     "@graph": {
53         "@id": "URIforprovondataset",
54         "prov:wasGeneratedAtTime": {
55             "@value": "dateDone",
56             "@type": "xsd:dateTime"
57         },
58         "prov:wasDerivedFrom": {
59             "@id": "IDHERE"
60         },
61         "prov:wasAttributedTo": {
62             "@id": "IDHERE"
63         }
64     }
65 },
66 {
67     "@id": "gleaner:nanopub/XID#pubInfo",
68     "@graph": {
69         "@id": "IDHERE",
70         "prov:wasAttributedTo": {
71             "@id": "gleaner:tool/gleaner"
72         },
73         "prov:generatedAtTime": {
74             "@value": "2019-10-23T14:38:00Z",
75             "@type": "xsd:dateTime"
76         }
77     }
78 }
79 ]
80 }

```

```
<graphviz.dot.Digraph at 0x7f63d44c5850>
```

## 24.4 Refs

Nanopubs Guidance





## ALTERNATIVES

### 25.1 Options

While [Gleaner](#) will be used during initial OIH development it is not the only approach. There are many other tools that can be used and might be leveraged in a production environment including:

- [Extrunct](#)
- [BioSchemas Tools](#)
- [LDSpider](#)
- [Squirrel](#)
- [Nutch \(Apache\)](#)
- [Laundromat](#)
- [DataArchiver](#)
- [OD Archiver](#)

These different tools may better fit into the workflow and available skill sets for a group.



# **Part IV**

## **Tooling**



## 26.1 About

A collection of notes and material on tools.

## 26.2 OpenRefine

### 26.2.1 About

Some examples of using OpenRefine to generate a list of things. This is done as an example for those who may have a more manual workflow and wish to explore some tools to help automating that.

As the output from the template export in OpenRefine is an array it give us a chance to look at both [DataFeed](#) and [ItemList](#).

In the wonderfully open world of [Schema.org](#) thee is also [Series](#) and its subtypes of [CreativeWorkSeries](#) and [EventSeries](#).

Since the majority of what we work with are not data sets, we will focus on the ItemList and CreativeWorkSeries. The comparison to RSS is valid and casting to RSS is likely easy should it be desired.

Note, in the context of OIH this also raises the option of leveraging these approaches for the publishing and indexing of resources that align with this model.

### 26.2.2 Generic Template

#### About

The following are the templates sections for the OpenRefine export template command.

There are four sections

### PREFIX

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": ["ItemList", "CreativeWork"],
  "name": "Creative work list",
  "author": "Author of the list",
  "about": {
    "@type": "Course"
  },
  "itemListElement": [
```

### ROW TEMPLATE

```
{
  "@context": {
    "@vocab": "https://schema.org/",
    "endDate": {
      "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
    },
    "startDate": {
      "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
    }
  },
  "@id": {{jsonize(cells["ID"].value)}},
  "@type": {{jsonize(cells["type"].value)}},
  "description": {{jsonize(cells["description"].value)}},
  "name": {{jsonize(cells["name"].value)}},
  "hasCourseInstance": {
    "@type": {{jsonize(cells["CourseInstance"].value)}},
    "courseMode": {{jsonize(cells["courseMode"].value)}},
    "endDate": {{jsonize(cells["enddata"].value)}},
    "startDate": {{jsonize(cells["startdate"].value)}}
  },
  "provider": {
    "@type": "CollegeOrUniversity",
    "name": {{jsonize(cells["provider.name"].value)}},
    "url": {
      "@id": {{jsonize(cells["provider.url"].value)}}
    }
  }
}
```

## ROW SEP

```
,
```

## SUFFIX

```
]
}
```

## NOTES

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "Course",
  "courseCode": "F300",
  "name": "Physics",
  "provider": {
    "@type": "CollegeOrUniversity",
    "name": "University of Bristol",
    "url": {
      "@id": "/provider/324/university-of-bristol"
    }
  }
}

{
  "ID" : {{jsonize(cells["ID"].value)}},
  "type" : {{jsonize(cells["type"].value)}},
  "description" : {{jsonize(cells["description"].value)}},
  "name" : {{jsonize(cells["name"].value)}},
  "provider.name" : {{jsonize(cells["provider.name"].value)}},
  "provider.url" : {{jsonize(cells["provider.url"].value)}},
  "CourseInstance" : {{jsonize(cells["CourseInstance"].value)}},
  "courseMode" : {{jsonize(cells["courseMode"].value)}},
  "enddata" : {{jsonize(cells["enddata"].value)}},
  "startdate" : {{jsonize(cells["startdate"].value)}}
}
```

### 26.2.3 Sargassum Project Template

#### About

The following is a simple proof of concept page. It uses data from the [Sargassum Projects](#) page and is presented here simply as an example of this workflow.

If we visit the page referenced above, we can see some map interfaces. In the first, we are able to view and export the data in the map as a table. This will be downloaded as a CSV file. We can load this .csv file directly into [OpenRefine](#) and from there use the [templating exporter](#) to generate a valid JSON-LD document.

#### Templating Export

Prefix

```
{
  "@context": "https://schema.org",
  "@type": ["ItemList", "ResearchProject"],
  "name": "Sargassum Information Hub Projects",
  "author": "Sargassum Information Hub",
  "about": {
    "@type": "ResearchProject"
  },
  "itemListElement": [
```

Row Template

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "ResearchProject",
  "description": "{{jsonize(cells["Organization Description"]
  "name": "{{jsonize(cells["Organization name"].value}}",
  "url": "{{jsonize(cells["Sargassum Activity Website"].value)
  "geosparql:hasGeometry": {
    "@type": "http://www.opengis.net/ont/sf#Point",
    "geosparql:asWKT": {
      "@type": "http://www.opengis.net/ont/geosparql#wkt",
      "@value": "POINT({{y}} {{x}})"
    },
    "geosparql:crs": {
```

Row Separator

```
,
```

Suffix

```
]
}
```

Reset Template

Export

Cancel

```
{
  "@context": "https://schema.org",
  "@type": ["ItemList", "ResearchProject"],
  "name": "Sargassum Information Hub Projects",
  "author": "Sargassum Information Hub",
  "about": {
    "@type": "ResearchProject"
  },
  "itemListElement": [
    {
      "@context": {
        "@vocab": "https://schema.org/"
      },
      "@type": "ResearchProject",
      "description": "SME - Environmental studies, res",
      "name": null,
      "url": "http://www.novablue-environment.org",
      "geosparql:hasGeometry": {
        "@type": "http://www.opengis.net/ont/sf#Point",
        "geosparql:asWKT": {
          "@type": "http://www.opengis.net/ont/geosparql#wkt",
          "@value": "POINT(14.70476783255596 14.70476783255596)"
        },
        "geosparql:crs": {
          "@id": "http://www.opengis.net/def/crs/OGC/1.3/CRS84"
        }
      }
    }
  ],
  "@context": {
    "@vocab": "https://schema.org/"
  }
}
```

Fig. 1: A view of the template export in OpenRefine with the following sections inserted

The following are the templates sections for the OpenRefine export template command.

There are four sections



## PREFIX

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": ["ItemList", "ResearchProject"],
  "name": "Sargassum Information Hub Projects",
  "author": "Sargassum Information Hub",
  "about": {
    "@type": "ResearchProject"
  },
  "itemListElement": [
```

## ROW TEMPLATE

### Note

In the following row template the entries such as:

```
{{jsonize(cells["Organization Description"].value)}}
```

will present a value with quotes around it.

Where

```
${x}
```

will present the value from the noted colum, x, with no quotes around it.

When generating the JSON we may or may not need to include quotes depending on where we are in the serialization.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "ResearchProject",
  "description": {{jsonize(cells["Organization Description"].value)}},
  "name" : {{jsonize(cells["Organization name"].value)}},
  "url": {{jsonize(cells["Sargassum Activity Website"].value)}},
  "geosparql:hasGeometry": {
    "@type": "http://www.opengis.net/ont/sf#Point",
    "geosparql:asWKT": {
      "@type": "http://www.opengis.net/ont/geosparql#wktLiteral",
      "@value": "POINT(${y} ${y}) "
    },
  },
  "geosparql:crs": {
    "@id": "http://www.opengis.net/def/crs/OGC/1.3/CRS84"
  }
}
```

### ROW SEP

```
,
```

### SUFFIX

```
]
}
```

### NOTES

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "Course",
  "courseCode": "F300",
  "name": "Physics",
  "provider": {
    "@type": "CollegeOrUniversity",
    "name": "University of Bristol",
    "url": {
      "@id": "/provider/324/university-of-bristol"
    }
  }
}

{
  "ID" : {{jsonize(cells["ID"].value)}},
  "type" : {{jsonize(cells["type"].value)}},
  "description" : {{jsonize(cells["description"].value)}},
  "name" : {{jsonize(cells["name"].value)}},
  "provider.name" : {{jsonize(cells["provider.name"].value)}},
  "provider.url" : {{jsonize(cells["provider.url"].value)}},
  "CourseInstance" : {{jsonize(cells["CourseInstance"].value)}},
  "courseMode" : {{jsonize(cells["courseMode"].value)}},
  "enddata" : {{jsonize(cells["enddata"].value)}},
  "startdate" : {{jsonize(cells["startdate"].value)}}
}
```

## 26.3 OIH Notebooks

### 26.3.1 About

Notebooks related to the OIH project. These are early documents and they and this documentation will expand. These notebooks are developed and run using the Jupyter Data Science notebooks server (<https://hub.docker.com/r/jupyter/datascience-notebook>). They will likely run on Google Colab or Binder but additional installs may be required. We will attempt to document all requirements in the notebooks.

#### OIH Editor

This notebook will hopefully provide an introduction to JSON-LD and how it can be generated and worked with. Its goal is to give those comfortable with programming an introduction.

#### OIH Graphs

A notebook to load and test queries across one or more graphs

#### OIH Graph Analytics

Load and analyses a graph

### 26.3.2 Graph Analytics Exploration

- <https://stackoverflow.com/questions/39274216/visualize-an-rdflib-graph-in-python>
- [https://networkx.org/documentation/stable/reference/algorithms/link\\_analysis.html](https://networkx.org/documentation/stable/reference/algorithms/link_analysis.html)

#### Steps

- Pull in the OIH RDF graph and load

```
!pip -q install pydotplus
!pip -q install graphviz
# !pip -q install rdflib
# !pip install -q -e git+https://github.com/RDFLib/rdflib.git#egg=rdflib
!pip -q install pydotplus
!pip -q install mimesis
!pip -q install minio
!pip -q install s3fs
!pip -q install SPARQLWrapper
!pip -q install boto3
!pip -q install 'fsspec>=0.3.3'
!pip -q install rdflib
!pip -q install rdflib-jsonld
!pip -q install PyLD==2.0.2
```

```
ERROR: pip's dependency resolver does not currently take into account all the
↳packages that are installed. This behaviour is the source of the following
↳dependency conflicts.
aiobotocore 1.3.1 requires botocore<1.20.50,>=1.20.49, but you have botocore 1.20.102
↳which is incompatible.
```

```
import rdflib
from rdflib.extras.external_graph_libs import rdflib_to_networkx_multidigraph
from rdflib.extras.external_graph_libs import rdflib_to_networkx_digraph
import networkx as nx
import matplotlib.pyplot as plt
import gzip

with gzip.open('./oceanexperts_graph.nq.gz', 'rb') as f:
    file_content = f.read()

g = rdflib.Graph()
g.parse(data = file_content, format="nquads")

G = rdflib_to_networkx_digraph(g)
# G = rdflib_to_networkx_multidigraph(result)

# # Plot Networkx instance of RDF Graph
# pos = nx.spring_layout(G, scale=2)
# edge_labels = nx.get_edge_attributes(G, 'r')b
# #nx.draw_networkx_edge_labels(G, pos, labels=edge_labels)
# nx.draw_networkx_edge_labels(G, pos)
# nx.draw(G, with_labels=True)
```

```
pr = nx.pagerank(G,alpha=0.9)
# for key, value in pr.items():
#     print(key, ' : ', value)
```

```
import pandas as pd
prdf = pd.DataFrame.from_dict(pr, orient='index')
```

```
prdf.dtypes
```

```
0      float64
dtype: object
```

```
prdf.sort_values(by=0,ascending=False, inplace=True,)
prdf.head(20)
```

```

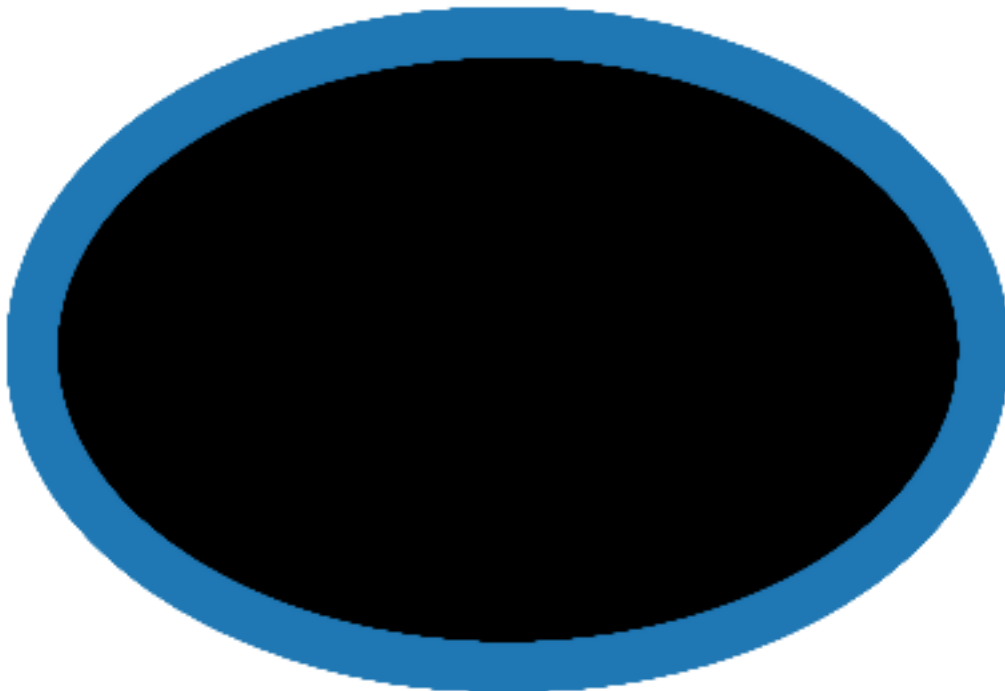
0
https://schema.org/Place      0.058482
https://schema.org/CourseInstance  0.018446
https://schema.org/Course      0.016571
UNESCO/IOC Project Office for IODE Wandelaarka...  0.007863
Russia      0.007497
UNESCO/IOC Project Office for IODE Wandelaarka...  0.006203
Wandelaarkaai 7 8400 Oostende Belgium      0.003761
Belgium      0.002931
RV Professor Logachev Russia      0.002624
UNESCO / IOC Project Office for IODE Wandelaar...  0.002360
```

(continues on next page)

(continued from previous page)

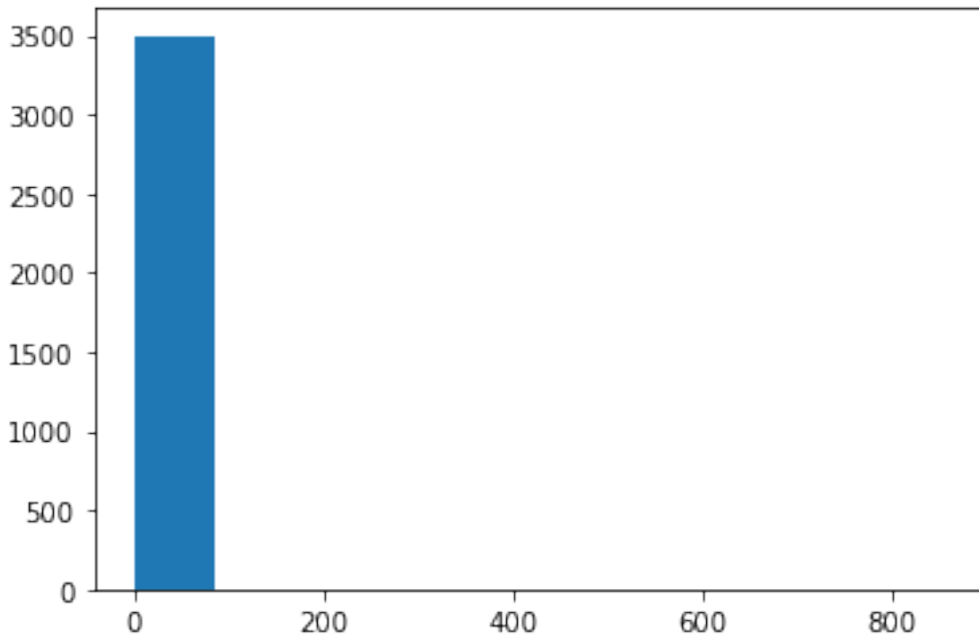
IOC Science and Communication Centre on Harmfu...	0.001830
Instituto de Investigaciones Marinas y Costera...	0.001812
"Ocean Valley", Pragathi Nagar (BO),...	0.001812
Kenya Marine and Fisheries Research Institute,...	0.001548
Calle 25 No. 2-55, Playa Salguero, Rodadero S...	0.001542
, Colombia	0.001271
Institute of Oceanography and Environment Univ...	0.001271
Australia	0.001271
Wandelaarkaai 7 Oostende Belgium	0.001018
Qingdao China	0.001018

```
nx.draw_circular(G, with_labels = False)
plt.show() # display
```



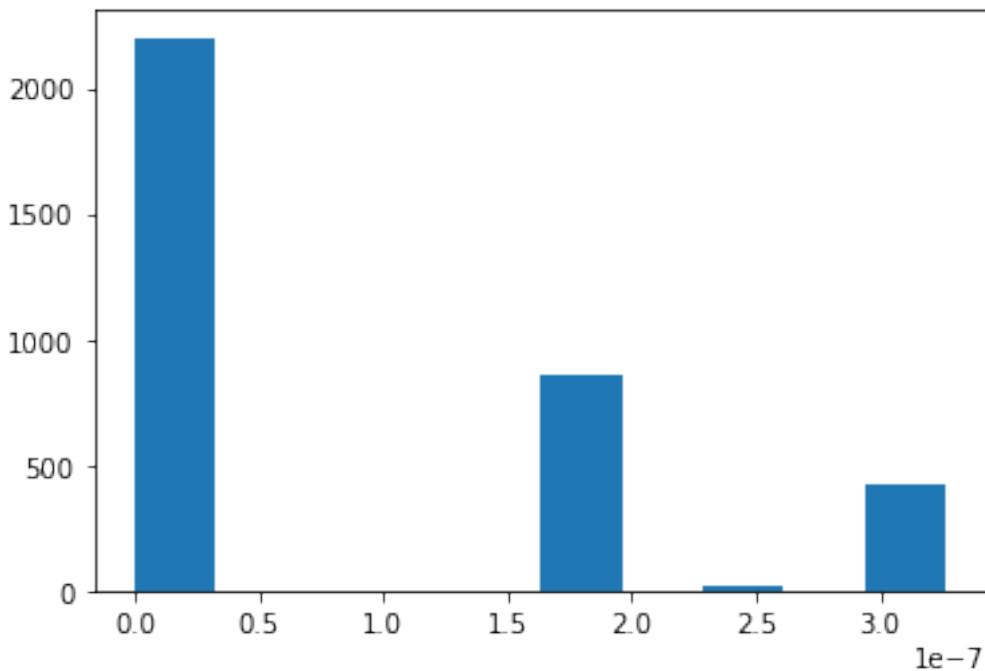
```
plt.hist([v for k,v in nx.degree(G)])
```

```
(array([3.499e+03, 0.000e+00, 0.000e+00, 0.000e+00, 0.000e+00, 2.000e+00,
        0.000e+00, 0.000e+00, 0.000e+00, 1.000e+00]),
 array([ 1. , 86.5, 172. , 257.5, 343. , 428.5, 514. , 599.5, 685. ,
        770.5, 856. ]),
 <BarContainer object of 10 artists>)
```



```
plt.hist(nx centrality.betweenness centrality(G).values())
```

```
(array([2200., 0., 0., 0., 0., 858., 0., 20., 0.,
        424.]),
 array([0.00000000e+00, 3.26437344e-08, 6.52874689e-08, 9.79312033e-08,
        1.30574938e-07, 1.63218672e-07, 1.95862407e-07, 2.28506141e-07,
        2.61149876e-07, 2.93793610e-07, 3.26437344e-07]),
 <BarContainer object of 10 artists>)
```



### 26.3.3 OIH Queries

```
import rdflib
import gzip

with gzip.open('./oceanexperts_graph.nq.gz', 'rb') as f:
    file_content = f.read()

g = rdflib.Graph()
g.parse(data = file_content, format="nquads")
```

```
<Graph identifier=N1bd41c55297e4e5fa4d8197321449083 (<class 'rdflib.graph.Graph'>)>
```

```
gres = g.query(
    """prefix schema: <https://schema.org/>
    SELECT DISTINCT ?s ?name
    WHERE {
        ?s a schema:Course .
        ?s schema:name ?name
    }
    LIMIT 10""")

for row in gres:
    print("%s Course Name: %s" % row)
```

```
Nada751c262da4aaea74ecfeecb456ff9 Course Name: Field training course &quot;Flysch
↳domain of the NW Rif belt, Morocco&quot;
N81febf0e07924f1d8ee16504e143c0d5 Course Name: INVEMAR: Carbono Azul y adaptación
↳basada en ecosistemas: rol de humedales marinos y estuarinos frente a Cambio
↳Climatico (énfasis manglares y pastos marinos)
N5a559f8fd2ad42149a19a42bc75d2347 Course Name: Educating about the Ocean: Waves of
↳change
N31b29ddb10b449a48080c6ae87d4d57f Course Name: GLOSS Training Course for Operators of
↳Sea Level Stations
N6d041d9804164a7e9f1cc376e7dc7a98 Course Name: Final ODINAFRICA II Training Course in
↳Marine Data Management
Nc54856a9d710405b801bdc5b47d8809f Course Name: II Baltic Floating University Cruise,
↳1994
N3c5f54e7273a4aeda1a83f0e0aab24eb Course Name: OTGA/INOS/UMT Training Course: Marine
↳GIS Applications for Coastal Zone Management
Ndcc18d7d7ca64bf683761d561bd677e5 Course Name: ITIC TRAINING PROGRAMME -
↳INTERNATIONAL (ITP-INTL) TSUNAMI EVACUATION MAPS, PLANS AND PROCEDURES (TEMPP):
↳TEMPP 1: INUNDATION MODELING AND MAPPING
N4132ad71d726482a9d9c946edd908142 Course Name: ITP-Fiji: Training on Improving
↳Tsunami Warning and Response using PTWS Enhanced Products
N59848a16a75645989193b25f48b36ad0 Course Name: OTGA/MAPCO/INVEMAR Training Course:
↳Áreas Marinas Protegidas (MPAs)
```

### Oceans UI test query

```
defaulttrq = """prefix prov: <http://www.w3.org/ns/prov#>
  PREFIX con: <http://www.ontotext.com/connectors/lucene#>
  PREFIX luc: <http://www.ontotext.com/owlim/lucene#>
  PREFIX con-inst: <http://www.ontotext.com/connectors/lucene/instance#>
  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
  PREFIX schema: <https://schema.org/>
  PREFIX schemaold: <http://schema.org/>
  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

  SELECT DISTINCT ?g ?s ?wat ?orgname ?domain ?type ?score ?name ?url ?lit ?
  ↳description ?headline
  WHERE
  {

    ?lit bds:search "coral" .
    ?lit bds:matchAllTerms "false" .
    ?lit bds:relevance ?score .
    graph ?g {
      ?s ?p ?lit .
      ?s rdf:type ?type .
      OPTIONAL { ?s schema:name ?name . }
      OPTIONAL { ?s schema:headline ?headline . }
      OPTIONAL { ?s schema:url ?url . }
      OPTIONAL { ?s schema:description ?description . }
    }
    ?sp prov:generated ?g .
    ?sp prov:used ?used .
    ?used prov:hadMember ?hm .
    ?hm prov:wasAttributedTo ?wat .
    ?wat rdf:name ?orgname .
    ?wat rdfs:seeAlso ?domain

  }
  ORDER BY DESC(?score)
  LIMIT 30
  OFFSET 0
  """
```

### AquaDocs Alignment Test Query

```
adrq = """prefix prov: <http://www.w3.org/ns/prov#>
  PREFIX con: <http://www.ontotext.com/connectors/lucene#>
  PREFIX luc: <http://www.ontotext.com/owlim/lucene#>
  PREFIX con-inst: <http://www.ontotext.com/connectors/lucene/instance#>
  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
  PREFIX schema: <https://schema.org/>
  PREFIX schemaold: <http://schema.org/>
  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

  SELECT DISTINCT ?g ?s ?wat ?orgname ?domain ?type ?score ?name ?url ?lit ?
  ↳description ?headline
  WHERE
```

(continues on next page)



(continued from previous page)

```
{

    ?lit bds:search "coral reef" .
    ?lit bds:matchAllTerms "false" .
    ?lit bds:relevance ?score .
    ?s ?p ?lit .

    graph ?g {
        ?s ?p ?lit .
        ?s rdf:type ?type .
        OPTIONAL { ?s schema:name ?name . }
        OPTIONAL { ?s schema:headline ?headline . }
        OPTIONAL { ?s schema:url ?url . }
        OPTIONAL { ?s schema:description ?description . }
    }

}

ORDER BY DESC(?score)
LIMIT 30
OFFSET 0
"""
```

## 26.3.4 OIH Editor Playground

For more background on the larger ODIS Ocean Info Hub and related referneces please visit out GitHub repository.

### Some package and imports

#### pip installs

```
##%quiet
#!pip install -q gastrodon
#!pip install -q ontospy
!pip install -q anytree
!pip install -q PyJSONViewer
!pip install -q ConceptNet
!pip install -q qwikidata
!pip install -q SPARQLWrapper
!pip install -q Wikidata
!pip install -q pySHACL
!pip install -q 'PyLD>=2.0.3'
!pip install -q rdflib
```

```
ERROR: Command errored out with exit status 1:
  command: /opt/conda/bin/python3.9 -u -c 'import io, os, sys, setuptools, tokenize;
  sys.argv[0] = '"'"'/tmp/pip-install-u2r_bu8s/marisa-trie_
  1d36873853dd40b6a8f9b670c86a37e7/setup.py'"'"'; __file__='"'"'/tmp/pip-install-u2r_
  bu8s/marisa-trie_1d36873853dd40b6a8f9b670c86a37e7/setup.py'"'"';f =
  getattr(tokenize, '"'"'open'"'"', open)(__file__) if os.path.exists(__file__) else
  io.StringIO('"'"'from setuptools import setup; setup()'"'"');code = f.read().
  replace('"'"'\r\n'"'"', '"'"'\n'"'"');f.close();exec(compile(code, __file__, '"'"'
  'exec'"'"'))' bdist_wheel -d /tmp/pip-wheel-6b4m_5t_
```

(continues on next page)

(continued from previous page)

```

cwd: /tmp/pip-install-u2r_bu8s/marisa-trie_1d36873853dd40b6a8f9b670c86a37e7/
Complete output (406 lines):
running bdist_wheel
running build
running build_clib
building 'libmarisa-trie' library
creating build
creating build/temp.linux-x86_64-3.9
creating build/temp.linux-x86_64-3.9/marisa-trie
creating build/temp.linux-x86_64-3.9/marisa-trie/lib
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/io
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/trie
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/vector
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/agent.cc -o build/temp.linux-x86_64-3.9/marisa-trie/lib/
↳marisa/agent.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/keyset.cc -o build/temp.linux-x86_64-3.9/marisa-trie/lib/
↳marisa/keyset.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/trie.cc -o build/temp.linux-x86_64-3.9/marisa-trie/lib/
↳marisa/trie.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/io/reader.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/reader.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/io/mapper.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/mapper.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/io/writer.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/writer.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/trie/tail.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/trie/tail.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/trie/louds-trie.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/trie/louds-trie.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/vector/bit-vector.cc -o build/temp.linux-x86_64-3.
↳9/marisa-trie/lib/marisa/grimoire/vector/bit-vector.o

```

(continued from previous page)

```

ar rcs build/temp.linux-x86_64-3.9/liblibmarisa-trie.a build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/agent.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/
↳keyset.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/trie.o build/temp.linux-
↳x86_64-3.9/marisa-trie/lib/marisa/grimoire/io/reader.o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/mapper.o build/temp.linux-x86_64-3.9/marisa-trie/
↳lib/marisa/grimoire/io/writer.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/
↳grimoire/trie/tail.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/
↳trie/louds-trie.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/
↳vector/bit-vector.o
running build_ext
building 'marisa_trie' extension
creating build/temp.linux-x86_64-3.9/src
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/agent.cpp -o build/temp.linux-x86_64-3.9/src/agent.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/base.cpp -o build/temp.linux-x86_64-3.9/src/base.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/iostream.cpp -o build/temp.linux-x86_64-3.9/src/iostream.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/key.cpp -o build/temp.linux-x86_64-3.9/src/key.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/keyset.cpp -o build/temp.linux-x86_64-3.9/src/keyset.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/marisa_trie.cpp -o build/temp.linux-x86_64-3.9/src/marisa_
↳trie.o
src/marisa_trie.cpp: In function 'int __Pyx_modinit_type_init_code()':
src/marisa_trie.cpp:17944:34: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
17944 |     __pyx_type_11marisa_trie__Trie.tp_print = 0;
      |                                     ^~~~~~
src/marisa_trie.cpp:17968:39: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
17968 |     __pyx_type_11marisa_trie_BinaryTrie.tp_print = 0;
      |                                     ^~~~~~
src/marisa_trie.cpp:17981:46: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
17981 |     __pyx_type_11marisa_trie__UnicodeKeyedTrie.tp_print = 0;
      |                                     ^~~~~~
src/marisa_trie.cpp:17995:33: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
17995 |     __pyx_type_11marisa_trie_Trie.tp_print = 0;
      |                                     ^~~~~~
src/marisa_trie.cpp:18014:38: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18014 |     __pyx_type_11marisa_trie_BytesTrie.tp_print = 0;

```

(continues on next page)

(continued from previous page)

```

|                                     ^~~~~~
src/marisa_trie.cpp:18039:40: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18039 |   __pyx_type_11marisa_trie__UnpackTrie.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18052:39: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18052 |   __pyx_type_11marisa_trie_RecordTrie.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18070:57: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18070 |   __pyx_type_11marisa_trie__pyx_scope_struct___init__.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18076:57: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18076 |   __pyx_type_11marisa_trie__pyx_scope_struct_1_genexpr.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18082:58: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18082 |   __pyx_type_11marisa_trie__pyx_scope_struct_2_iterkeys.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18088:63: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18088 |   __pyx_type_11marisa_trie__pyx_scope_struct_3_iter_prefixes.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18094:59: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18094 |   __pyx_type_11marisa_trie__pyx_scope_struct_4_iteritems.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18100:63: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18100 |   __pyx_type_11marisa_trie__pyx_scope_struct_5_iter_prefixes.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18106:59: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18106 |   __pyx_type_11marisa_trie__pyx_scope_struct_6_iteritems.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18112:58: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18112 |   __pyx_type_11marisa_trie__pyx_scope_struct_7___init__.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18118:57: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18118 |   __pyx_type_11marisa_trie__pyx_scope_struct_8_genexpr.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18124:59: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18124 |   __pyx_type_11marisa_trie__pyx_scope_struct_9_iteritems.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18130:59: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18130 |   __pyx_type_11marisa_trie__pyx_scope_struct_10_iterkeys.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18136:59: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18136 |   __pyx_type_11marisa_trie__pyx_scope_struct_11___init__.tp_print = 0;
|                                     ^~~~~~

```

(continues on next page)

(continued from previous page)

```

src/marisa_trie.cpp:18142:58: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18142 |   __pyx_type_11marisa_trie__pyx_scope_struct_12_genexpr.tp_print = 0;
      |                                     ^~~~~~
src/marisa_trie.cpp:18148:60: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18148 |   __pyx_type_11marisa_trie__pyx_scope_struct_13_iteritems.tp_print = 0;
      |                                     ^~~~~~
src/marisa_trie.cpp:18154:58: error: 'PyObject' {aka 'struct _typeobject'} has
↳no member named 'tp_print'
18154 |   __pyx_type_11marisa_trie__pyx_scope_struct_14_genexpr.tp_print = 0;
      |                                     ^~~~~~
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
                  from /opt/conda/include/python3.9/Python.h:97,
                  from src/marisa_trie.cpp:4:
src/marisa_trie.cpp: In function 'int __Pyx_ParseOptionalKeywords(PyObject*,
↳PyObject**, PyObject*, PyObject**, Py_ssize_t, const char*)':
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
      |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
261 |     PyUnicode_WSTR_LENGTH(op) :
      |     ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
      |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
261 |     PyUnicode_WSTR_LENGTH(op) :
      |     ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE*
↳PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |     ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
      |     ^
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(

```

(continues on next page)

(continued from previous page)

```

|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE*
PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
|                                     ^
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
|         ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
|         ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
|         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~

```

(continues on next page)

(continued from previous page)

```

/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
  451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
      |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
  261 |         PyUnicode_WSTR_LENGTH(op) :
      |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
  446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE*
↳PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
  262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
      |                                     ^
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
  580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE*
↳PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
  262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
      |                                     ^
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
  580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
  451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
      |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
  264 |         PyUnicode_WSTR_LENGTH(op))
      |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↳? 1 :
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
  446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
  451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
      |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'

```

(continues on next page)

(continued from previous page)

```

264 |         PyUnicode_WSTR_LENGTH(op))
    |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |             (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_SIZE(key))
↪? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↪PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
    |             ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↪macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
    |         ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |             (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↪PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
    |             ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↪macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
    |         ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |             (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE*
↪PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |             ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)), \
    |             ^
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |             (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE*
↪PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |             ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)), \
    |             ^
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |             (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :

```

(continues on next page)



(continued from previous page)

```

|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of _
macro 'PyUnicode_WSTR_LENGTH'
264 |     PyUnicode_WSTR_LENGTH(op))
|     ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
PyUnicode_GET_SIZE(key)) ? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of _
macro 'PyUnicode_WSTR_LENGTH'
264 |     PyUnicode_WSTR_LENGTH(op))
|     ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
PyUnicode_GET_SIZE(key)) ? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of _
macro 'PyUnicode_WSTR_LENGTH'
261 |     PyUnicode_WSTR_LENGTH(op) : \
|     ^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
PyUnicode_GET_SIZE(key)) ? 1 :
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
|                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of _
macro 'PyUnicode_WSTR_LENGTH'
261 |     PyUnicode_WSTR_LENGTH(op) : \
|     ^~~~~~

```

(continues on next page)

(continued from previous page)

```

src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |                                     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE* _
↳PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
    |                                     ^
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |                                     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_UNICODE* _
↳PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
    |                                     ^
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |                                     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
    |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of _
↳macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
    |         ^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |                                     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t _
↳PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_length((PyObject*)op)
    |                                     ^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of _
↳macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
    |         ^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |                                     (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
    |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here

```

(continues on next page)

(continued from previous page)

```

446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |                               ^~~~~~
src/marisa_trie.cpp: In function 'PyObject* __Pyx_PyUnicode_Substring(PyObject*, Py_
↪ssize_t, Py_ssize_t)':
src/marisa_trie.cpp:20109:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
↪UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20109 |         return PyUnicode_FromUnicode(NULL, 0);
    |                               ^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
    |                 from /opt/conda/include/python3.9/Python.h:97,
    |                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
    |                               ^~~~~~
src/marisa_trie.cpp:20109:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
↪UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20109 |         return PyUnicode_FromUnicode(NULL, 0);
    |                               ^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
    |                 from /opt/conda/include/python3.9/Python.h:97,
    |                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
    |                               ^~~~~~
src/marisa_trie.cpp: In function 'PyObject* __Pyx_decode_c_string(const char*, Py_
↪ssize_t, Py_ssize_t, const char*, const char*, PyObject* (*)(const char*, Py_ssize_
↪t, const char*))':
src/marisa_trie.cpp:20142:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
↪UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20142 |         return PyUnicode_FromUnicode(NULL, 0);
    |                               ^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
    |                 from /opt/conda/include/python3.9/Python.h:97,
    |                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
    |                               ^~~~~~
src/marisa_trie.cpp:20142:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
↪UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20142 |         return PyUnicode_FromUnicode(NULL, 0);
    |                               ^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
    |                 from /opt/conda/include/python3.9/Python.h:97,
    |                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
    |                               ^~~~~~
error: command '/usr/bin/gcc' failed with exit code 1
-----
ERROR: Failed building wheel for marisa-trie
ERROR: Command errored out with exit status 1:
    command: /opt/conda/bin/python3.9 -u -c 'import io, os, sys, setuptools, _
↪tokenize; sys.argv[0] = '"/tmp/pip-install-u2r_bu8s/marisa-trie_
↪1d36873853dd40b6a8f9b670c86a37e7/setup.py'"; __file__='"/tmp/pip-install-u2r_
↪bu8s/marisa-trie_1d36873853dd40b6a8f9b670c86a37e7/setup.py'"; f = _
↪getattr(tokenize, 'open', open)(__file__) if os.path.exists(__file__) else _
↪io.StringIO(''''from setuptools import setup; setup()''); code = f.read().
↪replace(''\n'', '\n\n'); f.close(); exec(compile(code, __file__, 'exec'
↪'exec'))' install --record /tmp/pip-record-zgrqbimv/install-record.txt --single-
↪version-externally-managed --compile --install-headers /opt/conda/include/python3.9/
↪marisa-trie

```

(continued from previous page)

```

    cwd: /tmp/pip-install-u2r_bu8s/marisa-trie_1d36873853dd40b6a8f9b670c86a37e7/
Complete output (406 lines):
running install
running build
running build_clib
building 'libmarisa-trie' library
creating build
creating build/temp.linux-x86_64-3.9
creating build/temp.linux-x86_64-3.9/marisa-trie
creating build/temp.linux-x86_64-3.9/marisa-trie/lib
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/io
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/trie
creating build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/vector
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/agent.cc -o build/temp.linux-x86_64-3.9/marisa-trie/lib/
↳marisa/agent.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/keyset.cc -o build/temp.linux-x86_64-3.9/marisa-trie/lib/
↳marisa/keyset.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/trie.cc -o build/temp.linux-x86_64-3.9/marisa-trie/lib/
↳marisa/trie.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/io/reader.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/reader.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/io/mapper.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/mapper.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/io/writer.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/writer.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/trie/tail.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/trie/tail.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/trie/louds-trie.cc -o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/trie/louds-trie.o
gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/lib -Imarisa-trie/include -
↳c marisa-trie/lib/marisa/grimoire/vector/bit-vector.cc -o build/temp.linux-x86_64-3.
↳9/marisa-trie/lib/marisa/grimoire/vector/bit-vector.o

```

(continued from previous page)

```

ar rcs build/temp.linux-x86_64-3.9/liblibmarisa-trie.a build/temp.linux-x86_64-3.
↳9/marisa-trie/lib/marisa/agent.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/
↳keyset.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/trie.o build/temp.linux-
↳x86_64-3.9/marisa-trie/lib/marisa/grimoire/io/reader.o build/temp.linux-x86_64-3.9/
↳marisa-trie/lib/marisa/grimoire/io/mapper.o build/temp.linux-x86_64-3.9/marisa-trie/
↳lib/marisa/grimoire/io/writer.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/
↳grimoire/trie/tail.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/
↳trie/louds-trie.o build/temp.linux-x86_64-3.9/marisa-trie/lib/marisa/grimoire/
↳vector/bit-vector.o
    running build_ext
    building 'marisa_trie' extension
    creating build/temp.linux-x86_64-3.9/src
    gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/agent.cpp -o build/temp.linux-x86_64-3.9/src/agent.o
    gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/base.cpp -o build/temp.linux-x86_64-3.9/src/base.o
    gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/iostream.cpp -o build/temp.linux-x86_64-3.9/src/iostream.o
    gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/key.cpp -o build/temp.linux-x86_64-3.9/src/key.o
    gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/keyset.cpp -o build/temp.linux-x86_64-3.9/src/keyset.o
    gcc -pthread -B /opt/conda/compiler_compat -Wl,--sysroot=/ -Wno-unused-result -
↳Wsign-compare -DNDEBUG -fwrapv -O2 -Wall -fPIC -O2 -isystem /opt/conda/include -
↳fPIC -O2 -isystem /opt/conda/include -fPIC -Imarisa-trie/include -I/opt/conda/
↳include/python3.9 -c src/marisa_trie.cpp -o build/temp.linux-x86_64-3.9/src/marisa_
↳trie.o
    src/marisa_trie.cpp: In function 'int __Pyx_modinit_type_init_code()':
    src/marisa_trie.cpp:17944:34: error: 'PyObject' {aka 'struct _typeobject'}_
↳has no member named 'tp_print'
    17944 |     __pyx_type_11marisa_trie__Trie.tp_print = 0;
          |                                     ^~~~~~
    src/marisa_trie.cpp:17968:39: error: 'PyObject' {aka 'struct _typeobject'}_
↳has no member named 'tp_print'
    17968 |     __pyx_type_11marisa_trie_BinaryTrie.tp_print = 0;
          |                                     ^~~~~~
    src/marisa_trie.cpp:17981:46: error: 'PyObject' {aka 'struct _typeobject'}_
↳has no member named 'tp_print'
    17981 |     __pyx_type_11marisa_trie__UnicodeKeyedTrie.tp_print = 0;
          |                                     ^~~~~~
    src/marisa_trie.cpp:17995:33: error: 'PyObject' {aka 'struct _typeobject'}_
↳has no member named 'tp_print'
    17995 |     __pyx_type_11marisa_trie_Trie.tp_print = 0;
          |                                     ^~~~~~
    src/marisa_trie.cpp:18014:38: error: 'PyObject' {aka 'struct _typeobject'}_
↳has no member named 'tp_print'
    18014 |     __pyx_type_11marisa_trie_BytesTrie.tp_print = 0;

```

(continues on next page)

(continued from previous page)

```

|                                     ^~~~~~
src/marisa_trie.cpp:18039:40: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18039 |     __pyx_type_11marisa_trie__UnpackTrie.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18052:39: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18052 |     __pyx_type_11marisa_trie__RecordTrie.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18070:57: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18070 |     __pyx_type_11marisa_trie__pyx_scope_struct____init____.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18076:57: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18076 |     __pyx_type_11marisa_trie__pyx_scope_struct_1_genexpr.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18082:58: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18082 |     __pyx_type_11marisa_trie__pyx_scope_struct_2_iterkeys.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18088:63: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18088 |     __pyx_type_11marisa_trie__pyx_scope_struct_3_iter_prefixes.tp_print =
↳0;
|                                     ^~~~~~
src/marisa_trie.cpp:18094:59: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18094 |     __pyx_type_11marisa_trie__pyx_scope_struct_4_iteritems.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18100:63: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18100 |     __pyx_type_11marisa_trie__pyx_scope_struct_5_iter_prefixes.tp_print =
↳0;
|                                     ^~~~~~
src/marisa_trie.cpp:18106:59: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18106 |     __pyx_type_11marisa_trie__pyx_scope_struct_6_iteritems.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18112:58: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18112 |     __pyx_type_11marisa_trie__pyx_scope_struct_7____init____.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18118:57: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18118 |     __pyx_type_11marisa_trie__pyx_scope_struct_8_genexpr.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18124:59: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18124 |     __pyx_type_11marisa_trie__pyx_scope_struct_9_iteritems.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18130:59: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18130 |     __pyx_type_11marisa_trie__pyx_scope_struct_10_iterkeys.tp_print = 0;
|                                     ^~~~~~
src/marisa_trie.cpp:18136:59: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'

```

(continues on next page)

(continued from previous page)

```

18136 |   __pyx_type_11marisa_trie__pyx_scope_struct_11__init__ .tp_print = 0;
      |   ^~~~~~
src/marisa_trie.cpp:18142:58: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18142 |   __pyx_type_11marisa_trie__pyx_scope_struct_12_genexpr.tp_print = 0;
      |   ^~~~~~
src/marisa_trie.cpp:18148:60: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18148 |   __pyx_type_11marisa_trie__pyx_scope_struct_13_iteritems.tp_print = 0;
      |   ^~~~~~
src/marisa_trie.cpp:18154:58: error: 'PyObject' {aka 'struct _typeobject'}
↳has no member named 'tp_print'
18154 |   __pyx_type_11marisa_trie__pyx_scope_struct_14_genexpr.tp_print = 0;
      |   ^~~~~~
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
               from /opt/conda/include/python3.9/Python.h:97,
               from src/marisa_trie.cpp:4:
src/marisa_trie.cpp: In function 'int __Pyx_ParseOptionalKeywords(PyObject*,
↳PyObject**, PyObject*, PyObject**, Py_ssize_t, const char*)':
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↳_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↳length((PyObject*) op)
      |
↳^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
261 |   PyUnicode_WSTR_LENGTH(op) :
      |   ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |   (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
      |   ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |   ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↳_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↳length((PyObject*) op)
      |
↳^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
261 |   PyUnicode_WSTR_LENGTH(op) :
      |   ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |   (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
      |   ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |   ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↳UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |   ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)), \
      |   ^

```

(continues on next page)



(continued from previous page)

```

src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |                                     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↳UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)), \
      |                                     ^
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |                                     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↳_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↳length((PyObject*)op)
      |                                     ^
↳^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
      |         ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |                                     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↳_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↳length((PyObject*)op)
      |                                     ^
↳^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
↳macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
      |         ^~~~~~
src/marisa_trie.cpp:18940:22: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |                                     (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↳SIZE(key)) ? 1 :
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↳_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↳length((PyObject*)op)
      |                                     ^
↳^

```

(continues on next page)



(continued from previous page)

```

/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of ↵
↵macro 'PyUnicode_WSTR_LENGTH'
    261 |         PyUnicode_WSTR_LENGTH(op) :                \
        |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
    18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↵SIZE(key)) ? 1 :
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
    446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↵_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
    451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↵length((PyObject*)op)
        |
↵^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of ↵
↵macro 'PyUnicode_WSTR_LENGTH'
    261 |         PyUnicode_WSTR_LENGTH(op) :                \
        |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
    18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↵SIZE(key)) ? 1 :
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
    446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↵UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
    262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)), \
        |         ^
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
    18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↵SIZE(key)) ? 1 :
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
    580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↵UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
    262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)), \
        |         ^
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
    18940 |         (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↵SIZE(key)) ? 1 :
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
    580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
        |         ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↵_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
    451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↵length((PyObject*)op)
        |
↵^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of ↵
↵macro 'PyUnicode_WSTR_LENGTH'

```

(continues on next page)

(continued from previous page)

```

264 |         PyUnicode_WSTR_LENGTH(op))
    |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |             (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*) op)
    |
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of_
↪macro 'PyUnicode_WSTR_LENGTH'
264 |         PyUnicode_WSTR_LENGTH(op))
    |         ^~~~~~
src/marisa_trie.cpp:18940:52: note: in expansion of macro 'PyUnicode_GET_SIZE'
18940 |             (PyUnicode_GET_SIZE(**name) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*) op)
    |
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of_
↪macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
    |         ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |             (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
    |             ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*) op)
    |
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of_
↪macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
    |         ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |             (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
    |             ^~~~~~

```

(continues on next page)

(continued from previous page)

```

/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
  446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↪UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
  262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
      |                                   ^
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
  18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
  580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↪UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
  262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
      |                                   ^
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
  18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
  580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
  451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*)op)
      |
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of_
↪macro 'PyUnicode_WSTR_LENGTH'
  264 |         PyUnicode_WSTR_LENGTH(op))
      |         ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
  18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
  446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t_
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
  451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*)op)
      |
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of_
↪macro 'PyUnicode_WSTR_LENGTH'
  264 |         PyUnicode_WSTR_LENGTH(op))
      |         ^~~~~~
src/marisa_trie.cpp:18956:26: note: in expansion of macro 'PyUnicode_GET_SIZE'
  18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
      |                               ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here

```

(continues on next page)

(continued from previous page)

```

446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*)op)
|                                     ^~~~~~
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↪macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
|         ^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
|                                     ^~~~~~
↪~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
↪_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
↪length((PyObject*)op)
|                                     ^~~~~~
↪^
/opt/conda/include/python3.9/cpython/unicodeobject.h:261:7: note: in expansion of
↪macro 'PyUnicode_WSTR_LENGTH'
261 |         PyUnicode_WSTR_LENGTH(op) :
|         ^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
|                                     ^~~~~~
↪~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↪UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
|                                     ^
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :
|                                     ^~~~~~
↪~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
|                                     ^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:262:52: warning: 'Py_
↪UNICODE* PyUnicode_AsUnicode(PyObject*)' is deprecated [-Wdeprecated-declarations]
262 |         ((void)PyUnicode_AsUnicode(_PyObject_CAST(op)),\
|                                     ^
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 |         (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
↪SIZE(key)) ? 1 :

```

(continues on next page)

(continued from previous page)

```

|
^~~~~~
~
/opt/conda/include/python3.9/cpython/unicodeobject.h:580:45: note: declared here
580 | Py_DEPRECATED(3.3) PyAPI_FUNC(Py_UNICODE *) PyUnicode_AsUnicode(
|
^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
length((PyObject*)op)
|
^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
macro 'PyUnicode_WSTR_LENGTH'
264 | PyUnicode_WSTR_LENGTH(op))
|
^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 | (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
SIZE(key)) ? 1 :
|
^~~~~~
~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|
^~~~~~
/opt/conda/include/python3.9/cpython/unicodeobject.h:451:75: warning: 'Py_ssize_t
_PyUnicode_get_wstr_length(PyObject*)' is deprecated [-Wdeprecated-declarations]
451 | #define PyUnicode_WSTR_LENGTH(op) _PyUnicode_get_wstr_
length((PyObject*)op)
|
^
/opt/conda/include/python3.9/cpython/unicodeobject.h:264:8: note: in expansion of
macro 'PyUnicode_WSTR_LENGTH'
264 | PyUnicode_WSTR_LENGTH(op))
|
^~~~~~
src/marisa_trie.cpp:18956:59: note: in expansion of macro 'PyUnicode_GET_SIZE'
18956 | (PyUnicode_GET_SIZE(**argname) != PyUnicode_GET_
SIZE(key)) ? 1 :
|
^~~~~~
~
/opt/conda/include/python3.9/cpython/unicodeobject.h:446:26: note: declared here
446 | static inline Py_ssize_t _PyUnicode_get_wstr_length(PyObject *op) {
|
^~~~~~
src/marisa_trie.cpp: In function 'PyObject* __Pyx_PyUnicode_Substring(PyObject*,
Py_ssize_t, Py_ssize_t)':
src/marisa_trie.cpp:20109:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20109 | return PyUnicode_FromUnicode(NULL, 0);
|
^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
from /opt/conda/include/python3.9/Python.h:97,
from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
|
^~~~~~
src/marisa_trie.cpp:20109:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20109 | return PyUnicode_FromUnicode(NULL, 0);
|
^

```

(continues on next page)

(continued from previous page)

```

In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
                 from /opt/conda/include/python3.9/Python.h:97,
                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
 551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
      |                                     ^~~~~~
src/marisa_trie.cpp: In function 'PyObject* __Pyx_decode_c_string(const char*, Py_
↪ssize_t, Py_ssize_t, const char*, const char*, PyObject* (*)(const char*, Py_ssize_
↪t, const char*))':
src/marisa_trie.cpp:20142:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
↪UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20142 |         return PyUnicode_FromUnicode(NULL, 0);
      |                                     ^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
                 from /opt/conda/include/python3.9/Python.h:97,
                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
 551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
      |                                     ^~~~~~
src/marisa_trie.cpp:20142:45: warning: 'PyObject* PyUnicode_FromUnicode(const Py_
↪UNICODE*, Py_ssize_t)' is deprecated [-Wdeprecated-declarations]
20142 |         return PyUnicode_FromUnicode(NULL, 0);
      |                                     ^
In file included from /opt/conda/include/python3.9/unicodeobject.h:1026,
                 from /opt/conda/include/python3.9/Python.h:97,
                 from src/marisa_trie.cpp:4:
/opt/conda/include/python3.9/cpython/unicodeobject.h:551:42: note: declared here
 551 | Py_DEPRECATED(3.3) PyAPI_FUNC(PyObject*) PyUnicode_FromUnicode(
      |                                     ^~~~~~
error: command '/usr/bin/gcc' failed with exit code 1
-----
ERROR: Command errored out with exit status 1: /opt/conda/bin/python3.9 -u -c 'import_
↪io, os, sys, setuptools, tokenize; sys.argv[0] = ''''/tmp/pip-install-u2r_bu8s/
↪marisa-trie_1d36873853dd40b6a8f9b670c86a37e7/setup.py''''; __file__='''''/tmp/pip-
↪install-u2r_bu8s/marisa-trie_1d36873853dd40b6a8f9b670c86a37e7/setup.py''''; f =_
↪getattr(tokenize, ''''open''''', open)(__file__) if os.path.exists(__file__) else_
↪io.StringIO('''''from setuptools import setup; setup()''''');code = f.read().
↪replace('''''\r\n''''', '''''\n''''');f.close();exec(compile(code, __file__, ''''
↪'exec'''''))' install --record /tmp/pip-record-zgrqbimv/install-record.txt --single-
↪version-externally-managed --compile --install-headers /opt/conda/include/python3.9/
↪marisa-trie Check the logs for full command output.

```

## imports

```

#@title
# General imports
import json
import rdflib
import requests
from rdflib import Graph, plugin
from rdflib.serializer import Serializer
from bs4 import BeautifulSoup
import urllib.request

```

(continues on next page)

(continued from previous page)

```

from rdflib.extras.external_graph_libs import rdflib_to_networkx_multidigraph
from rdflib.extras.external_graph_libs import rdflib_to_networkx_graph
import networkx as nx
from networkx import Graph as NXGraph
import matplotlib.pyplot as plt
import statistics
import collections

# for JSON-LD work
from pyld import jsonld

# For SHACL
from pyshacl import validate

# concept net
import graphviz
from conceptnet5.uri import join_uri, split_uri
API_ROOT = 'http://api.conceptnet.io'

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-4-05965b657363> in <module>
    24
    25 # concept net
--> 26 import graphviz
    27 from conceptnet5.uri import join_uri, split_uri
    28 API_ROOT = 'http://api.conceptnet.io'

ModuleNotFoundError: No module named 'graphviz'

```

## functions

```

import graphviz
from conceptnet5.uri import join_uri, split_uri
API_ROOT = 'http://api.conceptnet.io'

def short_name(value, max_length=40):
    """
    Convert an RDF value (given as a dictionary) to a reasonable label.
    """
    if value['type'] == 'blank node':
        return '_'
    elif value['type'] == 'IRI':
        url = value['value']
        if '#' in url:
            # Show just the fragment of URLs with a fragment
            # (it's probably a property name)
            return url.split('#')[-1]

        # Give URLs relative to the root of our API
        if url.startswith(API_ROOT):
            short_url = url[len(API_ROOT):]

```

(continues on next page)

(continued from previous page)

```

        # If the URL is too long, hide it
        if len(short_url) > max_length:
            pieces = split_uri(short_url)
            return join_uri(pieces[0], '...')
        else:
            return short_url
    else:
        return url.split('/://')[ -1]
else:
    # Put literal values in quotes
    text = value['value'].replace(':', ' ')
    if len(text) > max_length:
        text = text[:max_length] + '...'
    return "{}{}".format(text)

def show_graph(url, size=10):
    """
    Show the graph structure of a ConceptNet API response.
    """
    rdf = jsonld.normalize(url)['@default']
    graph = graphviz.Digraph(
        strict=False, graph_attr={'size': str(size), 'rankdir': 'LR'}
    )
    for edge in rdf:
        subj = short_name(edge['subject'])
        obj = short_name(edge['object'])
        pred = short_name(edge['predicate'])
        if subj and obj and pred:
            # Apply different styles to the nodes based on whether they're
            # literals, ConceptNet URLs, or other URLs
            if obj.startswith('"'):
                # Literal values
                graph.node(obj, penwidth='0')
            elif obj.startswith('/'):
                # ConceptNet nodes
                graph.node(obj, style='filled', fillcolor="#ddeeff")
            else:
                # Other URLs
                graph.node(obj, color="#558855")
            graph.edge(subj, obj, label=pred)

    return graph

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-5-b348409f0077> in <module>
----> 1 import graphviz
      2 from conceptnet5.uri import join_uri, split_uri
      3 API_ROOT = 'http://api.conceptnet.io'
      4
      5

ModuleNotFoundError: No module named 'graphviz'

```



```
#@title
def get_sparql_dataframe(service, query):
    """
    Helper function to convert SPARQL results into a Pandas data frame.
    """
    sparql = SPARQLWrapper(service)
    sparql.setQuery(query)
    sparql.setReturnFormat(JSON)
    result = sparql.query()

    processed_results = json.load(result.response)
    cols = processed_results['head']['vars']

    out = []
    for row in processed_results['results']['bindings']:
        item = []
        for c in cols:
            item.append(row.get(c, {}).get('value'))
        out.append(item)

    return pd.DataFrame(out, columns=cols)
```

```
# Fetch a single <1MB file using the raw GitHub URL.
!curl --remote-name \
    --location https://raw.githubusercontent.com/ESIPFed/science-on-schema.org/
master/examples/dataset/full.jsonld
```

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current
			Dload Upload	Total	Spent	Left	Speed
100	9241	100	9241	0	0	48894	0

## Editor playground

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "CreativeWork",
  "@id": "https://example.org/id/XYZ",
  "name": "Name or title of the document",
  "description": "Description of the creative work to aid in searching",
  "url": "https://www.sample-data-repository.org/creativework/report.pdf"
}
```

```
{'@context': {'@vocab': 'https://schema.org/'},
 '@type': 'CreativeWork',
 '@id': 'https://example.org/id/XYZ',
 'name': 'Name or title of the document',
 'description': 'Description of the creative work to aid in searching',
 'url': 'https://www.sample-data-repository.org/creativework/report.pdf'}
```

### 26.3.5 Introduction to JSON-LD files

```
#@title Google Data Set Required+

name = 'Data Set Name one' #@param {type:"string"}
sdotype = 'http://schema.org/Dataset' #@param ["http://schema.org/Dataset", "http://
↪schema.org/DataCatalog"]
description = 'Descriptive text of the dataset.' #@param {type:"string"}
url = 'http://foo.org/data/distribution' #@param {type:"string"}
version = 'version' #@param {type:"string"}
license = 'CC-BY-4.0' #@param ["CC-BY-4.0", "CC-0"]
keywords = 'geochemistry, Earth System Modeling, climate change' #@param {type:"string"
↪"" }

print(name)
print(description)
```

```
Data Set Name one
Descriptive text of the dataset.
```

```
#@title
from pyld import jsonld
import json

doc = {}
doc["https://schema.org/name"] = name
doc["@type"] = sdotype
doc["@id"] = "http://cooldata.io/id/doc/1"
doc["https://schema.org/description"] = description
doc["https://schema.org/url"] = url
doc["https://schema.org/version"] = version
doc["https://schema.org/license"] = license

# parse comma seperated keywords, clean white spaces
k = keywords.split(",")
kp = []
for i in k:
    j = i.strip()
    kp.append(j)

doc["http://schema.org/keywords"] = kp

context = {
    "@vocab": "https://schema.org/",
}

# compact a document according to a particular context
# see: http://json-ld.org/spec/latest/json-ld/#compact-document-form
compactd = jsonld.compact(doc, context)

jd = json.dumps(compactd, indent=4)
print(jd)
```

```
{
```

(continues on next page)

(continued from previous page)

```

"@context": {
  "@vocab": "https://schema.org/"
},
"@id": "http://cooldata.io/id/doc/1",
"@type": "http://schema.org/Dataset",
"http://schema.org/keywords": [
  "geochemistry",
  "Earth System Modeling",
  "climate change"
],
"description": "Descriptive text of the dataset.",
"license": "CC-BY-4.0",
"name": "Data Set Name one",
"url": "http://foo.org/data/distribution",
"version": "version"
}

```

```
show_graph(doc)
```

```

-----
NameError                                Traceback (most recent call last)
<ipython-input-11-aef8eaa8fa4e> in <module>
----> 1 show_graph(doc)

NameError: name 'show_graph' is not defined

```

## Framing

Understanding Framing is not a first order concern. However, understanding it and what it does can help you to think about how your data graph will be used.

Let's make a frame that allows us to view only the elements of the JSON-LD data graph that we are interested in. In this case let's target the keywords.

```

urlf = "https://raw.githubusercontent.com/ESIPFed/science-on-schema.org/master/
examples/dataset/minimal.jsonld"

frame = {
  "@context": {"@vocab": "http://schema.org/"},
  "@explicit": "true",
  "@type": "Dataset",
  "keywords": "",
}

framed = jsonld.frame(doc, frame)
print(framed)

show_graph(framed)

```

```

{'@context': {'@vocab': 'http://schema.org/'}, '@id': 'http://cooldata.io/id/doc/1',
  '@type': 'Dataset', 'keywords': ['geochemistry', 'Earth System Modeling', 'climate_
change']}

```

```
<graphviz.dot.Digraph at 0x7fb564a17e80>
```

### Parse out the keywords

At this point we can now take out resulting JSON-LD graph and extract the items we are interested in.

```
#g = framed['@graph'] # get the graph
#kw = g[0]['keywords'] # get the keywords (you could do this in 1 line, 2 here for
↳exposition)

kw = framed['keywords'] # get the keywords (you could do this in 1 line, 2 here for
↳exposition)

print("We have the individual keywords, what shall we do with them?")
for w in kw:
    print(w)
```

```
We have the individual keywords, what shall we do with them?
geochemistry
Earth System Modeling
climate change
```

### Wikidata

So, we have used framing to arrive at a way to link to other graphs. A nice use of framing but perhaps more important is the linking. This exercise has shown how linking across graph gets us from strings to things.

Let's see if we can query wikidata and pull back some concepts. We will see if anything we use as a keyword aligns with a JSTOR (<https://www.jstor.org/>) topic.

Reference also: [https://www.wikidata.org/wiki/Wikidata:WikidataCon\\_2019/Program/Sessions/Lightning\\_talks\\_2](https://www.wikidata.org/wiki/Wikidata:WikidataCon_2019/Program/Sessions/Lightning_talks_2)

```
from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd
import json

dbsparql = "http://dbpedia.org/sparql"
ufokn = "http://graph.openknowledge.network/blazegraph/namespace/demo/sparql"
wikidata = "https://query.wikidata.org/sparql"
```

In the example below we can BIND in keywords like geochemistry but also other terms we may extract from the name, keywords and description.

```
rq = '''SELECT ?term ?topic WHERE {{ BIND ( "{var}" as ?term ) ?topic wdt:P3827 ?term
↳. }}''' .format(var="geochemistry")
test = get_sparql_dataframe(wikidata, rq)
test.head()
```

```
      term                                     topic
0  geochemistry  http://www.wikidata.org/entity/Q161764
```

```
#df = pd.DataFrame(columns=['term', 'topic'])

for w in range(len(kw)):
    print(kw[w])
    rq = '''SELECT ?term ?topic WHERE {{ BIND ( "{var}" as ?term ) ?topic wdt:P3827 ?
    term . }}''' .format(var=kw[w])
    sdf = get_sparql_dataframe(wikidata, rq)
    #df.append(sdf)
    x = sdf.head()
    print(x)
    print("-----")

#df.head()
```

```
geochemistry
      term
0 geochemistry http://www.wikidata.org/entity/Q161764
-----
Earth System Modeling
Empty DataFrame
Columns: [term, topic]
Index: []
-----
climate change
Empty DataFrame
Columns: [term, topic]
Index: []
-----
```

```
from wikidata.client import Client
from urllib.parse import urlparse

client = Client()

for row in test.iterrows():
    o = urlparse(row[1]['topic'])
    e = o.path.rsplit('/', 1)[-1]
    entity = client.get('Q161764', load=True)
    print('---\nTerm:{} \nURL:{} \ndescription:{}'.format(row[1]['term'], row[1][
    'topic'], entity.description))
```

```
---
Term:geochemistry
URL:http://www.wikidata.org/entity/Q161764
description:science that applies chemistry to geological systems
```

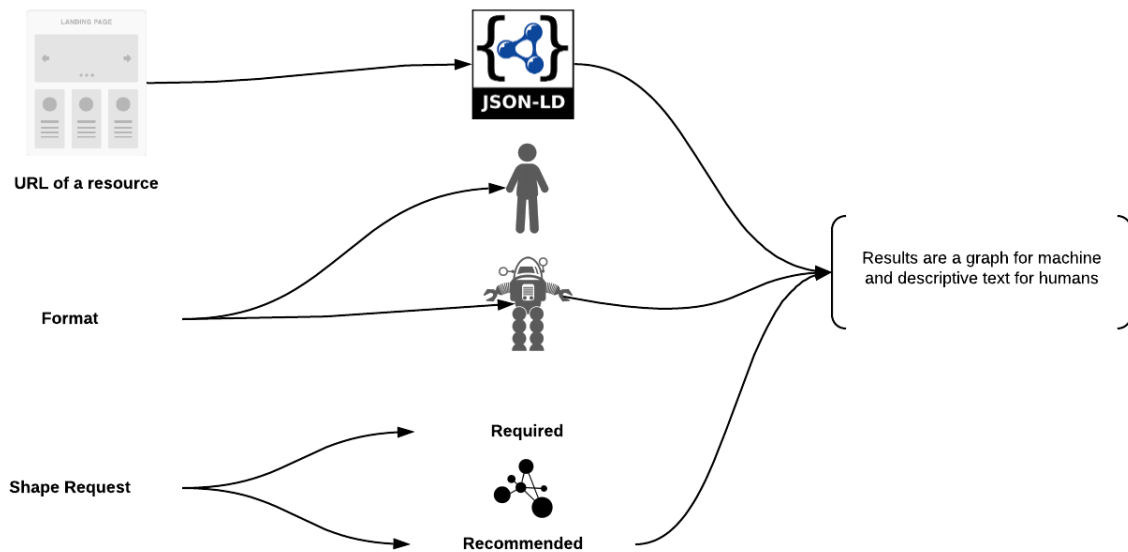
## 26.3.6 Validation

### About

The code below invokes pySHACL on some data and shape graphs out of GitHub. Note, we could edit these local to this notebook too. The human output is a bit hard to read since some of the encoding is off.

It might actually work to use the graph output and route it through the graph package and into Pandas too. It might let us parse and present the results a bit better.

The image below is just a test of putting images into this document. We can also upload and associate a document with the notebook and use it locally too.



```
from pyshacl import validate
import json
import rdflib
from rdflib.extras.external_graph_libs import rdflib_to_networkx_multidigraph
import requests
from rdflib import Graph, plugin
from rdflib.serializer import Serializer

from bs4 import BeautifulSoup
import urllib.request

dg = 'https://raw.githubusercontent.com/ESIPFed/science-on-schema.org/master/examples/
dataset/minimal.jsonld'
sg = 'https://raw.githubusercontent.com/geoschemas-org/geoshapes/master/shapegraphs/
googleRecommendedCoverageCheck.ttl'

s = rdflib.Graph()
sr = s.parse(sg, format="ttl")
d = rdflib.Graph()
dr = d.parse(dg, format="json-ld")

conforms, v_graph, v_text = validate(dr, shacl_graph=sr,
```

(continues on next page)

(continued from previous page)

```
data_graph_format="json-ld",
shacl_graph_format="ttl",
inference='none', debug=False,
serialize_report_graph=False)

print('{} {}'.format(conforms, v_text))
```

```
True Validation Report
Conforms: True
```

## 26.3.7 Thematic topic: Documents

### About

A testing area for the work on type Document

```
# Load examples from ODIS-Arch
!curl --remote-name \
      --location https://raw.githubusercontent.com/iodepo/odis-arch/master/schema/
      <thematics/docs/graphs/doc.json
```

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current
			Dload Upload	Total	Spent	Left	Speed
100	888	100	888	0	0	2439	0 --:--:-- --:--:-- --:--:-- 2439

```
# read into var

with open('/content/doc.json', 'r') as file:
    docstring = file.read()

docjson = json.loads(docstring)

# could I %load the file via magic commands? see all via %lsmagic
```

### Dev note

During development it would good to edit in the notebook and process the results through testing and viz. So we have to edit the JSON in the notebook, which is hideous.

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/XYZ",
  "@type": "Dataset",
  "description": "Description of the dataset to aid in searching",
  "distribution": {
    "@type": "DataDownload",
    "contentUrl": "https://www.sample-data-repository.org/dataset/472032.tsv",
    "encodingFormat": "text/tab-separated-values"
  },
}
```

(continues on next page)

(continued from previous page)

```

    "maintainer": {
      "@id": "https://link.to/PID_like_re3_or_others",
      "@type": "Organization",
      "description": "Organization or Person who maintains the creative work"
    },
    "name": "Name or title of the document",
    "subjectOf": {
      "@type": "DataDownload",
      "dateModified": "2019-06-12T14:44:15Z",
      "description": "EML metadata describing the dataset",
      "encodingFormat": [
        "application/xml",
        "https://eml.ecoinformatics.org/eml-2.2.0"
      ],
      "name": "eml-metadatafile.xml"
    }
  }
}

```

```

{ '@context': { '@vocab': 'https://schema.org/' },
  '@id': 'https://example.org/id/XYZ',
  '@type': 'Dataset',
  'description': 'Description of the dataset to aid in searching',
  'distribution': { '@type': 'DataDownload',
    'contentUrl': 'https://www.sample-data-repository.org/dataset/472032.tsv',
    'encodingFormat': 'text/tab-separated-values',
    'maintainer': { '@id': 'https://link.to/PID_like_re3_or_others',
      '@type': 'Organization',
      'description': 'Organization or Person who maintains the creative work',
      'name': 'Name or title of the document',
      'subjectOf': { '@type': 'DataDownload',
        'dateModified': '2019-06-12T14:44:15Z',
        'description': 'EML metadata describing the dataset',
        'encodingFormat': ['application/xml',
          'https://eml.ecoinformatics.org/eml-2.2.0'],
        'name': 'eml-metadatafile.xml' } } }
  }
}

```

```

context = {
  "@vocab": "https://schema.org/",
}

# compact a document according to a particular context
# see: http://json-ld.org/spec/latest/json-ld/#compacted-document-form
compacted = jsonld.compact(_, context) # CAUTION.. note _ which is reading the
↳previous cell output...

jd = json.dumps(compacted, indent=4)
print(jd)

```

```

{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@id": "https://example.org/id/XYZ",
  "@type": "Dataset",

```

(continues on next page)



(continued from previous page)

```

    "description": "Description of the dataset to aid in searching",
    "distribution": {
        "@type": "DataDownload",
        "contentUrl": "https://www.sample-data-repository.org/dataset/472032.tsv",
        "encodingFormat": "text/tab-separated-values"
    },
    "maintainer": {
        "@id": "https://link.to/PID_like_re3_or_others",
        "@type": "Organization",
        "description": "Organization or Person who maintains the creative work"
    },
    "name": "Name or title of the document",
    "subjectOf": {
        "@type": "DataDownload",
        "dateModified": "2019-06-12T14:44:15Z",
        "description": "EML metadata describing the dataset",
        "encodingFormat": [
            "application/xml",
            "https://eml.ecoinformatics.org/eml-2.2.0"
        ],
        "name": "eml-metadatafile.xml"
    }
}

```

```
show_graph(docjson)
```

```
<graphviz.dot.Digraph at 0x7fb558dd0a20>
```

## 26.3.8 SHACL validation

Ocean Info Hub SHACL validation on S3(minio) objects

### Flow

- get an object (use the dask notebook)
- process the object against OIH SHACL shapes

```

%%capture
!pip install pyshacl
!pip install 'PyLD>=2.0.3'
!pip install flatten_json
!pip install 'fsspec>=0.3.3'
!pip install s3fs
!pip install boto3
!pip install seaborn
!pip install dask

```

```

def label_status (row):
    result = row['http://www.w3.org/ns/shacl#resultSeverity']
    if result == "nan":
        return "NA"
    elif "Warning" in result:

```

(continues on next page)

(continued from previous page)

```

        return "Warning"
    elif "Violation" in result:
        return "Violation"
    else:
        return result

def source_shape (row):
    result = row['http://www.w3.org/ns/shacl#sourceShape']
    if type(result) is list:
        return result[0]['@id']
    else:
        return "NA"

```

## Gleaner Data

First lets load up some of the data Gleaner has collected. This is just simple data graph objects and not any graphs or other processed products from Gleaner.

```

# Set up our S3FileSystem object
import s3fs

oss = s3fs.S3FileSystem(
    anon=True,
    key="",
    secret="",
    client_kwargs = {"endpoint_url": "https://oss.collaborium.io"}
)

```

```

# Create the Dask tasks..  created..  not run..
import json
import dask, boto3
import dask.dataframe as dd

@dask.delayed()
def read_a_file(fn):
    # or preferably open in text mode and json.load from the file
    with oss.open(fn, 'rb') as f:
        #return json.loads(f.read().replace('\n', ' '))
        return json.loads(f.read().decode("ascii", "ignore").replace('\n', ' '))

# List of buckets to work with..  if you don't know them, you could print out above
buckets = ['gleaner/summoned/oceanexperts']
filenames = []

for d in range(len(buckets)):
    print("indexing {}".format(buckets[d]))
    f = oss.ls(buckets[d])
    filenames += f

#filenames = oss.cat('gleaner/summoned/opentopo', recursive=True)
output = [read_a_file(f) for f in filenames]
print(len(filenames))
# print(filenames)

```

```
indexing gleaner/summoned/oceanexperts
```

```
-----
ClientError                                Traceback (most recent call last)
/opt/conda/lib/python3.9/site-packages/s3fs/core.py in _lsdir(self, path, refresh,
    ↳ max_items, delimiter, prefix)
    571         dircache = []
--> 572         async for i in it:
    573             dircache.extend(i.get("CommonPrefixes", []))

/opt/conda/lib/python3.9/site-packages/aiobotocore/paginate.py in __anext__(self)
    31         while True:
---> 32             response = await self._make_request(current_kwargs)
    33             parsed = self._extract_parsed_response(response)

/opt/conda/lib/python3.9/site-packages/aiobotocore/client.py in _make_api_call(self,
    ↳ operation_name, api_params)
    153         error_class = self.exceptions.from_code(error_code)
--> 154         raise error_class(parsed_response, operation_name)
    155     else:

ClientError: An error occurred (AccessDenied) when calling the ListObjectsV2
    ↳ operation: Access Denied.
```

The above exception was the direct cause of the following exception:

```
PermissionError                            Traceback (most recent call last)
<ipython-input-11-775ae512ea02> in <module>
    17 for d in range(len(buckets)):
    18     print("indexing {}".format(buckets[d]))
--> 19     f = oss.ls(buckets[d])
    20     filenames += f
    21

/opt/conda/lib/python3.9/site-packages/fsspec/asyn.py in wrapper(*args, **kwargs)
    85     def wrapper(*args, **kwargs):
    86         self = obj or args[0]
--> 87         return sync(self.loop, func, *args, **kwargs)
    88
    89     return wrapper

/opt/conda/lib/python3.9/site-packages/fsspec/asyn.py in sync(loop, func, timeout,
    ↳ *args, **kwargs)
    66         raise FSTimeoutError
    67     if isinstance(result[0], BaseException):
--> 68         raise result[0]
    69     return result[0]
    70

/opt/conda/lib/python3.9/site-packages/fsspec/asyn.py in _runner(event, coro, result,
    ↳ timeout)
    22         coro = asyncio.wait_for(coro, timeout=timeout)
    23     try:
--> 24         result[0] = await coro
    25     except Exception as ex:
    26         result[0] = ex
```

(continues on next page)

(continued from previous page)

```

/opt/conda/lib/python3.9/site-packages/s3fs/core.py in _ls(self, path, detail, refresh)
    775         files = await self._lsbuckets(refresh)
    776     else:
--> 777         files = await self._lsdir(path, refresh)
    778         if not files and "/" in path:
    779             files = await self._lsdir(self._parent(path), refresh=refresh)

/opt/conda/lib/python3.9/site-packages/s3fs/core.py in _lsdir(self, path, refresh, max_items, delimiter, prefix)
    593         f["name"] = f["Key"]
    594     except ClientError as e:
--> 595         raise translate_boto_error(e)
    596
    597     if delimiter and files:

PermissionError: Access Denied.

```

```

%%time
from pyshacl import validate
from os import path
from pandas import json_normalize
import pandas as pd
import json
import rdflib
import seaborn as sns
import matplotlib.pyplot as plt

gldf = pd.DataFrame(columns=["id", "status", "shape"])

for ndx in range(len(output)):
    # for ndx in range(10):

        if "/.jsonld" not in filenames[ndx] :
            try:
                jld = output[ndx].compute()  ## Now pull from dask.. In REAL version, move
                # this logic into Dask! to get the parallel approach
            except:
                print(filenames[ndx])
                print("Doc has bad encoding")

        jd = json.dumps(jld, sort_keys=True, indent=4)

        try:
            conforms, v_graph, v_text = validate(jd,
                                                  shacl_graph='./oih_learning.ttl',
                                                  data_graph_format="json-ld",
                                                  shape_graph_format="ttl",
                                                  inference='none',
                                                  serialize_report_graph="json-ld")

            gd = v_graph.decode("ascii")
            df = pd.DataFrame(json.loads(gd))
            conforms = df["http://www.w3.org/ns/shacl#conforms"]
            tf = conforms[0][0]['@value']

```

(continues on next page)

(continued from previous page)

```

if "False" in str(tf):
    df['http://www.w3.org/ns/shacl#resultSeverity'] = df['http://www.w3.org/ns/
    shacl#resultSeverity'].astype(str)
    df['ID'] = filenames[ndx] # 'Object:{}'.format(ndx)
    df['Status'] = df.apply (lambda row: label_status(row), axis=1)
    df['Shape'] = df.apply (lambda row: source_shape(row), axis=1)

    data = [df["ID"], df["Status"], df['Shape']]
    headers = ["id", "status", "shape"]
    df3 = pd.concat(data, axis=1, keys=headers)
    gldf = gldf.append(df3, ignore_index=True)
elif "True" in str(tf):
    df['ID'] = filenames[ndx] # 'Object:{}'.format(ndx)
    df['Status'] = "Valid"
    df['Shape'] = "AllPassed"

    data = [df["ID"], df["Status"], df['Shape']]
    headers = ["id", "status", "shape"]
    df3 = pd.concat(data, axis=1, keys=headers)
    gldf = gldf.append(df3, ignore_index=True)

#     print("-----")
#     print(conforms)
#     print(v_graph)
#     print(v_text)

except:
    print("ERROR")
    df = pd.DataFrame()
    df['ID'] = filenames[ndx] # 'Object:{}'.format(ndx)
    df['Status'] = "ErrorProcessing"
    df['Shape'] = "ErrorProcessing"

    data = [df["ID"], df["Status"], df['Shape']]
    headers = ["id", "status", "shape"]
    df3 = pd.concat(data, axis=1, keys=headers)
    gldf = gldf.append(df3, ignore_index=True)
    print("PySHACL decode error: {}".format(filenames[ndx]))

```

```

gldf.info()
gldf.head(5)

```

```

-----
NameError                                Traceback (most recent call last)
<ipython-input-5-582109644996> in <module>
----> 1 gldf.info()
      2 gldf.head(5)

NameError: name 'gldf' is not defined

```

```

pd.value_counts(gldf['shape'])

```

```

-----
NameError                                Traceback (most recent call last)
<ipython-input-6-592d91c5781b> in <module>

```

(continues on next page)

(continued from previous page)

```
----> 1 pd.value_counts(gldf['shape'])
```

```
NameError: name 'pd' is not defined
```

```
pd.value_counts(gldf['shape']).plot.barh()
```

```
plt.xticks(rotation=45)
```

```
-----  
NameError                                Traceback (most recent call last)
```

```
<ipython-input-7-8967d9175060> in <module>
```

```
----> 1 pd.value_counts(gldf['shape']).plot.barh()
```

```
      2 plt.xticks(rotation=45)
```

```
NameError: name 'pd' is not defined
```

### 26.3.9 ISO 19139 To Schema.org

This work is an implementation of Steve Richard's work at: <https://github.com/usgin/metadataTransforms/tree/master/iso-19139-to-HTMLwSDO>

It demonstrates a simple transform from an ISO record to JSON-LD and [schema.org](https://schema.org). There are alternative paths to HTML + embedded JSON-LD that can be found in the examples directory of the repository.

Refs:

- <https://lxml.de/index.html>
- [https://www.seadatanet.org/content/download/4534/file/CDI\\_ISO19139\\_full\\_example\\_12.2.0.xml](https://www.seadatanet.org/content/download/4534/file/CDI_ISO19139_full_example_12.2.0.xml)
- <https://raw.githubusercontent.com/usgin/metadataTransforms/master/iso-19139-to-HTMLwSDO/ISO19139ToSchemaOrgDataset1.0.xslt>

```
!pip install -q lxml
```

```
import lxml.etree as ET
```

```
import urllib.request
```

```
!wget https://raw.githubusercontent.com/usgin/metadataTransforms/master/iso-19139-to-HTMLwSDO/ISO19139ToSDODatasetStandalone1.0.xslt
```

```
--2021-06-29 20:58:05-- https://raw.githubusercontent.com/usgin/metadataTransforms/  
master/iso-19139-to-HTMLwSDO/ISO19139ToSDODatasetStandalone1.0.xslt  
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133,  
185.199.111.133, 185.199.110.133, ...  
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.108.  
133|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 79137 (77K) [text/plain]  
Saving to: 'ISO19139ToSDODatasetStandalone1.0.xslt.1'  
  
ISO19139ToSDODataset 100%[=====] 77.28K --.-KB/s in 0.01s  
  
2021-06-29 20:58:05 (5.85 MB/s) - 'ISO19139ToSDODatasetStandalone1.0.xslt.1' saved  
[79137/79137]
```

```
dom = ET.parse("./CDI_ISO19139_full_example_12.2.0.xml")
xslt = ET.parse("./ISO19139ToSDODatasetStandalone1.0.xslt") ## convert to JSON-LD
↳schema.org voc
transform = ET.XSLT(xslt)
newdom = transform(dom)
```

```
print (newdom)
```

```
{
  "@context": {
    "@vocab": "http://schema.org/",
    "datacite": "http://purl.org/spar/datacite/",

    "earthcollab": "https://library.ucar.edu/earthcollab/schema#",

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

    "vivo": "http://vivoweb.org/ontology/core#",

    "dcat": "http://www.w3.org/ns/dcat#"

  },
  "@id": "urn:urnSDNCIDILOCALMARIS-TEST",
  "@type": "Dataset",
  "additionalType": [
    "geolink:Dataset",
    "vivo:Dataset"
  ],
  "name": "Test record with full coverage",
  "alternateName": "MARIS-TEST",
  "citation": "not provided, not provided, not provided, not provided, not provided",
  ↳(2012-04-16), Test record with full coverage, urn:urnSDNCIDILOCALMARIS-TEST.",
  "creator":
  [{
    "@type": "Role",
    "roleName": "originator",
    "creator": {

      "@type": "Role",
      "roleName": "originator"
    }
  },
  {
    "@type": "Role",
    "roleName": "originator",
    "creator": {

      "@type": "Role",
      "roleName": "originator"
    }
  },
  {
    "@type": "Role",
    "roleName": "originator",
    "creator": {

      "@type": "Role",

```

(continues on next page)

(continued from previous page)

```

        "roleName": "originator"
    },
    {
        "@type": "Role",
        "roleName": "originator",
        "creator": {
            "@type": "Role",
            "roleName": "originator"
        }
    },
    {
        "@type": "Role",
        "roleName": "originator",
        "creator": {
            "@type": "Role",
            "roleName": "originator"
        }
    }
],
    "datePublished": "2012-04-16",
    "description": "This record is meant for test purposes. It contains a value for
every field and multiple values wherever possible.",
    "distribution": [
        {
            "@id": "http://www.sdn-taskmanager.org/",
            "@type": "DataDownload",
            "additionalType": "dcat:distribution",
            "dcat:accessURL": "http://www.sdn-taskmanager.org/",
            "url": "http://www.sdn-taskmanager.org/",
            "description": "DBTEST. Service Protocol: DBTEST. Link Function:
downloadRegistration-- manual interaction with an on-line system by registered
users following successful authentication and authorisation. ",
            "provider": {
                "@type": "Role",
                "roleName": "distributor",
                "provider": {
                    "@type": "Role",
                    "roleName": "distributor"
                }
            }
        },
        {
            "fileFormat": [
                "Ocean Data View ASCII input v.0.3", "MEDATLAS ASCII v.1"
            ],
            "contentType": "text/plain"
        }
    ],
    {
        "@id": "http://geoservice.maris2.nl/wms/seadatanet/seadatanet/?",
        "@type": "DataDownload",
        "additionalType": "dcat:distribution",
        "dcat:accessURL": "http://geoservice.maris2.nl/wms/seadatanet/seadatanet/?",
        "url": "http://geoservice.maris2.nl/wms/seadatanet/seadatanet/?",
        "description": "WMS example url. Service Protocol: WMS example url. Link
Function: URL-- online resource locator for accessing data using a specific web
protocol. ",
        "provider": {

```

(continues on next page)



(continued from previous page)

```

    "@type": "Role",
    "roleName": "distributor",
"provider": {
    "@type": "Role",
    "roleName": "distributor"
},
    "fileFormat": [
"Ocean Data View ASCII input v.0.3", "MEDATLAS ASCII v.1"]
    ],
    "identifier":
    [
    {
        "@type": "PropertyValue",
        "propertyID": "dataset identifier",
        "value": "urn:urnSDNCIDILOCALMARIS-TEST"
    }
],
    "includedInDataCatalog": {
        "@type": "DataCatalog",
        "name": "Name of catalog source for record being transformed",
        "url": "not defined"
    },
    "keywords": [
"Oceanographic geographical features", "Atmospheric visibility and
↪transparency", "Ammonium concentration parameters in the water column",
↪"Atmospheric humidity", "aerosol samplers", "Differential Global Positioning System
↪receivers", "cetacean", "Integrated Ocean Drilling Program (IODP) - Artic
↪expedition (ACEX) {acronym=IODP organisation=Natural Environment Research Council
↪(NERC) country=United Kingdom}", "National Coastal Data Co-ordinator {acronym=
↪organisation=Department for Environment, Food and Rural Affairs (DEFRA)
↪country=United Kingdom}", "GEOWARN - Geo-spatial warning system Nisyros volcano
↪(Greece). An emergency case study. {acronym=GEOWARN organisation=Hellenic Centre
↪for Marine Research, Institute of Oceanography (HCMR/IO) country=Greece}"],
    "license": [{
        "@type": "DigitalDocument", "name": "MD_Constraints",
        "description": "useLimitation: Not applicable.  "},
        {
            "@type": "DigitalDocument", "name": "MD_LegalConstraints",
            "description": "accessConstraints: otherRestrictions.    otherConstraints: "},
        {
            "@type": "DigitalDocument", "name": "MD_LegalConstraints",
            "description": "accessConstraints: otherRestrictions.    otherConstraints: "},
        {
            "@type": "DigitalDocument", "name": "MD_LegalConstraints",
            "description": "accessConstraints: otherRestrictions.    otherConstraints: "
        }
    ],
    "publisher": "publisher not specified",
    "spatialCoverage": {
        "@type": "Place",
        "geo":
        [
        {
            "@type": "GeoShape",
            "box": "-68.548849, 59.400296 -49.007153, 73.889864"
        }
        ],
    },

```

(continues on next page)

(continued from previous page)

```
{
  "@type": "GeoShape",
  "box": "106.574831, 5.916728 114.842479, 17.636232"
},
{
  "@type": "GeoShape",
  "box": "-80.198605, -57.079131 -70.052045, -36.777203"
}]
}}
```

### 26.3.10 Mapping

Testing approaches to mapping DCAT to [schema.org](https://schema.org/)

Current thinking

- JSON-LD Frame with default values
- SPARQL construct on these resulting frame to generate the new triples

Mapping references

- [https://www.w3.org/2015/spatial/wiki/ISO\\_19115\\_-DCAT-\\_Schema.org\\_mapping](https://www.w3.org/2015/spatial/wiki/ISO_19115_-DCAT-_Schema.org_mapping)
- <https://ec-jrc.github.io/dcat-ap-to-schema-org/>
- <https://data.gov.au/data/dataset/67ca5de1-8774-4678-9d1b-8b1cb70ab33c.jsonld>

### Methodology

We will load the DCAT JOSH-LD example and explore approaches to converting this to a form that can be used for [schema.org](https://schema.org/).

Possible approaches include

- Inferencing
  - ref: <https://derwen.ai/docs/kgl/infer/>
- SPARQL CONSTRUCT
  - <https://rdflib.readthedocs.io/en/stable/apidocs/rdflib.html>
  - [https://derwen.ai/docs/kgl/ex4\\_0/](https://derwen.ai/docs/kgl/ex4_0/)
- JSON-LD APIs
  - <https://w3c.github.io/json-ld-framing/#omit-default-flag>
- Context modification

```
!pip install -q kglab
```

```
ERROR: pip's dependency resolver does not currently take into account all the
↳ packages that are installed. This behaviour is the source of the following
↳ dependency conflicts.
boto3 1.17.102 requires botocore<1.21.0,>=1.20.102, but you have botocore 1.20.49
↳ which is incompatible.
```

```
import kglab
import json
import rdflib
```

```
# load our JSON into a var to use later
f = open('dcatEx.json',)
j = json.load(f)
f.close()
```

## JSON-LD

Use a frame to pull the elements we want to map, then alter the context for that frame or otherwise cast to new namespace.

Frame with defaults and then work to convert to new names space with SPARQL construct

## SPARQL CONSTRUCT example

Refs:

- [https://derwen.ai/docs/kgl/ex4\\_0/](https://derwen.ai/docs/kgl/ex4_0/)

```
from icecream import ic
from pathlib import Path

txt = Path('dcatEx.json').read_text()

g = rdflib.Graph()
g.parse(data=txt, format="json-ld")
```

```
<Graph identifier=Nad1628ac8eb84e5881b07f2b9ac96afd (<class 'rdflib.graph.Graph'>)>
```

```
sparql = """
    SELECT ?s ?p ?o
    WHERE {
        ?s ?p ?o .
    }
    LIMIT 1
    """
```

```
for row in g.query(sparql):
    ic(row.asdict())
```

```
ic| row.asdict(): {'o': rdflib.term.Literal('LAND-Cover'),
                  'p': rdflib.term.URIRef('http://www.w3.org/ns/dcat#keyword'),
                  's': rdflib.term.URIRef('https://data.gov.au/dataset/67ca5de1-8774-4678-9d1b-8b1cb70ab33c')}
```

```
sparqlc = """
PREFIX dbpedia: <http://dbpedia.org/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX mo: <http://purl.org/ontology/mo/>
```

(continues on next page)

(continued from previous page)

```
PREFIX schema: <https://schema.org/>

CONSTRUCT {
    ?s schema:identifier ?o .
}
WHERE {
    ?s dct:identifier ?o .
}
"""

gres = g.query(sparqlc)
context = {"@vocab": "https://schema.org/", "@language": "en"}
print(gres.serialize(format='json-ld', context=context, indent=4))

# g.parse(gres, format="nt")

# for row in gres:
#     print("-----")
#     print(row)
```

```
b'{"@context": {"@language": "en", "@vocab": "https://schema.org/",
  "id": "https://data.gov.au/dataset/67ca5de1-8774-4678-9d1b-8b1cb70ab33c",
  "identifier": {"@value": "67ca5de1-8774-4678-9d1b-8b1cb70ab33c"}
}
```

```
import kglab

namespaces = {
    "adms": "http://www.w3.org/ns/adms#",
    "dcat": "http://www.w3.org/ns/dcat#",
    "dct": "http://purl.org/dc/terms/",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "gsp": "http://www.opengis.net/ont/geosparql#",
    "locn": "http://www.w3.org/ns/locn#",
    "owl": "http://www.w3.org/2002/07/owl#",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "schema": "http://schema.org/",
    "skos": "http://www.w3.org/2004/02/skos/core#",
    "time": "http://www.w3.org/2006/time",
    "vcard": "http://www.w3.org/2006/vcard/ns#",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
}

kg = kglab.KnowledgeGraph(
    name = "DCAT example",
    base_uri = "https://www.example.org/",
    namespaces = namespaces,
)

kg.load_jsonld("dcatEx.json")
```

```
<kglab.kglab.KnowledgeGraph at 0x7f702ccf86a0>
```

```
sparql2 = """
```

(continues on next page)

(continued from previous page)

```
SELECT ?s ?o
WHERE {
    ?s dct:description ?o .
}
"""
```

```
import pandas as pd
pd.set_option("max_rows", None)

df = kg.query_as_df(sparql2)
df.head(20)
```

```

                                s \
0  <https://data.gov.au/dataset/67ca5de1-8774-467...
1  <https://data.gov.au/dataset/67ca5de1-8774-467...

                                o
0                                Data File
1  ## **Abstract** \n\nThis dataset and its metad...
```

```
pyvis_graph = kg.visualize_query(sparql2, notebook=True)

pyvis_graph.force_atlas_2based()
pyvis_graph.show("tmp.fig06.html")
```

```
<IPython.lib.display.IFrame at 0x7f702ccc4d00>
```

## SHACL Rules

```
import pyshacl
```

```
from pyshacl import validate

conforms, v_graph, v_text = validate(data_graph="./learning.jsonld",
                                     shacl_graph='./oih_learning.ttl',
                                     data_graph_format="json-ld",
                                     shape_graph_format="ttl",
                                     inference='none',
                                     serialize_report_graph="json-ld")
```

```
print(conforms)
print(v_graph)
print(v_text)
```

```
True
b'[\n {\n   "@id": "_:N8be15736f0b945d19b55b266f8475c54",\n   "@type": [\n
  ↪ "http://www.w3.org/ns/shacl#ValidationReport"\n   ],\n   "http://www.w3.org/ns/
  ↪ shacl#conforms": [\n     {\n       "@value": true\n     }\n   ]\n }]\n'
```

Validation Report  
Conforms: True

```
from pyshacl import Validator

# v = Validator(data_graph=dg_basin, shacl_graph=rule, options={"inference": "rdfs"},
#               ↪ont_graph=ont)
# conforms, report_graph, report_text = v.run()
# expanded_graph = v.target_graph

df = Path('data.ttl').read_text()
dg = rdflib.Graph()
dg.parse(data=df, format="ttl")

sf = Path('shape.ttl').read_text()
sg = rdflib.Graph()
sg.parse(data=sf, format="ttl")

v = Validator(data_graph=dg, shacl_graph=sg, options={"inference": "none", "advanced
#               ↪": True}) # turn off rdfs inferencing
conforms, report_graph, report_text = v.run()
expanded_graph = v.target_graph
```

```
# print(conforms)
# print(v_graph)
# print("-----")
# print(v_text)
# print(expanded_graph)
```

```
print(expanded_graph.serialize(format="ttl").decode("utf-8"))
```

```
@prefix ex: <http://example.com/ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

ex:InvalidRectangle a ex:Rectangle .

ex:NonSquareRectangle a ex:Rectangle ;
    ex:height 2 ;
    ex:width 3 .

ex:SquareRectangle a ex:Rectangle,
    ex:Square ;
    ex:height 4 ;
    ex:width 4 .
```

### Notes on SHACL AF Rules

We need to add in PROV triples in this process to note the generation of these triples and the source IRI that results in the product IRI and the actor (?reference)

Maybe review: <https://www.w3.org/TR/2013/REC-prov-o-20130430/#qualifiedPrimarySource>

```
df = Path('dcat.ttl').read_text()
dg = rdflib.Graph()
dg.parse(data=df, format="ttl")

sf = Path('dcatsdo.ttl').read_text()
sg = rdflib.Graph()
```

(continues on next page)

(continued from previous page)

```
sg.parse(data=sf, format="ttl")

v = Validator(data_graph=dg, shacl_graph=sg, options={"inference": "none", "advanced
↳": True}) # turn off rdfls inferencing
conforms, report_graph, report_text = v.run()
expanded_graph = v.target_graph

print(expanded_graph.serialize(format="ttl").decode("utf-8"))
```

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix ex: <http://example.com/ns#> .
@prefix ns1: <http://www.w3.org/ns/dcat#> .
@prefix ns2: <http://xmlns.com/foaf/0.1/> .
@prefix ns3: <http://www.w3.org/ns/locn#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<https://data.gov.au/dataset/67ca5de1-8774-4678-9d1b-8b1cb70ab33c> a ns1:Dataset ;
    ex:area2 "a descriptiono of this dataset" ;
    dct:description "a descriptiono of this dataset" ;
    dct:identifier "67ca5de1-8774-4678-9d1b-8b1cb70ab33c" ;
    dct:issued "2016-03-23T05:08:17.991412"^^xsd:dateTime ;
    dct:language "eng" ;
    dct:modified "2019-11-19T23:18:49.871451"^^xsd:dateTime ;
    dct:publisher <https://data.gov.au/organization/69f37b4c-bdf0-4c85-bd56-
↳82fa6d6b087a> ;
    dct:spatial [ a dct:Location ;
        ns3:geometry "POLYGON ((110.0012 -10.0012, 115.0080 -10.0012, 155.0080 -
↳45.0036, 110.0012 -45.0036, 110.0012 -10.0012))"^^<http://www.opengis.net/ont/
↳geosparql#wktLiteral>,
        "{\type\": \"Polygon\", \"coordinates\": [[[110.0012, -10.00117],
↳[115.008, -10.00117], [155.008, -45.00362], [110.0012, -45.00362], [110.0012, -10.
↳00117]]]}"^^<https://www.iana.org/assignments/media-types/application/vnd.geo+json>
↳] ;
    dct:title "Dynamic Land Cover Dataset" ;
    ns1:distribution <https://data.gov.au/dataset/67ca5de1-8774-4678-9d1b-
↳8b1cb70ab33c/resource/1f8174f8-573e-43f2-b110-3d1a13c380e8> ;
    ns1:keyword "Australia",
        "Cooper subregion",
        "LAND-Cover",
        "Maranoa-Balonne-Condamine subregion",
        "biota",
        "environment",
        "planningCadastre" .

<https://data.gov.au/dataset/67ca5de1-8774-4678-9d1b-8b1cb70ab33c/resource/1f8174f8-
↳573e-43f2-b110-3d1a13c380e8> a ns1:Distribution ;
    dct:description "Data File" ;
    dct:format "ZIP" ;
    dct:title "Dynamic Land Cover Dataset" ;
    ns1:accessURL <https://datagovau.s3.amazonaws.com/bioregionalassessments/BA_ALL/
↳ALL/DATA/Geography/LandCoverNDLC/1556b944-731c-4b7f-a03e-14577c7e68db.zip> ;
    ns1:byteSize 186838338.0 .

<https://data.gov.au/organization/69f37b4c-bdf0-4c85-bd56-82fa6d6b087a> a
↳ns2:Organization ;
    ns2:name "Bioregional Assessment Program" .
```

### 26.3.11 Georeference KG completion example

#### About

This is a simple test notebook to explore approaches to associating geometries in the OIH graph with named ocean regions. From Marine Regions (<https://www.marineregions.org/>) we downloaded the IHO Sea Areas dataset. This is a shapefile that is converted to WKT and loaded into a geopandas dataframe.

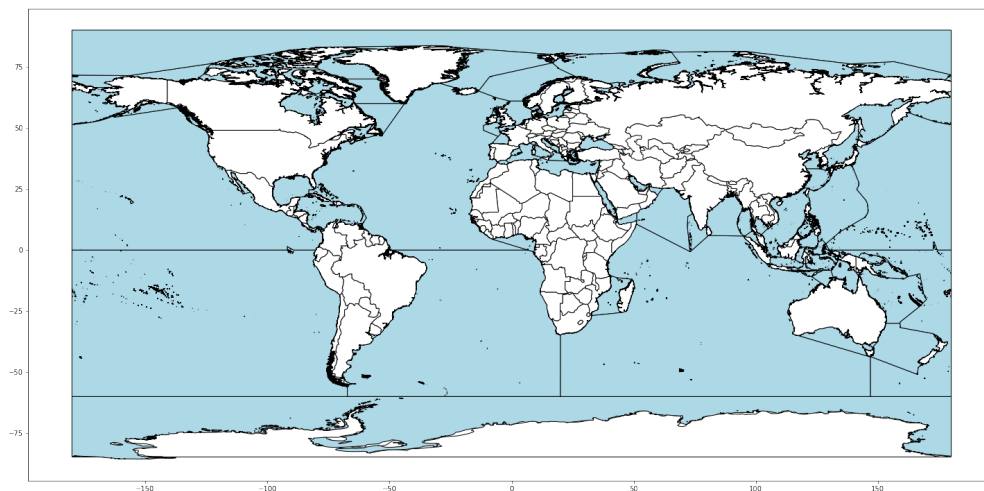
We then create a few points to compare against this. In this test we are only comparing points to the polygons. So this is what is called a “point in polygon” test. Later revisions to this could do polygon intersection or other tests. This is only a proof of concept notebook.

We then generate a new dataframe based on the matches. These matches could then be fed back into the graph as a sort of “Knowledge Graph Completion” workflow. Alternatively we can do `geosparql` calls based on the sea WKT strings and do the search in the `geosparql` aware triple store.

#### References

- [https://geopandas.org/docs/user\\_guide/set\\_operations.html](https://geopandas.org/docs/user_guide/set_operations.html)
- [https://geopandas.org/docs/user\\_guide/geocoding.html](https://geopandas.org/docs/user_guide/geocoding.html)
- <https://medium.com/analytics-vidhya/point-in-polygon-analysis-using-python-geopandas-27ea6788bfff>

#### Process





### Conclusion

The generated dataframe holds the matches of the test Lat Long pairs to the named seas from the reference shape file. These results could be fed back into the graph as keywords or items from a known list of terms for more explicate relation mapping.

Specifically, something like <https://schema.org/DefinedTerm> where the property <https://schema.org/DefinedTermSet> would point back to the Marine Regions source documents and URL. Similarly these resources could be connected up to WikiData in a similar manner.

```
{
  "@type": "DefinedTermSet",
  "@id": "http://geonetwork.vliz.be/geonetwork/srv/eng/catalog.search#/
↪metadata/f4cfa278-730f-4646-b6cc-a3dceaa3a1e5",
  "name": "IHO Sea Areas"
},
{
  "@type": "DefinedTerm",
  "name": "Bay of Bengal",
  "description": "IHO Sea Area Bay of Bengal",
  "inDefinedTermSet": "http://geonetwork.vliz.be/geonetwork/srv/eng/
↪catalog.search#/metadata/f4cfa278-730f-4646-b6cc-a3dceaa3a1e5"
},
```

For reference the WikiData resource is: <https://www.wikidata.org/wiki/Q38684> which is an instance of “body of water”. So leveraging this type and the IHO names should allow relatively reliable link detection. Leveraging the top level Thing class in [schema.org](https://schema.org) we would be looking at a simple

```
"sameAs": "https://www.wikidata.org/wiki/Q38684",
```

in the DefinedTerm type.

In the final listing below the “id” is the just the random string I associated with the test lat longs. I used a simple online map to just pick some random locations and gave them names. The “region” comes from the official Marine Regions file.

### Geospatial KG Completion

#### About

This is a simple test notebook to explore approaches to associating geometries in the OIH graph with named ocean regions. From Marine Regions (<https://www.marineregions.org/>) we downloaded the IHO Sea Areas dataset. This is a shapefile that is converted to WKT and loaded into a geopandas dataframe.

We then create a few points to compare against this. In this test we are only comparing points to the polygons. So this is what is called a “point in polygon” test. Later revisions to this could do polygon intersection or other tests. This is only a proof of concept notebook.

We then generate a new dataframe based on the matches. These matches could then be fed back into the graph as a sort of “Knowledge Graph Completion” workflow. Alternatively we can do `gepsparql` calls based on the sea WKT strings and do the search in the geosparql aware triple store.

### References

- [https://geopandas.org/docs/user\\_guide/set\\_operations.html](https://geopandas.org/docs/user_guide/set_operations.html)
- [https://geopandas.org/docs/user\\_guide/geocoding.html](https://geopandas.org/docs/user_guide/geocoding.html)
- <https://medium.com/analytics-vidhya/point-in-polygon-analysis-using-python-geopandas-27ea67888bff>

```
##@title
# !apt-get install libproj-dev proj-data proj-bin
# !apt-get install libgeos-dev
!pip install -q cython
!pip install -q cartopy
!pip install -q SPARQLWrapper
!pip install -q rdfliib
!pip install -q geopandas
!pip install -q contextily==1.0rc2
!pip install -q rtree
!pip install -q pygeos
```

```
from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd
import numpy as np
import json
import geopandas
import matplotlib.pyplot as plt
import shapely

#dbsparql = "http://dbpedia.org/sparql"
# ufokn = "http://graph.collaborium.io/blazegraph/namespace/oihdev/sparql"
```

```
# Point in Polygon function from the cited Vidhya reference. A few small changes
# to align with our dataframes here

def get_pip (gdf, regions):
    r_list = list(regions.NAME)
    #create empty dataframe
    df = pd.DataFrame().reindex_like(gdf).dropna()
    for r in r_list:
        #get geometry for specific region
        pol = (regions.loc[regions.NAME==r])
        pol.reset_index(drop = True, inplace = True)
        #identify those records from gdf that are intersecting with the region polygon
        pip_mask = gdf.within(pol.loc[0, 'WKT'])
        #filter gdf to keep only the intersecting records
        pip_data = gdf.loc[pip_mask].copy()
        #create a new column and assign the region name as the value
        pip_data['region']= r
        #append region data to empty dataframe
        df = df.append(pip_data)
    #checking there are no more than one region assigned to an event
    print('Original dataframe count=',len(gdf), '\nNew dataframe count=', len(df))
    if df.loc[df.id.duplicated() == True].shape[0] > 0:
        print("There are id's with more than one region")
    #checking all events have a region
    elif gdf.loc[~gdf.id.isin(df.id)].shape[0] > 0:
        print("There are id's without an assigned region")
```

(continues on next page)

(continued from previous page)

```

else:
    print("No discrepancies in results!")
df.reset_index(inplace=True, drop=True)
df = df.drop(columns='geometry')
return df

```

```

# load CSV into pandas that holds the WKT strings for the world seas dataset
df = pd.read_csv('./World_Seas_IHO_v3/out.wkt/World_Seas_IHO_v3.csv')
# df.head(2)

```

```

# convert the WKT strings to WKT geometry
from shapely import wkt
df['WKT'] = df['WKT'].apply(wkt.loads)

```

```

# Load and plot the ocean regions
import contextily as ctx

# Make geopandas from df
gdf = geopandas.GeoDataFrame(df, geometry='WKT', crs={"init": "epsg:3857"}) #_
↳ contextily requires 3875?
# gdf.head(2)

```

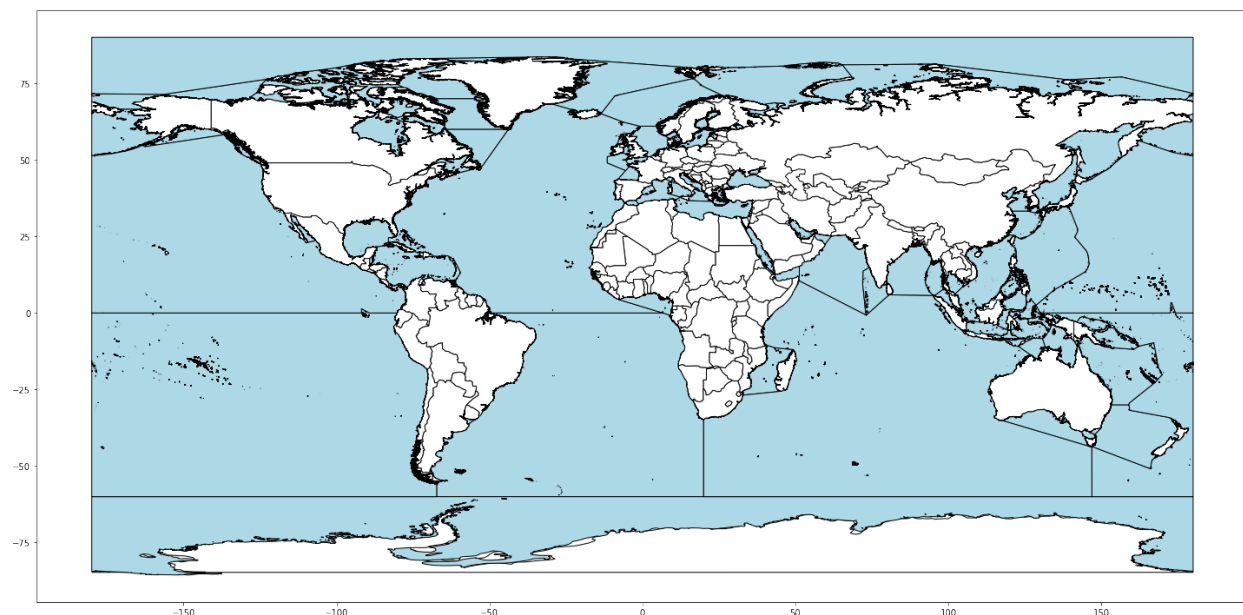
```

# Make geopandas from GeoJSON of the world for plot
import matplotlib.pyplot as plt

url = "https://raw.githubusercontent.com/johan/world.geo.json/master/countries.geo.
↳ json"
dftx = geopandas.read_file(url)

# plot
ax = dftx.plot(color='white', edgecolor='black', figsize=(25,15))
gdf.plot(ax=ax, edgecolor='black', color='lightblue')
plt.savefig('map.png')

```



```
# Make test datarame with some lat long pairs to test
# Later these will come from SPARQL calls to the OIH graph

df = pd.DataFrame({'Latitude': [38, -1, -37, -22, 14],
                   'Longitude': [-146, 0, 129, 37, 85],
                   'id': ["Off US West Coast", "Atlantic", "South of Australia", "Africa", "Near India"]})

testdf = geopandas.GeoDataFrame(df, geometry=geopandas.points_from_xy(df.Longitude, df.Latitude))
```

```
# Call the function and do pip calculations
eq_df = get_pip(testdf, gdf)
```

```
Original dataframe count= 5
New dataframe count= 5
No discrepancies in results!
```

### Conclusion

The generated dataframe holds the matches of the test Lat Long pairs to the named seas from the reference shape file. These results could be fed back into the graph as keywords or items from a known list of terms for more explicate relation mapping.

Specifically, something like <https://schema.org/DefinedTerm> where the property <https://schema.org/DefinedTermSet> would point back to the Marine Regions source documents and URL. Similarly these resources could be connected up to WikiData in a similar manner.

```
{
  "@type": "DefinedTermSet",
  "@id": "http://geonetwork.vliz.be/geonetwork/srv/eng/catalog.search#/metadata/f4cfa278-730f-4646-b6cc-a3dceaa3a1e5",
  "name": "IHO Sea Areas"
},
{
  "@type": "DefinedTerm",
  "name": "Bay of Bengal",
  "description": "IHO Sea Area Bay of Bengal",
  "inDefinedTermSet": "http://geonetwork.vliz.be/geonetwork/srv/eng/catalog.search#/metadata/f4cfa278-730f-4646-b6cc-a3dceaa3a1e5"
},
```

For reference the WikiData resource is: <https://www.wikidata.org/wiki/Q38684> which is an instance of “body of water”. So leveraging this type and the IHO names should allow relatively reliable link detection. Leveraging the top level Thing class in [schema.org](https://schema.org) we would be looking at a simple

```
"sameAs": "https://www.wikidata.org/wiki/Q38684",
```

in the DefinedTerm type.

In the final listing below the “id” is the just the random string I associated with the test lat longs. I used a simple online map to just pick some random locations and gave them names. The “region” comes from the official Marine Regions file.

```
eq_df.head()
```

	Latitude	Longitude	id	region
0	-37.0	129.0	South of Australia	Great Australian Bight
1	-22.0	37.0	Africa	Mozambique Channel
2	-1.0	0.0	Atlantic	South Atlantic Ocean
3	14.0	85.0	Near India	Bay of Bengal
4	38.0	-146.0	Off US West Coast	North Pacific Ocean



# **Part V**

## **Validation**





## VALIDATION

### 27.1 About

This section contains some initial work on developing some validation approaches for OIH. The focus initially is not on validating approaches with the full publishing guidance. Rather the focuses is on the the “info hub” as a search application and develops validation to support that.

Initial approach:

- Develop a base SHACL document that assess a data graph based on elements needed to support search
- Leverage [SHACL Playground](#)

To support this will need an initial data graph to work with. The type is not important. All types will need to satisfy the search needs.

Examples of these needs include:

- Have an @id
- Have a name
- Have a description
- Have a Distribution and contentURL
- Reference authority

### 27.2 Implementation

Work this implementation can be found in the notebooks section [pySHACL testing](#).




# **Part VI**

## **Interfaces**



## 28.1 About

The reference client used for development is currently hosted at [oceans.collaborium.io](https://oceans.collaborium.io). This is a fully client side Javascript based client to the OIH index. All the code is hosted at <https://github.com/iodepo/odis-arch/tree/master/schema/client/referenceclient/website>

 **OIH** [Home](#) [About](#) [Using](#)

Source: IODE OceanExpert  
Score: 0.4375  
# Course

### OBIS Asian nodes and Coral Reef

Overview The course provides an introduction to the Ocean Biogeographic Information System (OBIS). This includes best practices in marine biogeographic data management, data publication, data access, data analysis and data visualisation. Aims and Objectives - Reinforce and expand the OBIS network in the South-East Asian Region - Increase awareness on international standards and best practices related to marine biogeographic data management - Increase the amount and quality of open access biodiversity data published through OBIS and its OBIS nodes - Increase the use of data from OBIS for research, species conservation and area-based management applications for sustainable development Learning Outcomes - Knowledge and understanding of OBIS structure, mission and objectives - Use of Darwin Core standards for species occurrence records, taxonomy, event/sample records and additional...

Source: IODE OceanExpert  
Score: 0.375  
# Course

### OTGA/INIOAS: Remote Sensing of Coral Reefs

The course provides an introduction to the capabilities of the Remote Sensing for the coral reefs mapping. This includes practices in processing of high spatial resolution remotely sensed data for mapping the nearshore coral reef communities. After the workshop, the trainees will be able

## QUERYING SPARQL

### 29.1 About

This page will hold some information about the SPARQL queries we use and how they connect with some of the profile guidance in this document. We will show how this relates to and depends on the Gleaner prov as well as the Authoritative Reference elements of the patterns. It is expected that the Gleaner prov will be present, though this can be made optional in case other indexing systems are used that do not provide this prov shape. The SPARQL will be looking for both Gleaner prov and the Authoritative Reference elements.

This will be different for different patterns. For example, it might relate to the publisher provider elements for Creative-works, but to the identity element for People and Organizations.

Also here will be a SHACL shape to help validate a record as fit for use with the query. It will provide guidance on what is optional and and is required.

```
1 prefix prov: <http://www.w3.org/ns/prov#>
2 PREFIX con: <http://www.ontotext.com/connectors/lucene#>
3 PREFIX luc: <http://www.ontotext.com/owlim/lucene#>
4 PREFIX con-inst: <http://www.ontotext.com/connectors/lucene/instance#>
5 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
6 PREFIX schema: <https://schema.org/>
7 PREFIX schemaold: <http://schema.org/>
8 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
9 SELECT DISTINCT ?g ?s ?wat ?orgname ?domain ?type ?score ?name ?url ?lit ?
  ↳description ?headline
10 WHERE
11
12     ?lit bds:search "coral" .
13     ?lit bds:matchAllTerms "false" .
14     ?lit bds:relevance ?score .
15     graph ?g {
16         ?s ?p ?lit .
17         ?s rdf:type ?type .
18         OPTIONAL { ?s schema:name ?name . }
19         OPTIONAL { ?s schema:headline ?headline . }
20         OPTIONAL { ?s schema:url ?url . }
21         OPTIONAL { ?s schema:description ?description . }
22     }
23     ?sp prov:generated ?g .
24     ?sp prov:used ?used .
25     ?used prov:hadMember ?hm .
26     ?hm prov:wasAttributedTo ?wat .
27     ?wat rdf:name ?orgname .
28     ?wat rdfs:seeAlso ?domai
```

(continues on next page)

(continued from previous page)

```
29 }  
30 ORDER BY DESC(?score)  
31 LIMIT 30  
32 OFFSET 0
```



## **Part VII**

# **Appendix**



## 30.1 About

A collection of items related to the OIH development. Mostly related to examples around representing concepts in the graph such as date and time, language etc.

As these develop they may be moved into other sections of the book.

## 30.2 Known Issues

### 30.2.1 About

This document will collect some of the various issues we have encounter in publishing the JSON-LD documents.

#### control characters in URL string for sitemap or in the JSON-LD documents

Make sure there are no control characters such as new line, carriage returns, tabs or others in the document. These can be problematic both for processing and display.

#### context is a map (changed from 1.0)

Be sure to use a context style like:

```
"@context": {  
  "@vocab": "https://schema.org/"  
},
```

The context section must be a map starting in JSON-LD 1.1

### data graphs need @id

Be sure to include an @id in your graph that points to the identifier or the web address of the resource providing the metadata. This is not the material the metadata is about, but rather the metadata record itself.

### string literals must be valid

The string literals must be sure to not have quotation marks or other invalid characters without escaping or encoding them.

## 30.3 References

A broad collection of references.

### 30.3.1 General

Science on Schema

BioSchemas

Ocean Best Practices on Schema

PID policy for European Open Science Cloud

DCAT Schema.org mappings

DCAT US Data.gov reference

FAIR Semantics

### 30.3.2 Developer References

Schema.org releases

Schema.org RDF graph (turtle format)

JSON-LD Playground

Google Developers Search Gruid

Google Developers Fact Check

Structured Data Testing Tool

Rich Results Testing Tool

JSON-LD

Ruby JSON-LD

schema.org Java

Perl classes for schema.org markup

### 30.3.3 Content Management and Web Server support

Drupal Support

Wordpress Claim Review

### 30.3.4 Organizations

Google Open Source

DataCommons & DataCommons REST

### 30.3.5 Indexers

Gleaner BMUSE

### 30.3.6 Graph tools

Wikipedia SPARQL implementation list

RDFlib

Any23

rdfjs

shemaram

shemarama demo

validatingrdf

Structured Data Linter

### 30.3.7 Blogs and Press Releases

Yandex: What is Scheme.org

Bing: Fact Check Label

Bing: Contextual Awareness

Facebook: Marketing API

Facebook: fact checking

Amazon: Alex skills

Google Developers mail invoice

### 30.3.8 Not Categorized Yet

Lighthouse Plugins

Lighthouse

Science on Schema

BioSchemas

CodeMeta

Linter Structured Data

JSON-LD Playground

JSON-LD.org

SHACL playground

Google Structured Data testing tool (

Google Dataset for developers

Press article

Rich results

SchemaApp.com

Yandex

Schema dev

Chromeextension

Google Rich Results

Datashapes

ACL Web Alexa Meaning Representation Language

hash.aio Volcano schema

Yeast Structured Data Guide

Schema App

Springer: Scigraph

iPhylo Biodiversity KG

ozymandias

RDA group meeting notes

RDA Plenary meeting

## 30.4 Registries

### 30.4.1 Documents and Datasets

#### DOI

A not-for-profit membership organization that is the governance and management body for the federation of Registration Agencies providing Digital Object Identifier (DOI) services and registration, and is the registration authority for the ISO standard (ISO 26324) for the DOI system. The DOI system provides a technical and social infrastructure for the registration and use of persistent interoperable identifiers, called DOIs, for use on digital networks.

#### Datcite

Locate, identify, and cite research data with the leading global provider of DOIs for research data.

Archival Resource Key or ARK: and [N2T ARKs](#) and [Names to Thinkgs](#)

### 30.4.2 People

#### Orcid

ORCID's mission is to enable transparent and trustworthy connections between researchers, their contributions, and their affiliations by providing a unique, persistent identifier for individuals to use as they engage in research, scholarship, and innovation activities.

### 30.4.3 Organizations

#### re3data

A registry of research data repositories

#### ROR

ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organization in the world.

#### Grid

GRID is a free and openly available global database of research-related organisations, cataloging research-related organisations and providing each with a unique and persistent identifier. With GRID you have over 99,609 carefully curated records at hand, enabling you to identify and distinguish research-related institutions worldwide.

### 30.4.4 Physical Samples

#### IGSN

The objective of the IGSN e.V. is to implement and promote standard methods for identifying, citing, and locating physical samples with confidence by operating an international IGSN registration service.

## 30.5 Controlled Vocabularies

### 30.5.1 About

See also: [ODIS-ARCH Vocabularies](#)

A list of possible controlled vocabularies to use in the [schema.org](#) documents. Many such resources can be found by searching at [BARTOC.org](#) or the [UNESCO Thesaurus](#).

Note, at present this is an exploration and there is not yet a recommendation for use in OIH.

### 30.5.2 List

[Essential Climate Variables](#)

[Vocabulary based on DFG'S Classification of Subject Area, Review Board, Research Area and Scientific Discipline \(2016 - 2019\)](#)

[Nature Subjects Ontology](#)