JSON for Linking Data

Dataset-JSON as the compacted form of a graph

Jeremy Teoh
8th December 2022

https://github.com/TeMeta/Dataset-JSON_hackathon
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

```json
{
    "@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

```json
{
    "@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*

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

Context maps simple JSON fields to linked data:

- Structure
- IDs
- Types
Expanded Dataset-JSON output as RDF

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
- JSON-LD framing
- Custom document structure

Automation (metadata re-use, metaprogramming)

Searchable MDR
Single source of truth
Data contract / DTA
Metadata API
Browser-based Spec UI
Streaming
JS object Dataframe Python dict
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