What is Schema.org?
Hotel

A Schema.org Type

Thing > Organization > LocalBusiness > LodgingBusiness > Hotel
Thing > Place > LocalBusiness > LodgingBusiness > Hotel

A hotel is an establishment that provides lodging paid on a short-term basis (Source: Wikipedia, the free encyclopedia, see http://en.wikipedia.org/wiki/Hotel).

See also the dedicated document on the use of schema.org for marking up hotels and other forms of accommodations.

https://github.com/schemaorg/schemaorg/blob/main/data/releases/15.0/schemaorg-all-http.rdf
Wyndham Grand Pittsburgh Downtown

3.8 ★★★★ 4,316 Google reviews

3-star hotel

Address: 600 Commonwealth PI, Pittsburgh, PA 15222

Phone: (412) 391-4600

Hotel details

In the Central Business District, this upscale contemporary hotel is 0.7 miles from The Andy Warhol Museum, 1.6 miles from Heinz Field and 18.7 miles from Pittsburgh International Airport. ... MORE

- Free Wi-Fi
- Breakfast
- Parking
- Accessible
- Indoor pool
- Air-conditioned

View more amenities
Wyndham Grand Pittsburgh Downtown

Address: 600 Commonwealth PI, Pittsburgh, PA 15222
Phone: (412) 391-4600

Hotel details
In the Central Business District, this upscale contemporary hotel is 0.7 miles from The Andy Warhol Museum, 1.6 miles from Heinz Field and 18.7 miles from Pittsburgh International Airport.
The Raphael Hotel is an upscale historic US Renaissance revival-style building dating back to 1928.

Amenities: Free Wi-Fi.


Phone: +1-816-756-3800
Fax: +1-816-802-2150
Email: info@raphaelkc.com
325 Ward Parkway Kansas City MO 64112 USA

View Map

Map Position: 39.039536 -94.591296 Elevation: 1500ft
Check In: 6:00 PM
Check Out: 2:00 PM
Booking Details:
Best Available Rate
Enjoy our hotel's best available rate!
The Raphael Hotel is an upscale historic US Renaissance revival-style building dating back to 1928. Amenities: Free Wi-Fi.

Contact Details:
Telephone: +1-816-756-3800
Fax Number: +1-816-802-2150
Email: info@raphaelkc.com

Location:
Street Address: 325 Ward Parkway
City: Kansas City
State: MO
Postal Code: 64112

Amenities:
- Free Wi-Fi
- Free parking

Enjoy our hotel's best available rate!
Schema.org data on a webpage informs search engines, what your page content is **about**, 
Schema.org data on a webpage informs search engines, what your page content is about, using schemas agreed upon by publishers & consumers,
Schema.org data on a webpage informs search engines, what your page content is about, using schemas agreed upon by publishers & consumers, for making content actionable.
What is actionable content?
Wyndham Grand Pittsburgh Downtown

3.8 ★★★★★ 4,316 Google reviews

3-star hotel

Address: 600 Commonwealth Pl, Pittsburgh, PA 15222

Phone: (412) 391-4600

Hotel details

In the Central Business District, this upscale contemporary hotel is 0.7 miles from The Andy Warhol Museum, 1.6 miles from Heinz Field and 18.7 miles from Pittsburgh International Airport. ... MORE

Free Wi-Fi  Breakfast  Parking

Accessible  Indoor pool  Air-conditioned

View more amenities
Actions are *unique* to the content.
Hotel

A Schema.org Type

Thing > Organization > LocalBusiness > LodgingBusiness > Hotel

A hotel is an establishment that provides lodging paid on a short-term basis (Source: Wikipedia, the free encyclopedia, see http://en.wikipedia.org/wiki/Hotel).

See also the dedicated document on the use of schema.org for marking up hotels and other forms of accommodations.
ResearchProject
A Schema.org Type

This term is proposed for full integration into Schema.org, pending implementation feedback and adoption from applications and websites.

Thing > Organization > Project > ResearchProject
A Research project.

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>actionableFeedbackPolicy</td>
<td>CreativeWork or URL</td>
<td>For a NewsMediaOrganization or other news-related Organization, a statement about public engagement activities (for news media, the newsroom's), including involving the public - digitally or otherwise -- in coverage decisions, reporting and activities after publication.</td>
</tr>
<tr>
<td>address</td>
<td>PostalAddress or Text</td>
<td>The overall rating, based on a collection of reviews or ratings, of the item.</td>
</tr>
<tr>
<td>aggregateRating</td>
<td>AggregateRating</td>
<td>Alumni of an organization.</td>
</tr>
<tr>
<td>alumni</td>
<td>Person</td>
<td>Inverse property: alumniOf</td>
</tr>
<tr>
<td>areaServed</td>
<td>AdministrativeArea or GeoShape or Place or Text</td>
<td>The geographic area where a service or offered item is provided. Supersedes serviceArea.</td>
</tr>
<tr>
<td>award</td>
<td>Text</td>
<td>An award won by or for this item. Supersedes awards.</td>
</tr>
<tr>
<td>brand</td>
<td>Brand or Organization</td>
<td>The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.</td>
</tr>
<tr>
<td>contactPoint</td>
<td>ContactPoint</td>
<td>A contact point for a person or organization. Supersedes contactPoints.</td>
</tr>
</tbody>
</table>
## Dataset

A Schema.org Type

Thing > CreativeWork > Dataset

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

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>distribution</td>
<td>DataDownload</td>
<td>A downloadable form of this dataset, at a specific location, in a specific format.</td>
</tr>
<tr>
<td>includedInDataCatalog</td>
<td>DataCatalog</td>
<td>A data catalog which contains this dataset. Supersedes includedDataCatalog, catalog. Inverse property: dataset</td>
</tr>
<tr>
<td>issn</td>
<td>Text</td>
<td>The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication.</td>
</tr>
<tr>
<td>measurementTechnique</td>
<td>Text or URL</td>
<td>A technique or technology used in a Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (described using variableMeasured). This is oriented towards scientific and scholarly database publication but may have broader applicability; it is not intended as a full representation of measurement, but rather as a high level summary for dataset discovery. For example, if variableMeasured is: molecule concentration, measurementTechnique could be: &quot;mass spectrometry&quot; or &quot;nmr spectroscopy&quot; or &quot;colorimetry&quot; or &quot;immuno-fluorescence&quot;.</td>
</tr>
<tr>
<td>variableMeasured</td>
<td>PropertyValue or Text</td>
<td>The variableMeasured property can indicate (repeated as necessary) the variables that are measured in some dataset, either described as text or as pairs of identifier and description using PropertyValue.</td>
</tr>
<tr>
<td>about</td>
<td>Thing</td>
<td>The subject matter of the content. Inverse property: subjectOf</td>
</tr>
<tr>
<td>abstract</td>
<td>Text</td>
<td>An abstract is a short description that summarizes a CreativeWork.</td>
</tr>
<tr>
<td>accessMode</td>
<td>Text</td>
<td>The human sensory perceptual or cognitive faculty through which a person may process or perceive information. Values should be drawn from the approved vocabulary.</td>
</tr>
<tr>
<td>contactPoint</td>
<td>Text or ContactPoint</td>
<td>A list of single or combined accountIds that are sufficient to understand all the individual(s) involved in this interaction.</td>
</tr>
</tbody>
</table>

[more...]

### Properties from Dataset

- distribution
- includedInDataCatalog
- issn
- measurementTechnique
- variableMeasured

### Properties from CreativeWork

- about
- abstract
- accessMode
What does this mean for science data?
What *actions* does the science data community want?
What *actions* does the science data community want?
What actions does the science data community want?
Science-on-schema.org
ESIP Schema.org Cluster
Common publishing patterns for describing research data on your web pages using schema.org
Easy for publishers; harder for harvesters (to interpret)

Schema.org: Evolution of Structured Data on the Web by Guha & Brickley
DOI:10.1145/2844544

- make is easy for publishers (read lenient)
- the work is for harvesters to reconcile the data
- schema.org is a core vocabulary meant to be extended
Science On Schema.Org (SOSO) Guidance Documents

This repository provides guidance for repository operators and others to follow consistent conventions to provide schema.org markup in Dataset landing pages to improve data discovery through search engines. The main guides help describe Datasets and Data Repositories.

<table>
<thead>
<tr>
<th>Name</th>
<th>Guideline</th>
<th>Examples</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>guidelines for schema.org for a scientific Dataset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Repository</td>
<td>guidelines for schema.org for a research data repository</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Getting Started** - explains some useful techniques that will be common across all schema.org types
- **Experimental** - some suggestions for more advanced dataset documentation, proposed for inclusion in future SOSO guidelines.

Citation

Science on Schema.org (SOSO) releases are archived on Zenodo, and the current release should be cited as:

ESIP Schema.org Cluster

1. Monitor monthly to schema.org updates

2. Accept contributions and issues at Github

3. Maintain & publish updates to guidelines

Data Repository - a repository with research data

Dataset - the scientific dataset

mailing list | #sci-schemaorg (Slack) | Agendas (Google Doc)
Why guidelines?
WHY?

**Q:** How do we share patterns of use so that no one is left behind?

**schema.org**
- Flat descriptions
  - How are things connected?
- Limited examples
- Endless ways to publish

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>variableMeasured</td>
<td>The variableMeasured property can indicate (repeated as necessary) the variables that are measured in some dataset, either described as text or as pairs of identifier and description using PropertyValue.</td>
</tr>
<tr>
<td>properties from CreativeWork</td>
<td>The subject matter of the content. Inverse property: subjectOf.</td>
</tr>
<tr>
<td>about</td>
<td>An abstract is a short description that summarizes a CreativeWork.</td>
</tr>
<tr>
<td>abstract</td>
<td>The human sensory perceptual system or cognitive faculty through which a person may process or perceive information. Expected values include: auditory, tactile, textual, visual, colorDependent, chartOnVisual, chemOnVisual, diagramOnVisual, mathOnVisual, musicOnVisual, textOnVisual.</td>
</tr>
<tr>
<td>accessMode</td>
<td>A list of single or combined accessModes that are sufficient to understand all the intellectual content of a resource. Expected values include: auditory, tactile, textual, visual.</td>
</tr>
<tr>
<td>accessModeSufficient</td>
<td>Indicates that the resource is compatible with the referenced accessibility API (WebSchemas wiki lists possible values).</td>
</tr>
<tr>
<td>accessibilityAPI</td>
<td>Text</td>
</tr>
</tbody>
</table>

Why?
Guidelines for Shared Publishing Patterns

➢ Reliable, consistent **federation**

➢ Automate Validation

---

**Organization**

The sources for geodex come mostly from collaboration with the EarthCube Council of Data Facilities (CDF).

**Providers**

CDF members who express their resources via structured data on the web approaches can be indexed.

**Indexing**

Geodex uses the gleaner program (gleaner.io) to build the index and (GROW) as a server. See the about section for more.

---

**Dataset**

Photosymbiosis in planktonic foraminifera across the Palaeocene-Eocene Thermal Maximum

**Validation Failed**

5 errors. 10 / 23 tests applied.

**Violation**

Dataset must have an ID

**Warning**

It is recommended that a Dataset includes a sameAs URL Path: http://schema.org/sameAs

**Warning**

It is recommended that a Dataset indicates accessibility for free or otherwise Path: http://schema.org/isAccessibleForFree

**Violation**

Dataset must have a version as Literal or Number Path: http://schema.org/version

**Violation**

Dataset identifiers must be a URL, Text or PropertyValue Path: http://schema.org/identifier
Describing a Repository

- Basic Fields
- Identifier
- Funding Source
- Services
- Data Collections
- Advanced Publishing Techniques
  - How to use external vocabularies
Describing a Dataset

- Describing a Dataset
  - Common Properties
    - Identifier
    - Variables
    - Catalog
    - Metadata
  - Distributions
    - Accessing Data through a Service Endpoint
  - Temporal Coverage
  - Spatial Coverage
  - Roles of People
  - Publisher / Provider
  - Funding
  - License
  - Advanced Publishing Techniques
    - Attaching Physical Samples to a Dataset
Guidelines: Drawings & Examples

Once the linkage has been made, further details about the metadata can be provided. We recommend using `schema:encodingFormat` to indicate the metadata format/vocabulary to which the metadata record conforms. If it conforms to multiple formats, or to a specific and general format types, multiple types can be listed.

We use the `schema:DataDownload` class for Metadata files so that we can use the `schema:encodingFormat` to indicate the file format and encoding, etc.

It can be useful to aggregators and other consumers to indicate when the metadata record was created or last modified, which can be used to optimize harvesting schedules for search indices and other applications.

An example of a metadata reference to an instance of EML-formatted structured metadata:

```json
{
  "@context": "https://schema.org/",
  "@type": "Dataset",
  "name": "Removal of organic carbon by natural bacterioplankton communities as a function of pCO2 from laboratory experiments",
  "distribution": {
    "@type": "DataDownload",
    ...
  },
  "subjectOf": {
    "@type": "DataDownload",
    "name": "eml-metadata.xml",
    "description": "EML metadata describing the dataset",
    "encodingFormat": ["*application/xml", "https://ecoinformatics.org/eeml-2.2.0"],
    "dateModified": "2019-06-12T14:44:15Z"
  }
}
```

Alternatively, if the schema.org record is meant to describe the metadata record, one could use the inverse property `schema:about` to indicate the linkage back to the Dataset that it describes. This would be a more rare situation, as typically the schema.org record would be.
A point, or coordinate, would be defined in this way:

```json
{
   "@context": {
      "@vocab": "http://schema.org/",
      "datacite": "http://purl.org/spar/datacite/"
   },
   "@type": "Dataset",
   "name": "Removal of organic carbon by natural bacteriopla...
   "spatialCoverage": {
      "@type": "Place",
      "geo": {  
         "@type": "GeoCoordinates",
         "latitude": 39.3280
         "longitude": 120.1633
      }
   }
}
```

All other shapes are defined using the `schema:GeoShape`:

```json

"spatialCoverage": {
   "@type": "Place",
   "geo": {  
      "@type": "GeoShape",
      "line": [39.3280, 120.1633, 40.445, 123.7878]
   }
}
```

Guidelines: Drawings & Examples
How to Use Schema.org on your Dataset Web Pages

bit.ly/soso-tutorial

Takeaways:
● How to build Schema.org record
● Learn JSON-LD tips and techniques
● Introduce the latest updates to science-on-schema.org
  ○ Dataset Funding (Awards)
  ○ Variables
  ○ Geologic Time

Tags: schema-org, JSON-LD, metadata, tutorial

Want to learn more?
Contact: Adam Shepherd
Telecon: August 25th, 2:30pmET

- Build a Dataset (it’s not complicated)
  ○ #1. JSON-LD Context
  ○ #2. Basic Fields
  ○ #3. Keywords
  ○ #4. License
  ○ #5. Identifiers
    ▪ using sameAs
  ○ #6. Publisher
  ○ #7. Authors / Contributors
  ○ #8. Temporal Coverage
  ○ #9. Spatial Coverage
  ○ #10. Data Files

- Improving the Dataset
  ○ #11. Keywords
  ○ #12. Identifiers
  ○ #13. Authors / Contributors Improved
  ○ #14. Provider
  ○ #15. Funding & Awards
  ○ #16. Variables
  ○ #17. Metadata Records
  ○ #18. Checksums

- Advanced Techniques
  ○ Geologic Time
  ○ Provenance
  ○ Validating with SHACL
How are the guidelines managed?
Governance

architecture decision record (ADR)

noun

document that captures an important architectural decision made along with the context and consequences
Governance - Github Issues & Telecons

7 Feb 2022 @ 5pm EST

Attendees:
Matt, Megan, Stephen, Gary Berg-Cross, Dave, Rebecca R, Nick

Agenda

- ESIP Highlights Session Friday 1pm EST
  - https://docs.google.com/document/d/1GZ-zAgDE0cdfuccuOqMVAMNBGM-JANB5EgxUSVYyS/edit
  - 2 minute highlight
  - Stephen is available, will get template from Megan
  - Need to provide 1 slide following template by end of day tomorrow
- Finish v1.3 [github project]
  - Schema.org context namespace (Issue #151)
    - Reversed the recs in all examples to recommend pulling the context file
    - Harvesters will need to normalize for sanity
    - Matt closed this, need to be sure new examples follow the new context approach
  - Representing geo time (Issue #77)
    - Still need to incorporate text form the decision doc into the Dataset guide
  - variableMeasured (Issue #24)
    - PR still exists
  - Experimental.md (Issue #180)
    - Steve made updates following the recs last meetings
    - TODO: Examples need to follow the new context loading guidelines
    - TODO: the IGSN @id example should use the “igsn:” prefix
    - Needs review and merge
Using Architectural Decision Records

**Status**

Proposed

**Decision**

Store all design decisions using an ADR record in the /decisions directory. Discussion may happen across Github issues and pull requests related to documentation, code and ADRs themselves, but these discussions will be boiled down to the essentials of the decision and context to expedite onboarding of future members.

**Context**

*Architecture Decision Records (ADRs)* are a technique for capturing important architectural decisions along with their context and consequences stored in a version control system for the benefit of future team members as well as for external oversight. Storing important architectural decisions in version control systems are recommended as opposed to Wikis or a website so that they remain in sync with the code itself.

- See: https://www.thoughtworks.com/radar/techniques/lightweight-architecture-decision-records
- See: https://github.com/joelparkerhenderson/architecture_decision_record

**Consequences**

As views on certain topics change over time, it will be difficult without ADRs to reflect on the context for a decision as it changes with respect to the code. In the past, this meant wading through Github issues that speider web off into different directions to related it all together, which made it difficult to follow the progression of thought across a single idea.
How often are they updated?
Latest Release: v1.3

github.com/ESIPFed/science-on-schema.org/releases/tag/1.3.0

DOI 10.5281/zenodo.6502539

- Dataset Funding (Awards & Organizations)
- Dataset Variables (use of external vocabularies)
- Geologic time for temporal coverage
- Preserve item order (authors & contributors, etc.)
- More JSON-LD examples
- SHACL Validation

v1.3.0

RELEASE NOTES for 1.3.0

FIXES
- Namespace Consistency (decision, Issue #52)
- Moved experimental recommendations to a separate document

IMPROVEMENTS
- Describing dataset funding (decision, Issue #109)
- Describing dataset variables (decision, Issue #27)
- Improving temporal coverage of a dataset (decision, Issue #77)
- Preserving item order (Issue #135)
- Describe need for a sitemap (Issue #192)
- New JSON-LD examples
- Updated SHACL validation shape

NEW FEATURES
- Representing checksums as identifiers (decision, Issue #66)
Get Involved

**Telecons:**
- 4th Thursday at 2:30pm ET/ 11:30am PT

*Next Telecon: Monday, April 5th, 5pmET ([Zoom](#))*

**Mailing List:** esip-schema-dot-org@lists.esipfed.org

**Slack:** [https://esip-all.slack.com/messages/C8EACC8KT](https://esip-all.slack.com/messages/C8EACC8KT)

**Twitter:** [@ScienceOnSchema](https://twitter.com/ScienceOnSchema)
What adoption challenges do science data have?
Challenges

1. Temporal & spatial representations
2. Describing other related entities
   a. Deployments, Instrumentation & Vehicles, oh my!
   b. Methodologies
What have been the outcomes?
Google Dataset Search

Google Dataset Search By the Numbers

Google Dataset Search: Building a search engine for datasets in an open Web ecosystem

datasetsearch.research.google.com
• DataONE
• EarthCube
• Polder
• MagIC
• IOC-UNESCO IODE
• NASA
• BCO-DMO
• Arctic Data Center
• Netherlands Polar Data Center
• Dryad
• Hydroshare
• and more…
Questions?