

JSON for Linking Data



Dataset-JSON as the compacted form of a graph

Goal - show why JSON is the next-gen clinical data format

- FAIR (Findable, Accessible, Interoperable, Reusable)
- unambiguous, self-descriptive
- support Biomedical Concepts
- simple, lightweight

Summary

- Dataset-JSON is contextualised by JSON-LD into explicit RDF
- ODM expressed as a JSON-LD context provides a precise definition of each element in a JSON dataset in terms of web-compatible linked data
- Define-LD served online in ‘neutral zone’ provides a single source of metadata truth, a data contract between all trial data stakeholders

JSON-LD serialises linked data as JSON



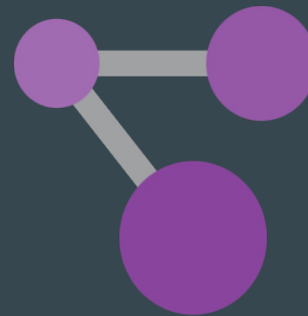
compacted

+



@context

=



expanded

JSON-LD applied to Dataset-JSON



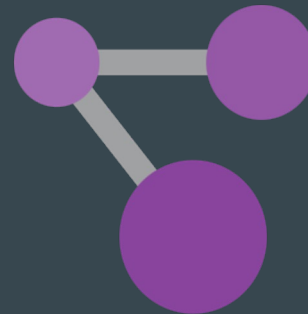
CDISC Data

+



Dataset-JSON
context

=



Semantic
Web content

This simple Dataset-JSON example becomes JSON-LD by adding @context

```
{
  "@context": "http://localhost:4000/transfer_104ab4/manifest#",
  "clinicalData": {
    "studyOID": "BS1234",
    "metaDataVersionOID": "v2",
    "itemGroupData": {
      "IG.DM": {
        "records": 600,
        "name": "DM",
        "label": "Demographics",
        "items": [
          {"OID": "ITEMGROUPDATASEQ", "name": "ITEMGROUPDATASEQ", "label": "Record identifier", "type": "integer"},
          {"OID": "IT.STUDYID", "name": "STUDYID", "label": "Study identifier", "type": "string", "length": 7},
          {"OID": "IT.USUBJID", "name": "USUBJID", "label": "Unique Subject Identifier", "type": "string", "length": 3},
          {"OID": "IT.DOMAIN", "name": "DOMAIN", "label": "Domain Identifier", "type": "string", "length": 2},
          {"OID": "IT.AGE", "name": "AGE", "label": "Subject Age", "type": "integer", "length": 2}
        ],
        "itemData": [
          [1, "MyStudy", "001", "DM", 56],
          [2, "MyStudy", "002", "DM", 26]
        ]
      }
    }
  }
}
```

JSON-LD context can reference other contexts e.g. explicit link to Define as a context

```
{
  "@context": [
    "http://localhost:4000/manifest#",
    { "@vocab": "http://localhost:4000/transfer_104ab4/define_BA1234_v2#" }
  ],

  "manifest": {
    "fileType": "Dataset-JSON",
    "fileOID": "transfer_104ab4",
    "priorFileOID": "transfer_25b200",
    "creationDateTime": "2012-04-23T18:25:43.511Z",
    "asOfDateTime": "2012-04-22T00:00:01.511Z",
    "originator": "COSA Dataset-JSON Hackathon",
    "studyOID": "BS1234",
    "sourceSystem": "node server",
    "sourceSystemVersion": "124.51.52.5552",
    "datasetJsonVersion": "v0.1"
  }
}
```

JSON-LD context complements schema by describing *meaning*

```
{  
  "@context": {  
    "@version": 1.1,  
    "schema": "http://schema.org/",  
    "xsd": "http://www.w3.org/2001/XMLSchema#",  
    "def": "http://localhost:4000/define#",  
    "@vocab": "def:",  
  }
```

Shared definitions such as schema.org are used to define ‘what is this thing?’

```
    "itemGroupData": {  
      "@container": "@id",  
      "@id": "schema:Dataset"  
    },  
    "items": {  
      "@container": "@index",  
      "@id": "schema:VariableMeasured"  
    },  
    "itemData": {  
      "@container": "@list",  
      "@id": "schema>DataFeed"  
    },  
    "records": {  
      "@id": "schema:maxValue",  
      "@type": "xsd:nonNegativeInteger"  
    },  
  }
```

Context maps simple JSON fields to linked data

- Structure
- IDs
- Types

Expanded Dataset-JSON output as RDF

```
/define_BS1234_v2/IG.DM> <http://schema.org/DataFeed> _:c14n1 .  
/define_BS1234_v2/IG.DM> <http://schema.org/VariableMeasured> <http://localhost:4000/transfer_104ab4/define_BS1234_v2/IT.AGE> .  
/define_BS1234_v2/IG.DM> <http://schema.org/VariableMeasured> <http://localhost:4000/transfer_104ab4/define_BS1234_v2/IT.DOMAIN> .  
/define_BS1234_v2/IG.DM> <http://schema.org/VariableMeasured> <http://localhost:4000/transfer_104ab4/define_BS1234_v2/IT.STUDYID> .  
/define_BS1234_v2/IG.DM> <http://schema.org/VariableMeasured> <http://localhost:4000/transfer_104ab4/define_BS1234_v2/IT.USUBJID> .  
/define_BS1234_v2/IG.DM> <http://schema.org/VariableMeasured> <http://localhost:4000/transfer_104ab4/define_BS1234_v2/ITEMGROUPDATA> .  
/define_BS1234_v2/IG.DM> <http://schema.org/description> "Demographics"^^<http://schema.org/PropertyValue> .  
/define_BS1234_v2/IG.DM> <http://schema.org/maxLength> "600"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger> .  
/define_BS1234_v2/IG.DM> <http://schema.org/name> "DM"^^<http://schema.org/PropertyValue> .  
/define_BS1234_v2/IT.AGE> <http://schema.org/DataType> "integer"^^<http://schema.org/PropertyValue> .  
/define_BS1234_v2/IT.AGE> <http://schema.org/description> "Subject Age"^^<http://schema.org/PropertyValue> .  
/define_BS1234_v2/IT.AGE> <http://schema.org/name> "AGE"^^<http://schema.org/PropertyValue> .  
/define_BS1234_v2/IT.AGE> <http://www.w3.org/2001/XMLSchema#length> "2"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger> .
```

Dataset-JSON elements expanded to IRIs

RDF data ready to be loaded and queried as a graph

JSON-LD applied to Define



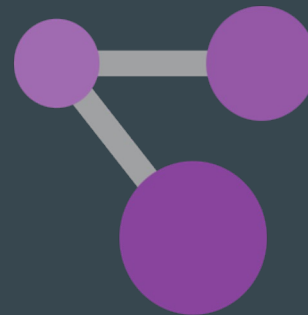
Define-JSON

+



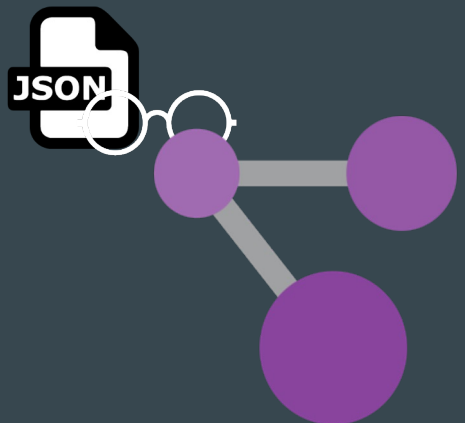
Define context

=



Linked
'Define-LD'
specification

Applications of a shared online Define



Define-LD
specification



Custom
document
structure

Searchable MDR
Single source of truth
Data contract / DTA

Metadata API

Browser-based
Spec UI

Streaming

JS object
Dataframe
Python dict

Automation (metadata re-use, metaprogramming)

Next concept: can JSON-LD link dataset *content* too?

Dataset-JSON <--> Biomedical Concepts



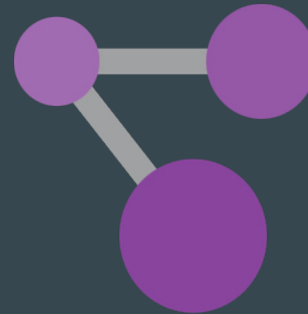
CDISC Dataset

+



JSON-LD
context built
from Define

=



Data in BC
metamodel

Summary

- Dataset-JSON is contextualised by JSON-LD into explicit RDF
- ODM expressed as a JSON-LD context provides a precise definition of each element in a JSON dataset in terms of web-compatible linked data
- Define-LD served online in ‘neutral zone’ provides a single source of metadata truth, a data contract between all trial data stakeholders