Adding Bioschemas markup to GitHub pages

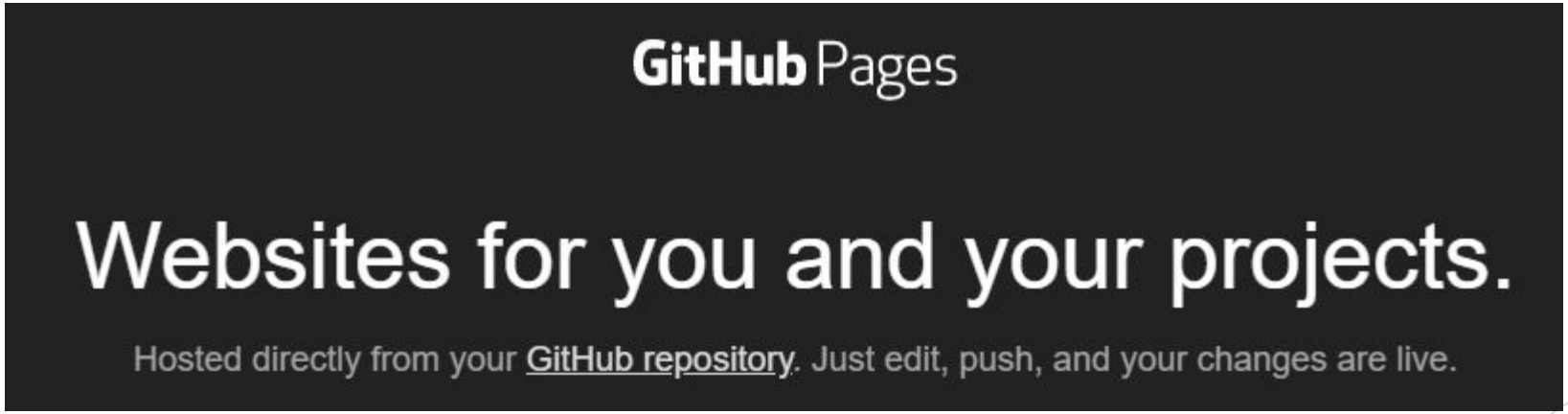
Leyla Jael Castro

Information Centre for Life Sciences
ZB MED
www.zbmed.de

elixir

Adding Bioschemas markup

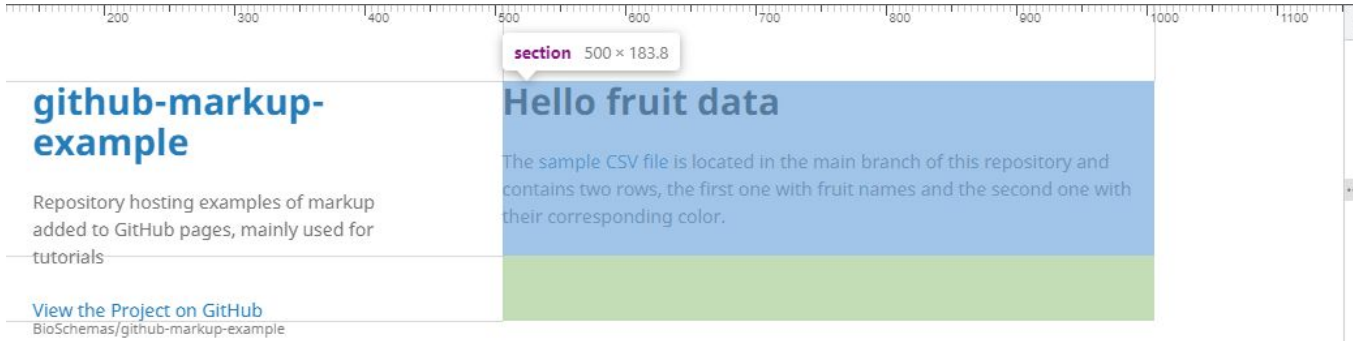
- We use and recommend markup in JSON-LD ▫ easy to read by machines and humans
- Markup can be added to any web page

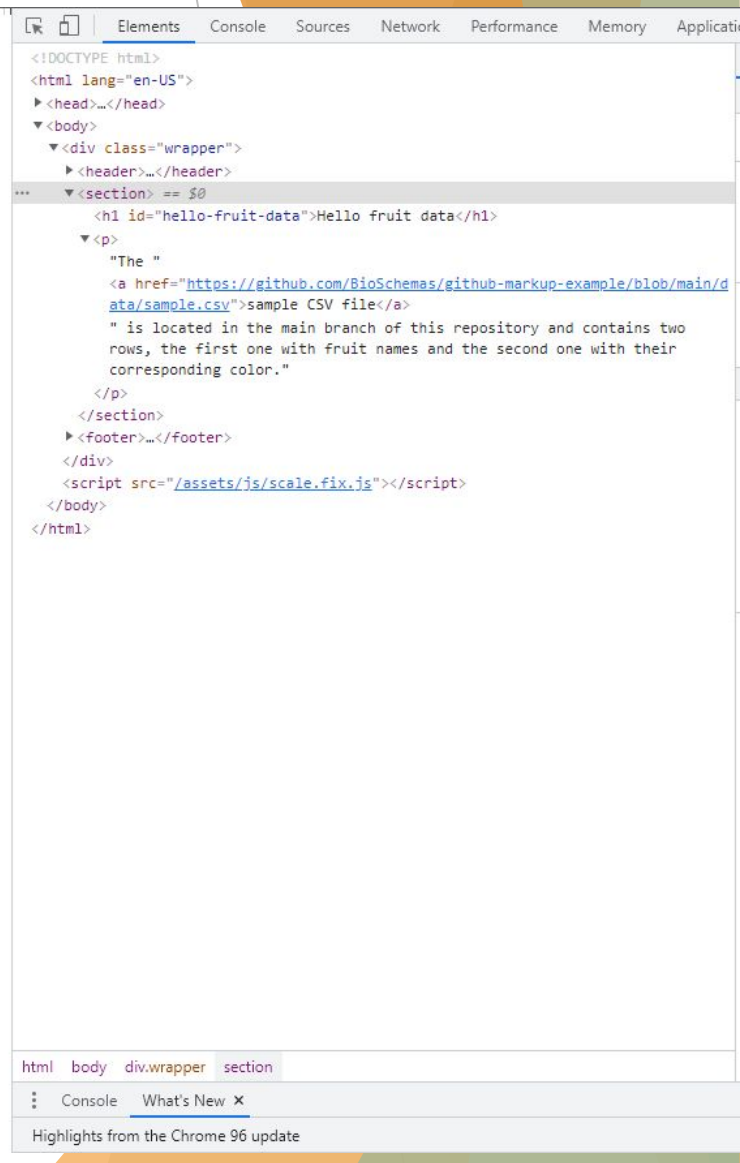


What can you expect from this lesson?

- Title: Adding Bioschemas markup to GitHub pages
- Keywords: Bioschemas, markup, GitHub pages, JSON-LD, Jekyll
- Description: In this lesson we will demo how to add Bioschemas markup to GitHub pages using Dataset (vr. Draft 0.4), ComputationalTool (vr. Release 1.0) and TrainingMaterial (vr. Draft 0.9) profiles
- Mentions:
  - This lesson reuses some information from an online Bioschemas tutorial
  - This lesson uses the repository (data, code and pages) at https://github.com/zbmed-semtec/bioschemas-github-markup-example
- Audience: This lesson is intended for anyone interested in Bioschemas markup
- Skills required:
  - How to use and run GitHub pages
  - Basic understanding of JSON-LD
  - Basic knowledge about Bioschemas (e.g., what a profile is)
- Learning outcomes:
  - Explaining different approaches to add Bioschemas markup to GitHub pages
  - Using Bioschemas markup on your own GitHub pages
  - Describing some of the minimum elements for the Bioschemas profiles Dataset, ComputationalTool and TrainingMaterial

Page without markup





Tasks of the day

Clone the repository

BioSchemas / github-markup-example Public

Select the branch

gh-pages-basic

Make sure your GitHub pages are running

Running?

No

Move to branch

gh-pages-with-markup

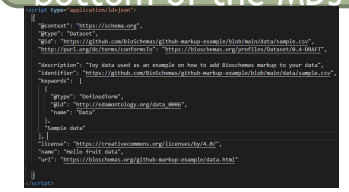
Look at

data.md

software.md

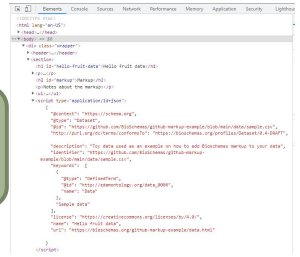
tutorial.md

Pay attention to JSON-LD on the bottom of the MDs



Refresh your pages and go to any of the ones with markup

Use the developer tools to see JSON-LD



Now move to branch

gh-pages-with-markup-and-layout

Look at

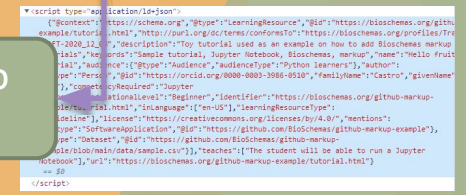
_layouts

tutorial.md

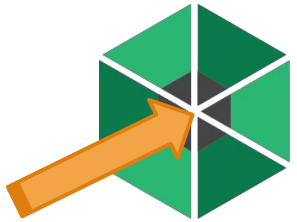
Refresh your pages and go to any the tutorial page

Look at the top, it has some markup data now

Look at the markup again



Try either way (with/out layout) on your own pages



Adding markup directly to the pages

Page with markup

github-markup-example

Repository hosting examples of markup added to GitHub pages, mainly used for tutorials

View the Project on GitHub
BioSchemas/github-markup-example

This project is maintained by
BioSchemas

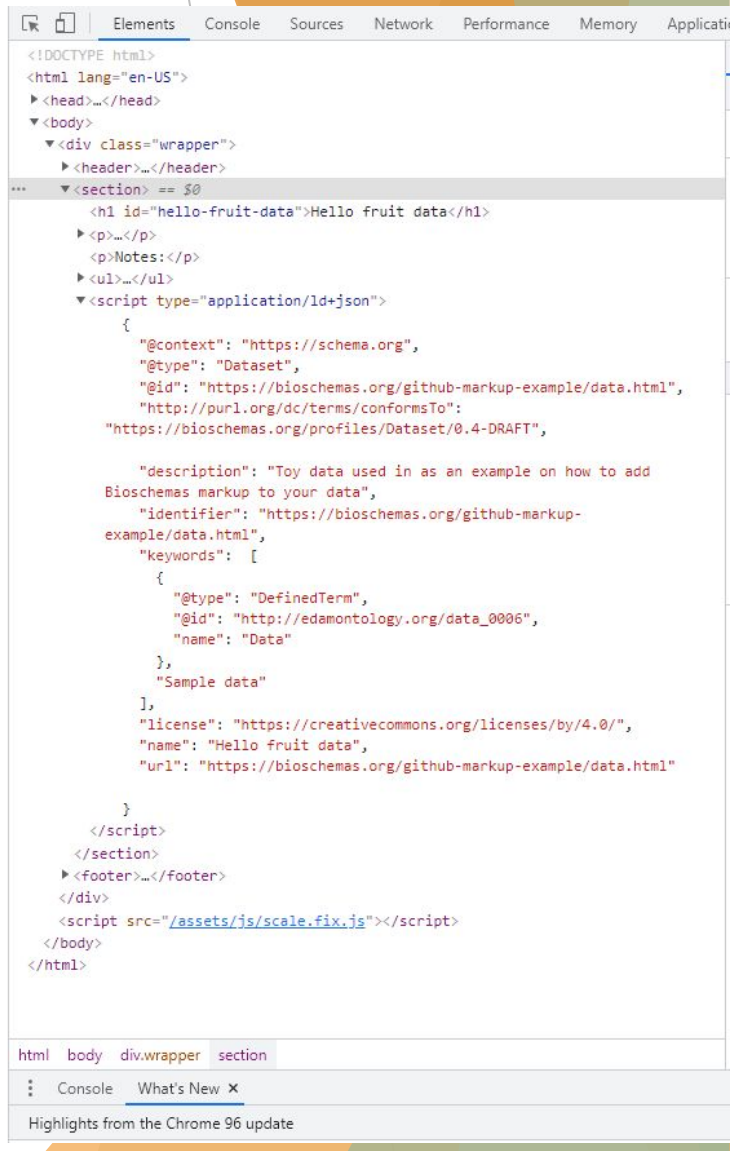
Hosted on GitHub Pages — Theme by orderedlist

Hello fruit data

The sample CSV file is located in the main branch of this repository and contains two rows, the first one with fruit names and the second one with their corresponding color.

Notes:

- We advised you to deposit your data on a data repository that will provide you with a Permanent Identifier (PID) such as a Digital Object Identifer (DOI). We have not done so as we are using toy data so we use the URL as the dataset identifier rather than a proper ID.
- We provide markup only for those properties defined as *minimum* by Bioschemas Dataset profile (version Draft 0.4) but you can improve your dataset description adding *recommended* and *optional* properties.
- Whenever you use an EDAM term to describe your data, try and find the one which describes it best. Here we use "Data" as an example.



Dataset DRAFT Profile

Version: 0.4-DRAFT (30 March 2021)

If you spot any errors or omissions with this type, please file an issue in our GitHub.

Description | Contributors | Links

Schema.org hierarchy

This Profile fits into the schema.org hierarchy as follows:

Thing > CreativeWork > Dataset

```
<script type="application/ld+json">
  {
    "@context": "https://schema.org",
    "@type": "Dataset",
    "@id": "https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv",
    "http://purl.org/dc/terms/conformsTo": "https://bioschemas.org/profiles/Dataset/0.4-DRAFT",

    "description": "Toy data used as an example on how to add Bioschemas markup to your data",
    "identifier": "https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv",
    "keywords":  [
      {
        "@type": "DefinedTerm",
        "@id": "http://edamontology.org/data_0006",
        "name": "Data"
      },
      "Sample data"
    ],
    "license": "https://creativecommons.org/licenses/by/4.0/",
    "name": "Hello fruit data",
    "url": "https://bioschemas.org/github-markup-example/data.html"

  }
</script>
```

description	Text	Schema: A description of the item. Bioschemas: A short summary describing a dataset.	ONE	
identifier	PropertyValue Text URL	Schema: The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details. Bioschemas: CURIEs that can be resolved using Identifiers.org should be used.	MANY	
keywords	DefinedTerm Text URL	Schema: Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas. Bioschemas: Keywords should be drawn from a controlled vocabulary, e.g. EDAM, and supplied as a DefinedTerm list.	ONE	
license	CreativeWork URL	Schema: A license document that applies to this content, typically indicated by URL. Bioschemas: A license under which the dataset is distributed.	ONE	
name	Text	Schema: The name of the item. Bioschemas: A descriptive name of the dataset.	ONE	
url	URL	Schema: URL of the item. Bioschemas: The location of a page describing the dataset.	ONE	

Marginality: Recommended.

ComputationalTool Profile

Version: 1.0-RELEASE (11 October 2021)

Bioschemas specification for describing a SoftwareApplication in the Life Sciences

If you spot any errors or omissions with this type, please file an issue in our GitHub.

Description | Contributors | Links

Schema.org hierarchy

This Profile fits into the schema.org hierarchy as follows:

Thing > CreativeWork > SoftwareApplication

```
<script type="application/ld+json">
  {
    "@context": "https://schema.org",
    "@type": "softwareApplication",
    "@id": "https://github.com/BioSchemas/github-markup-example",
    "http://purl.org/dc/terms/conformsTo": "https://bioschemas.org/profiles/ComputationalTool/1.0-RELEASE",

    "description": "Toy code used as an example on how to add Bioschemas markup to your software",
    "name": "Hello fruit tool",
    "url": "https://bioschemas.org/github-markup-example/software.html",

    "applicationCategory": [
      "Script"
    ],
    "applicationSubCategory":  [
      "Data management"
    ],
    "author": {
      "@type": "Person",
      "@id": "https://orcid.org/0000-0003-3986-0510",
      "familyName": "Castro",
      "givenName": "Leyla Jael"
    },
    "license": "https://opensource.org/licenses/MIT",
    "softwareVersion": "0.0.1"

  }
</script>
```

description	Text	**Schema:** A description of the item. **Bioschemas:** A short description of the tool.	ONE	
name	Text	**Schema:** The name of the item.		
url	URL	**Schema:** URL of the item. **Bioschemas:** Homepage of the tool.	ONE	
Marginality: Recommended.				
applicationCategory	Text URL	**Schema:** Type of software application, e.g. 'Game, Multimedia'. **Bioschemas:** Type of tool e.g. Command-line tool, Web application etc. **Note:** Bioschemas have changed URL to Text in the Expected Types. This will be reverted once Bio.Tools provides stable URIs for tool types.	MANY	Please use terms from the 'Tool type' table in the biotools documentation
applicationSubCategory	Text URL	**Schema:** Subcategory of the application, e.g. 'Arcade Game'. **Bioschemas:** Use an EDAM:Topic to describe the category of application	MANY	EDAM:Topic
author	Organization Person	**Schema:** The author of this content or rating. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.	MANY	
citation	CreativeWork URL	**Schema:** A citation or reference to another creative work, such as another publication, web page, scholarly article, etc. **Bioschemas:** Publication about this tool.	MANY	

TrainingMaterial DRAFT Profile

Version: 0.9-DRAFT-2020_12_08 (08 December 2020)

A specification for describing training materials in life sciences.

If you spot any errors or omissions with this type, please file an issue in our GitHub.

Description | Contributors | Links

Schema.org hierarchy

This Profile fits into the schema.org hierarchy as follows:

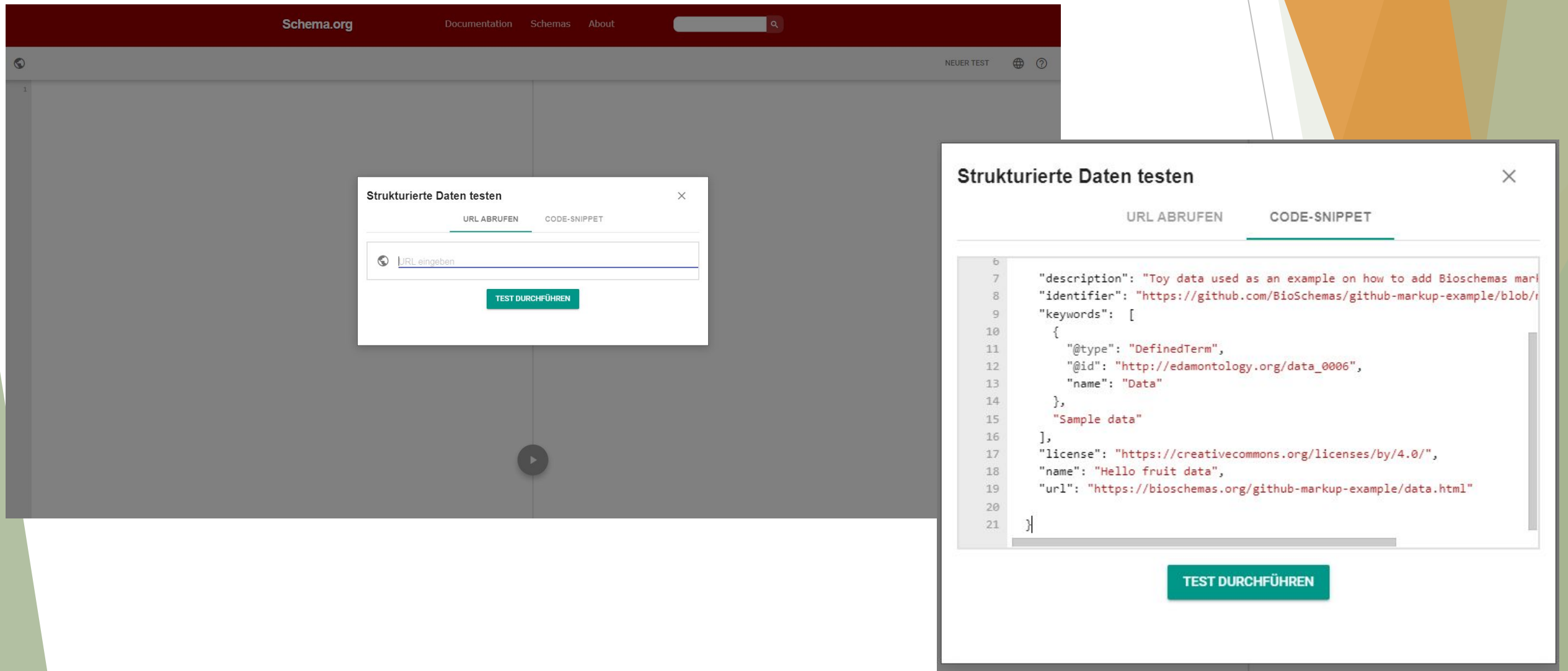
Thing > CreativeWork > LearningResource

```
<script type="application/ld+json">
 {
  "@context": "https://schema.org",
  "@type": "LearningResource",
  "@id": "https://bioschemas.org/github-markup-example/tutorial.html",
  "http://purl.org/dc/terms/conformsTo": "https://bioschemas.org/profiles/TrainingMaterial/0.9-DRAFT-2020_12_08",

  "description": "Toy tutorial used as an example on how to add Bioschemas markup to your tutorials",
  "keywords": "Sample tutorial, Jupyter Notebook, Bioschemas, markup",
  "name": "Hello fruit tutorial",

  "audience": {
    "@type": "Audience",
    "audienceType": "Python learners"
  },
  "author": {
    "@type": "Person",
    "@id": "https://orcid.org/0000-0003-3986-0510",
    "familyName": "Castro",
    "givenName": "Leyla Jael"
  },
  "competencyRequired": "Jupyter notebooks",
  "educationalLevel": "Beginner",
  "identifier": "https://bioschemas.org/github-markup-example/tutorial.html",
  "inLanguage": ["en-US"],
  "learningResourceType": ["guideline"],
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "mentions": [
    {
      "@type": "SoftwareApplication",
      "@id": "https://github.com/BioSchemas/github-markup-example"
    },{
      "@type": "Dataset",
      "@id": "https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv"
    }
  ],
  "teaches": [
    "The student will be able to run a Jupyter Notebook"
  ],
  "url": "https://bioschemas.org/github-markup-example/tutorial.html"

 }
</script>
```

description	Text	Schema: A description of the item.	ONE		
keywords	DefinedTerm Text URL	Schema: Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas.	MANY		
name	Text	Schema: The name of the item.	ONE		
Marginality: Recommended.					
about	Thing	Schema: The subject matter of the content. Inverse property: subjectOf. Bioschemas: The subject of this Training Material. Use the DefinedTerm type to add a controlled vocabulary term to describe the topic (such as from the EDAM ontology)	MANY	EDAM Topic Ontology	
abstract	Text	Schema: An abstract is a short description that summarizes a CreativeWork.	ONE		
audience	Audience	Schema: An intended audience, i.e. a group for whom something was created. Supersedes serviceAudience. Bioschemas: A succinct description of the intended target audience for your materials: e.g., graduates, postgraduates, clinicians.	MANY		
author	Organization Person	Schema: The author of this content or rating. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably. Bioschemas: Those involved in the preparation, creation and/or presentation of the published work, specifically writing the initial draft	MANY		
competencyRequired	DefinedTerm	Schema:	MANY		

Double check your markup with a validator

https://validator.schema.org/



So far so good

Schema.org Documentation Schemas About

NEUER TEST

```
{
   "@context": "https://schema.org",
   "@type": "Dataset",
   "@id": "https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv",
   "http://purl.org/dc/terms/conformsTo": "https://bioschemas.org/profiles/Dataset/0.4-DRAFT",
   "description": "Toy data used as an example on how to add Bioschemas markup to your data",
   "identifier": "https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv",
   "keywords":  [
     {
       "@type": "DefinedTerm",
       "@id": "http://edamontology.org/data_0006",
       "name": "Data"
     },
     "Sample data"
   ],
   "license": "https://creativecommons.org/licenses/by/4.0/",
   "name": "Hello fruit data",
   "url": "https://bioschemas.org/github-markup-example/data.html"
 }
```

Dataset

All (1)

Dataset

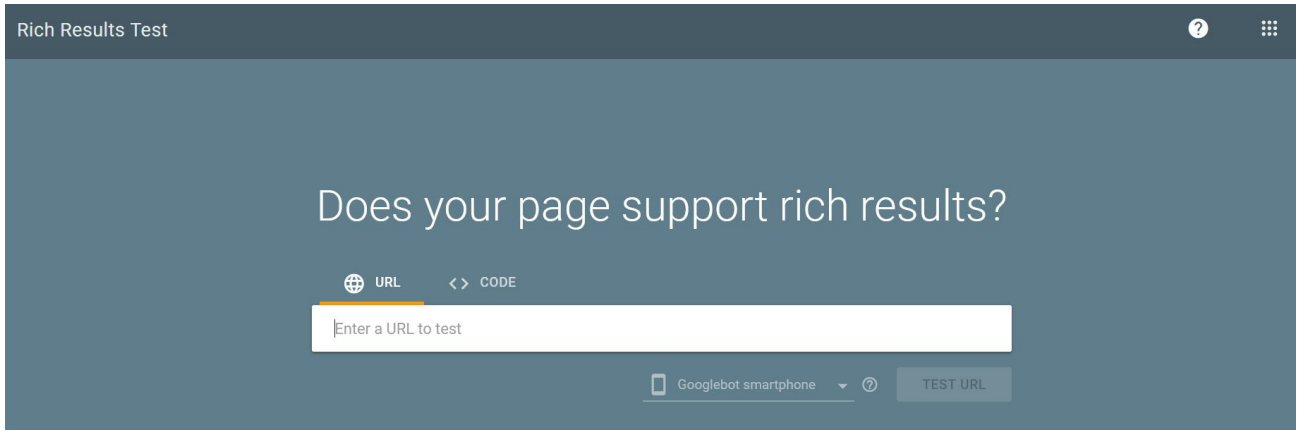
0 FEHLER 0 WARNUNGEN

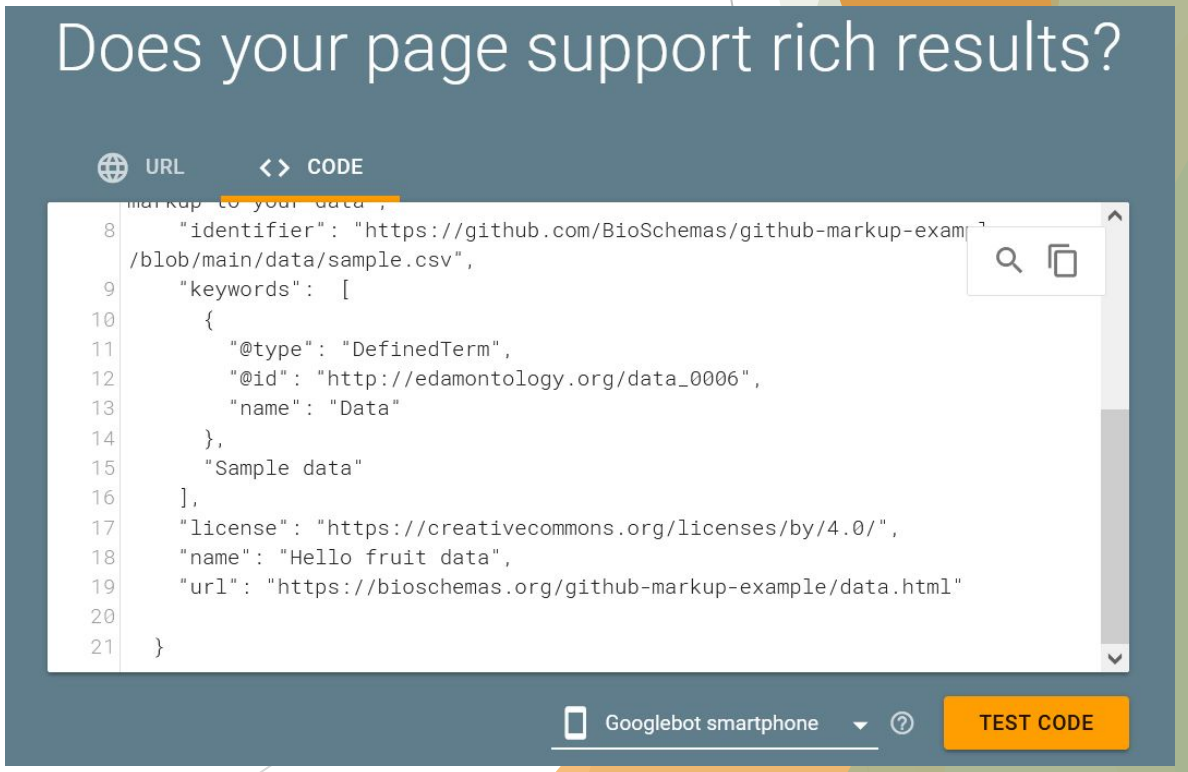
ID: https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv

@type	Dataset
@id	https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv
http://purl.org/dc/terms/conformsTo	https://bioschemas.org/profiles/Dataset/0.4-DRAFT
description	Toy data used as an example on how to add Bioschemas markup to your data
identifier	https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv
keywords	Sample data
license	https://creativecommons.org/licenses/by/4.0/
name	Hello fruit data
url	https://bioschemas.org/github-markup-example/data.html
keywords	
@type	DefinedTerm
@id	http://edamontology.org/data_0006
name	Data

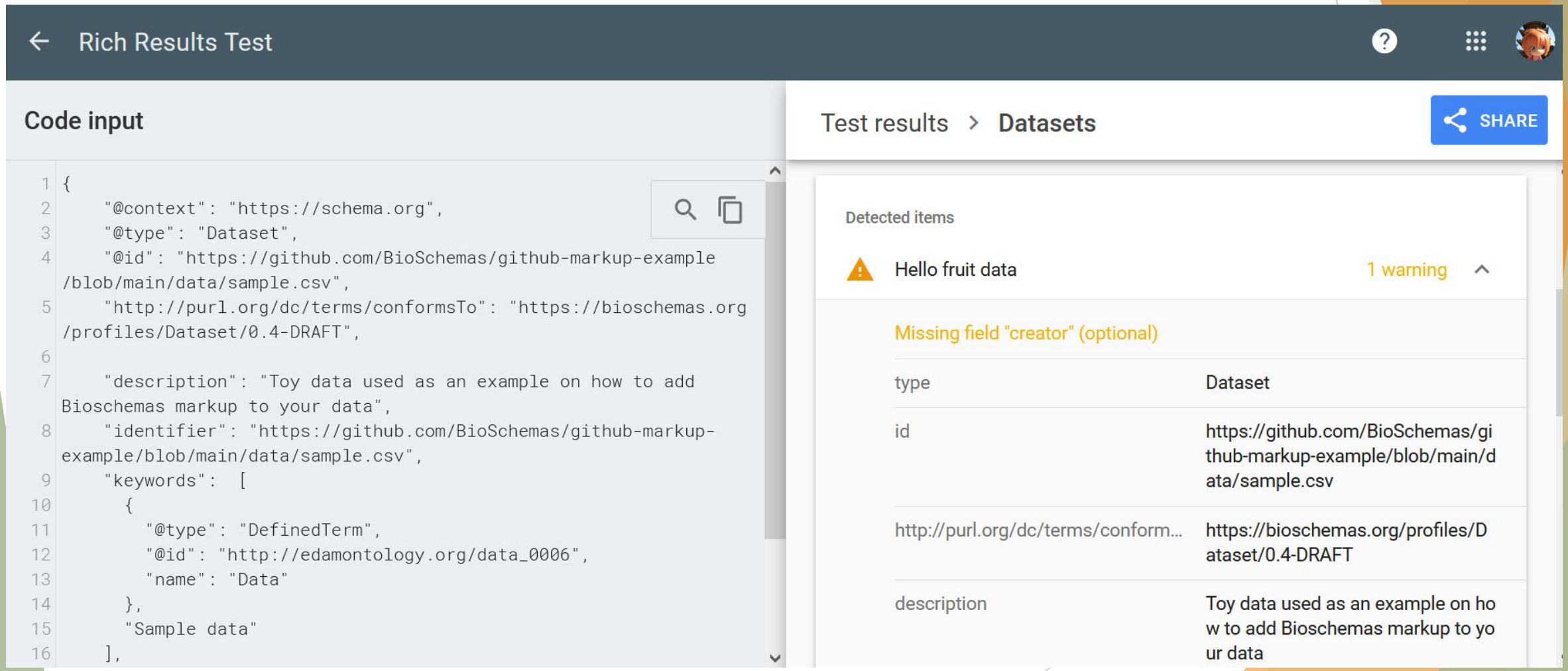
Double check your markup with a validator

https://search.google.com/test/rich-results



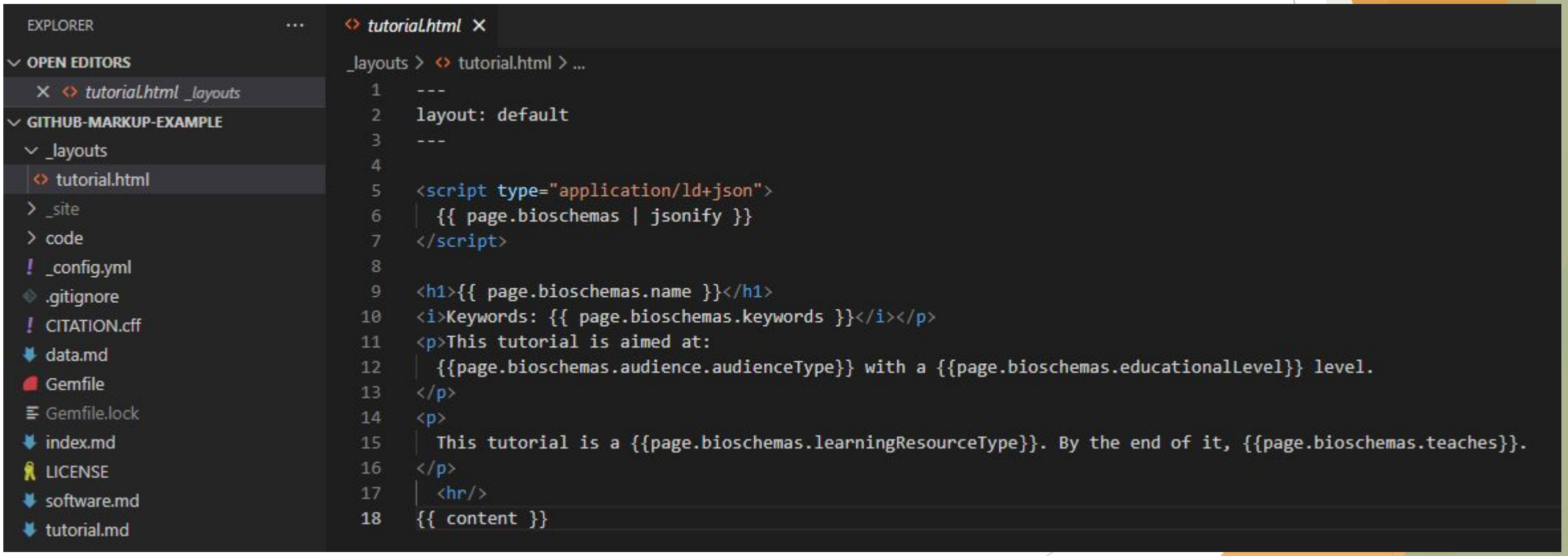


This one is a bit more strict but still we are good



Adding markup and rendering data via layouts

Adding a layout



Adding Bioschemas markup

Hello fruit tutorial

Keywords: Sample tutorial, Jupyter Notebook, Bioschemas, markup

This tutorial is aimed at: Python learners with a Beginner level.

This tutorial is a guideline. By the end of it, The student will be able to run a Jupyter Notebook.

Getting started

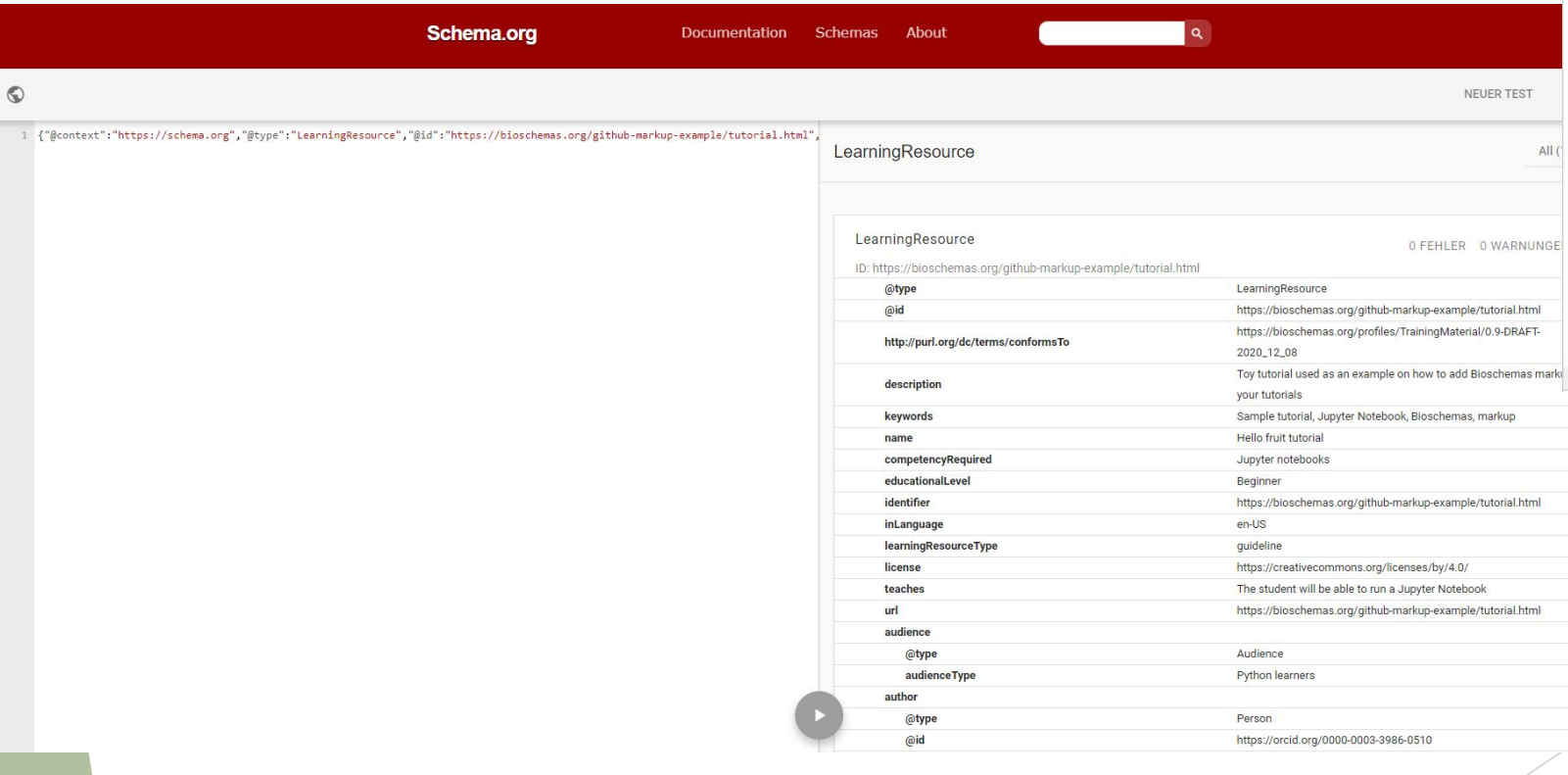
In this tutorial, you will learn how to use the Hello Fruit code and get a warm fruty and colorful greeting. We use Jupyter Notebooks so it is an advantage if you are familiar with it. If you have not installed yet Jupyter Notebooks, please do so.

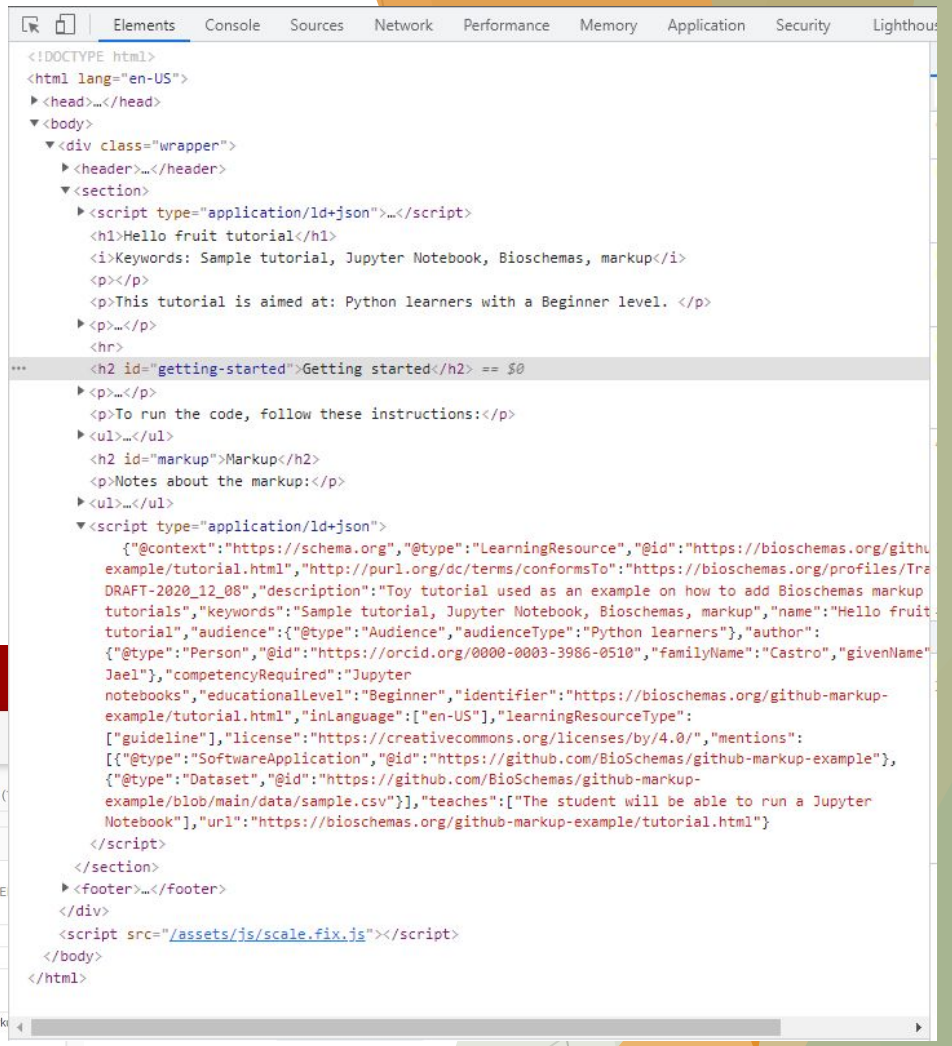
To run the code, follow these instructions:

- Clone the repository at https://github.com/BioSchemas/github-markup-example, we will use the main branch
- Init Jupyter Notebooks and make sure you can access the folder where you cloned de repo
- Open the notebook named "hello_fruit" under folder *code*
- Run the cells one by one to get a fuity colorful greeting

```
---
layout: tutorial
bioschemas:
  "@context": https://schema.org
  "@type": LearningResource
  "@id": https://bioschemas.org/github-markup-example/tutorial.html
  http://purl.org/dc/terms/conformsTo: https://bioschemas.org/profiles/TrainingMaterial/0.9-DRAFT-2020_12_08
  description: Toy tutorial used as an example on how to add Bioschemas markup to your
    tutorials
  keywords: Sample tutorial, Jupyter Notebook, Bioschemas, markup
  name: Hello fruit tutorial
  audience:
    "@type": Audience
    audienceType: Python learners
  author:
    "@type": Person
    "@id": https://orcid.org/0000-0003-3986-0510
    familyName: Castro
    givenName: Leyla Jael
  competencyRequired: Jupyter notebooks
  educationalLevel: Beginner
  identifier: https://bioschemas.org/github-markup-example/tutorial.html
  inLanguage:
  - en-US
  learningResourceType:
  - guideline
  license: https://creativecommons.org/licenses/by/4.0/
  mentions:
  - "@type": SoftwareApplication
    "@id": https://github.com/BioSchemas/github-markup-example
  - "@type": Dataset
    "@id": https://github.com/BioSchemas/github-markup-example/blob/main/data/sample.csv
  teaches:
  - The student will be able to run a Jupyter Notebook
  url: https://bioschemas.org/github-markup-example/tutorial.html

---
```

Rendered markup

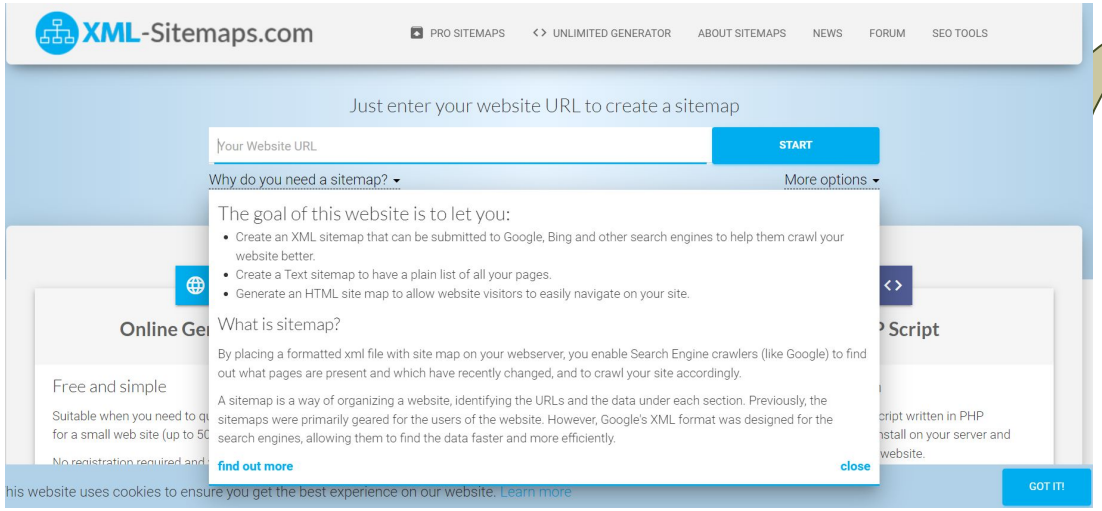




Make it easier for machines

Sitemap.xml and robots.txt

https://www.xml-sitemaps.com/
https://developers.google.com/search/docs/advanced/sitemaps/build-sitemap



sitemap.xml

```xml
<?xml version="1.0" encoding="UTF-8"?>
<urlset
      xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
            http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
<!-- created with Free Online Sitemap Generator www.xml-sitemaps.com -->

<url>
  <loc>https://zbmed-semtec.github.io/bioschemas-github-markup-example/</loc>
  <lastmod>2022-01-05T17:04:34+00:00</lastmod>
  <priority>1.00</priority>
</url>
<url>
  <loc>https://zbmed-semtec.github.io/bioschemas-github-markup-example/data.html</loc>
  <lastmod>2022-01-05T17:04:34+00:00</lastmod>
  <priority>0.80</priority>
</url>
<url>
  <loc>https://zbmed-semtec.github.io/bioschemas-github-markup-example/software.html</loc>
  <lastmod>2022-01-05T17:04:34+00:00</lastmod>
  <priority>0.80</priority>
</url>
<url>
  <loc>https://zbmed-semtec.github.io/bioschemas-github-markup-example/tutorial.html</loc>
  <lastmod>2022-01-05T17:04:34+00:00</lastmod>
  <priority>0.80</priority>
</url>
```

robots.txt

```
Sitemap: https://zbmed-semtec.github.io/bioschemas-github-markup-example/sitemap.xml
```

Time to add markup to your own pages

Questions?

- Website: bioschemas.org
- Mailing list: public-bioschemas@w3.org
- Slack: http://tiny.cc/bs-slack
- Example markup: https://github.com/BioSchemas/specifications/
- Live deployments: https://bioschemas.org/liveDeploys/

Information Centre for Life Sciences
ZB MED
www.zbmed.de