

FAIR datasets with Bioschemas

Tutorial: FAIRify a dataset using just enough metadata

Dataset profile in Bioschemas

<https://bioschemas.org/profiles/Dataset/1.0-RELEASE>



@context	URL	Used to provide the context (namespaces) for the JSON-LD file. Not needed in other serialisations.	ONE	
@type	Text	Schema.org/Bioschemas class for the resource declared using JSON-LD syntax. For other serialisations please use the appropriate mechanism. While it is permissible to provide multiple types, it is preferred to use a single type.	MANY	Schema.org, Bioschemas
@id	IRI	Used to distinguish the resource being described in JSON-LD. For other serialisations use the appropriate approach.	ONE	
dct:conformsTo	IRI	Used to state the Bioschemas profile that the markup relates to. The versioned URL of the profile must be used. Note that we use a CURIE in the table here but the full URL for Dublin Core terms must be used in the markup (http://purl.org/dc/terms/conformsTo), see example.	ONE	Bioschemas profile versioned URL

Property	Expected Type	Description	CD	Controlled Vocabulary
Marginality: Minimum.				
description	Text	<p>Schema: A description of the item.</p> <p>Bioschemas: A short summary describing a dataset.</p>	ONE	
identifier	PropertyValue	<p>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.</p> <p>Bioschemas: CURIEs that can be resolved using Identifiers.org should be used.</p>	MANY	
keywords	DefinedTerm	<p>Schema: Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas.</p>	MANY	





Property	Expected Type	Description	CD	Controlled Vocabulary
Marginality: Minimum.				
license	CreativeWork URL	<p>Schema: A license document that applies to this content, typically indicated by URL.</p> <p>Bioschemas: A license under which the dataset is distributed.</p>	ONE	
name	Text	<p>Schema: The name of the item.</p> <p>Bioschemas: A descriptive name of the dataset.</p>	ONE	
url	URL	<p>Schema: URL of the item.</p> <p>Bioschemas: The location of a page describing the dataset.</p>	ONE	

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.

Tutorial to add Bioschemas markup to GitHub pages

- [Repository](#)
- [Rendered pages](#)
- [Full tutorial](#) (including also validation, harvesting, deployment in Bioschemas)

8 lines (8 sloc) 103 Bytes		Raw	Blame				
🔍 Search this file...							
1	fruit	color					
2	apple	green					
3	apple	red					
4	apple	pink					
5	pear	yellow					
6	pineapple	yellow					
7	strawberry	red					
8	blueberry	blue					

Minimum descriptors for the Fruit Dataset

Property	Value	Additional info
description	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.	
identifier	https://github.com/zbmed-semtec/bioschemas-github-markup-example/blob/main/data/sample.csv	A PID would be better, e.g., DOI assigned by data repository
keywords	CSV, sample data, data	EDAM terms would be a good option
license	CC-By 4.0	A machine-readable version is a good option, e.g., SPDX
name	Hello Fruit Dataset	
url	https://zbmed-semtec.github.io/bioschemas-github-markup-example/data.html	

Minimum structured metadata for the Fruit Dataset

```
{
  "@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": {
    "@type": "CreativeWork",
    "@id": "https://bioschemas.org/profiles/Dataset/1.0-RELEASE"
  },
  "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": [
    "CSV", "Sample data",
    {
      "@type": "DefinedTerm",
      "@id": "http://edamontology.org/data_0006",
      "name": "Data"
    }
  ],
  "license": "https://spdx.org/licenses/CC-BY-4.0.html",
  "name": "Hello fruit data",
  "url": "https://bioschemas.org/github-markup-example/data.html"
}
```



Hands-on

Bring your own data

- Choose a dataset you want to describe
- Create a table with the descriptors
- Write the corresponding JSON-LD
- [Add markup to GitHub Pages](#)



OPEN MEET. SHARE. INSPIRE. CARE.
SCIENCE
FESTIVAL

#OSF2023DE

Thank you!