A very brief introduction to making metadata with JSON-LD

Stian Soiland-Reyes

The University of Manchester

RO-Crate co-lead

soiland-reyes@manchester.ac.uk

https://orcid.org/0000-0001-9842-9718

MANCHESTER
1824
The University of Manchester

Workshop: *Data exchange with RO-Crates and Knowledge Graphs*
Open Science Festival 2023, ZB Med, Köln, DE
2023-07-06

JSON for Linking Data

Data is messy and disconnected. JSON-LD organizes and connects it, creating a better Web.

https://json-ld.org/

⇄ Linked Data

Linked Data empowers people that publish and use information on the Web. It is a way to create a network of standards-based, machine-readable data across Web sites. It allows an application to start at one piece of Linked Data, and follow embedded links to other pieces of Linked Data that are hosted on different sites across the Web.

A Simple Example

```json
{
  "@context": "http://json-ld.org/contexts/person.jsonld",
  "@id": "http://dbpedia.org/resource/John_Lennon",
  "name": "John Lennon",
  "born": "1940-10-09",
  "spouse": "http://dbpedia.org/resource/Cynthia_Lennon"
}
```

{⋄} JSON-LD

JSON-LD is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for programming environments, REST Web services, and unstructured databases such as CouchDB and MongoDB.

https://www.w3.org/TR/json-ld/

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

  "@type": "LearningResource",

  "http://purl.org/dc/terms/conformsTo": {

    "@type": "CreativeWork",

    "@id": "https://bioschemas.org/profiles/TrainingMaterial/1.0-RELEASE"

  },

    "name": "Adding nanomaterial data",

    "version": "0.9.3",

    "description": "This tutorial describes how nanomaterial data can
be added to an eNanoMapper server using a RDF format.",

    "license": "https://creativecommons.org/licenses/by/4.0/",

    "keywords": "ontologies, enanomapper, RDF",

    "url": "https://nanocommons.github.io/tutorials/enteringData/",

    "provider": {

      "@type": "Organization",

      "name": "NanoCommons",

      "url": "https://www.nanocommons.eu/"

    },

  "…": {}
```

JSON-LD **context**

Type (identifier implied)

Bioschemas **profile**

Metadata about training
material

Nested object
with its own @type and
attributes

https://nanocommons.github.io/tutorials/enteringData/
by Egon Willighagen

```json
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [
```

```json
    { "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
      "about": { "@id": "./" }
    }
```

RO-Crate **metadata file** descriptor

```json
    { "@id": "./",
      "identifier": "https://doi.org/10.5281/zenodo.1009240",
      "@type": "Dataset",

      "hasPart": [
        { "@id": "cp7glop.ai" },
        { "@id": "lots_of_little_files/" },
        { "@id": "communities-2018.csv" },
        { "@id": "https://doi.org/10.4225/59/59672c09f4a4b" },
        { "@id": "SciDataCon-Presentations/AAA_Pilot_Abstract.html"}
      ],

      "author": { "@id": "https://orcid.org/0000-0002-8367-6908" },
      "publisher": { "@id": "https://ror.org/03f0f6041" },
      "citation": { "@id": "https://doi.org/10.1109/TCYB.2014.2386282"},
      "name": "Presentation of user survey 2018"
    },
```

RO-Crate **root** dataset

..collection of **Data entities**

..described w/ **contextual entities**

```json
    { "@id": "cp7glop.ai",
      "@type": "File",
      "name": "Diagram showing trend to increase",
      …
    },
    …
```

Flat list of metadata per entity

4

RO-Crate Metadata File

```json
{
  "@id": "figure.png",
  "@type": ["File", "ImageObject"],
  "name": "XXL-CT-scan of an XXL Tyrannosaurus rex skull",
  "identifier": "https://doi.org/10.5281/zenodo.3479743",
  "author": {"@id": "https://orcid.org/0000-0002-8367-6908"},
  "encodingFormat": "image/png"
}
```

Metadata

```json
{
  "@id": "https://orcid.org/0000-0002-8367-6908",
  "@type": "Person",
  "affiliation": { "@id": "https://ror.org/03f0f6041" },
  "name": "J. Xuan"
}
```

```json
{
  "@id": "https://ror.org/03f0f6041",
  "@type": "Organization",
  "name": "University of Technology Sydney",
  "url": "https://www.uts.edu.au/"
}
```

Linked Data: Reference by URI

Types and properties are expanded by context, e.g. http://schema.org/ImageObject

Entities can be **cross-referenced** with @id within the same JSON-LD document

Style *Flattened JSON-LD:* each entity is listed separately in @graph array. @id required

Style *Compacted JSON-LD:* the entity can be nested within any cross-reference, @id optional

Clients can still follow the links for potentially more data (e.g. ORCID lists publications)

5

Using common vocabularies

.. extending only when needed

Docs   Schemas   Validate   About

Dataset
A Schema.org Type

Thing > CreativeWork > Dataset

[more...]

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

Property	Expected Type	Description
Properties from Dataset		
distribution	DataDownload	A downloadable form of this dataset, at a specific location, in a specific format. This property can be repeated if different variations are available. There is no expectation that different downloadable distributions must contain exactly equivalent information (see also DCAT on this point). Different distributions might include or exclude different subsets of the entire dataset, for example.
includedInDataCatalog	DataCatalog	A data catalog which contains this dataset. Supersedes catalog, includedDataCatalog. Inverse property: dataset
issn	Text	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.
measurementTechnique	Text or URL	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 dataset 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: "mass spectrometry" or "nmr spectroscopy" or "colorimetry" or "immunofluorescence".

If the variableMeasured is "depression rating", the measurementTechnique could be "Zung Scale" or "HAM-D" or "Beck Depression Inventory".

If there are several variableMeasured properties recorded for some given data object, use a PropertyValue for each variableMeasured and attach the corresponding measurementTechnique. |
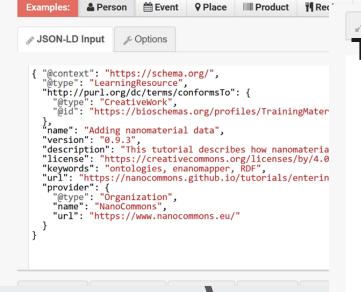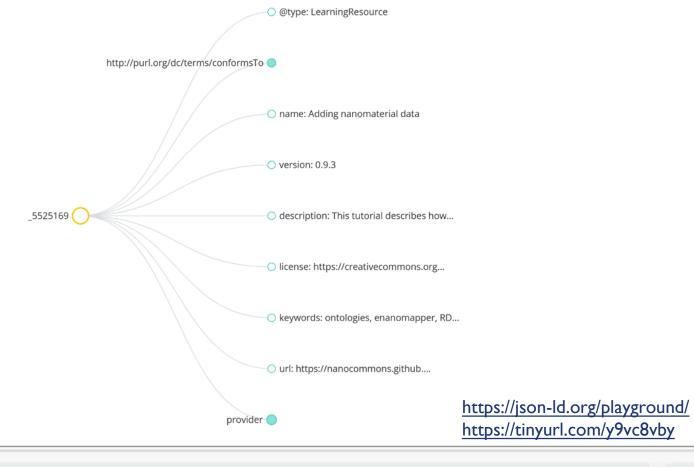| variableMeasured | PropertyValue or Text | The variableMeasured property can indicate (repeated as necessary) the variables that are measured in some dataset, either described as text or as pairs of identifier and description using PropertyValue. |
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: subjectOf
abstract	Text	An abstract is a short description that summarizes a CreativeWork.
accessMode	Text	The human sensory perceptual system or cognitive faculty through which a person may process or perceive information. Values should be drawn from the approved vocabulary.
accessModeSufficient	ItemList	A list of single or combined accessModes that are sufficient to understand all the intellectual

https://schema.org/Dataset
is based on W3C DCAT

7

Testing in the JSON-LD Playground

https://validator.schema.org/

OPEN SCIENCE FESTIVAL

MEET. SHARE. INSPIRE. CARE.

#OSF2023DE

Stian Soiland-Reyes
The University of Manchester
RO-Crate community co-chair

Thank you!