Simplified JSON-LD representations of linked data based on Allotrope Data Models

Jindřich Mynarz, Jan Rosecký, Vincent Antonucci, Jan Nešpor

R&D IT Data Infrastructure

September 20, 2021
Developer experience (DX) is user experience of developers, “focusing on perceptions of utility, ease of use, and efficiency”.

Good DX is seen as contributing to technology adoption and productivity, for example through gentle (or no) learning curve or faster ramp-up time.

DX is (largely) a property of interfaces through which developers work with a technology. For example, interfaces include APIs, programming languages, community support, or documentation.
Developer experience of Allotrope standards

Knowledge gap: unfamiliar data processing model. Black box binary data format.

Terms referred to via opaque identifiers (e.g., af-r:AFR_0002140) instead of human-readable names (e.g., qNMR purity result).

Verbose modelling patterns, for example due to BFO alignment.

Consequently, ADF data is difficult to understand and manipulate directly.

Can we expose a simpler interface for ADF data that can help developers with understanding and manipulating it indirectly while preserving the benefits of interoperability with the broader family of Allotrope standards?
JSON-LD: standing on the shoulders of standards

JSON-LD is a JSON syntax for RDF and an API for data manipulation standardized in W3C Recommendations.

It is interoperable with common data processing toolchains. JSON-LD API enables round-tripping back to other RDF serializations.

It preserves the key distinguishing feature of RDF: explicit links between data and metadata.

JSON-LD design goals and rationale explain that it aims to improve DX of semantic web standards.

It provides reach:

- to developers with limited knowledge of semantic web standards
- to tools that consume JSON
JSON-LD compaction reduces verbosity of data by referring to a (static) shared context of a data exchange.

The context defines local aliases or syntactical shortcuts that allow making data compact and unambiguous.

The context maps what data says to what it means. For example:

```json
{
   "@context": {
      "ADF": "Allotrope Data Format",
      "Allotrope Data Format": "https://docs.allotrope.org/Allotrope%20Data%20Format.html"
   }
}
```
JSON-LD framing

Graph ➞ Tree

JSON-LD framing coerces graph data into a hierarchical tree with a given structure.

It replaces links with nesting, allowing to access data by traversing rather than pattern-matching.

Makes data map to native data structures in popular programming languages (hash maps and arrays).

Not everyone wants to embed a SPARQL query engine in their application to navigate data.
Scope of work

Human-readable representation of **ADF Data Description** of simple **tabular** Allotrope data models.

AFO → JSON-LD context

ADMs → JSON-LD frames

A JSON-LD processor can use the context and the frames to produce JSON views of the ADF data for **straightforward consumption** by both people and applications.

The **running example** uses a fragment of test data for the qNMR model (Allotrope recommendation 2021/03) and shows some ways in which JSON-LD API can make ADF data easier to understand and manipulate.
ADF

A binary file format based on HDF5 that requires ADF API to be read.

- qnmr.hdf
  - data-cubes
  - data-description
    - dictionary
      - bytes
      - keys
      - nodes
  - index_GOSP
    - nodes
  - index_GPOS
    - nodes
  - index_GSPO
    - nodes
    - quads
HDF5 → RDF: ADF API maps HDF5 data structures for ADF Data Description to an RDF view that can be serialized as text in Turtle syntax.

```turtle
<urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d> a <http://purl.allotrope.org/ontologies/result#AFR_0002319> .


<urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5> a <http://purl.allotrope.org/ontologies/result#AFR_0002320> ; <http://purl.allotrope.org/ontologies/property#AFX_0002803> <urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36> ; <http://purl.obolibrary.org/obo/RO_0002350> <urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d> .
```
RDF/Turtle with namespace prefixes

@prefix af-r: <http://purl.allotrope.org/ontologies/result#> .
@prefix af-x: <http://purl.allotrope.org/ontologies/property#> .
@prefix obo: <http://purl.obolibrary.org/obo/> .
@prefix qudt: <http://qudt.org/schema/qudt#> .
@prefix unit: <http://qudt.org/vocab/unit#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d> a af-r:AFR_0002319 .

<urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36> a af-r:AFR_0002140 ;
  qudt:numericValue "58.40250856"^^xsd:double ;
  qudt:unit unit:Percent .

<urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5> a af-r:AFR_0002320 ;
  af-x:AFX_0002803 <urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36> ;
  obo:RO_0002350 <urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d> .
RDF → JSON: JSON-LD API maps RDF to a JSON view.

```json
[
    {
        "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
        "@type": ["http://purl.allotrope.org/ontologies/result#AFR_0002140"],
        "http://qudt.org/schema/qudt#numericValue": [{"@value": 58.40250856}],
        "http://qudt.org/schema/qudt#unit": [{"@id": "http://qudt.org/vocab/unit#Percent"}]
    },
    {
        "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
        "@type": ["http://purl.allotrope.org/ontologies/result#AFR_0002319"]
    },
    {
        "@id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
        "@type": ["http://purl.allotrope.org/ontologies/result#AFR_0002320"],
        "http://purl.allotrope.org/ontologies/property#AFX_0002803": [{
            "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36"
        }],
        "http://purl.obolibrary.org/obo/RO_0002350": [{
            "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d"
        }]
    }
]
```
Compaction: namespace prefixes

Context

```
{
  "af-r": "http://purl.allotrope.org/ontologies/result#",
  "af-x": "http://purl.allotrope.org/ontologies/property#",
  "obo": "http://purl.obolibrary.org/obo/",
  "qudt": "http://qudt.org/schema/qudt#",
  "unit": "http://qudt.org/vocab/unit#",
  "xsd": "http://www.w3.org/2001/XMLSchema#"
}
```

Data

```
{"@graph": [{
  "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
  "@type": "af-r:AFR_0002140",
  "qudt:numericValue": 58.40250856,
  "qudt:unit": {"@id": "unit:Percent"}
},
{
  "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bffff8dc39d",
  "@type": "af-r:AFR_0002319"
},
{
  "@id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
  "@type": "af-r:AFR_0002320",
  "af-x:AFX_0002803": {
    "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36"
  },
  "obo:RO_0002350": {
    "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bffff8dc39d"
  }
}]
```
Compaction: aliasing identifiers

Context

{"analyte aggregate document": {
"@id": "af-r:AFR_0002319"
},
"analyte document": {
"@id": "af-r:AFR_0002320"
},
"qNMR purity result": {
"@id": "af-r:AFR_0002140"
},
"has facet": {
"@id": "af-x:AFX_0002803",
"@type": "@id"
},
"has numeric value": {
"@id": "qudt:numericValue"
},
"has unit": {
"@id": "qudt:unit",
"@type": "@id"
},
"member of": {
"@id": "obo:RO_0002350",
"@type": "@id"}}

Data

{"@graph": [
{
"@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
"@type": "qNMR purity result",
"has numeric value": 58.40250856,
"has unit": "unit:Percent"
},
{
"@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
"@type": "analyte aggregate document"
},
{
"@id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
"@type": "analyte document",
"has facet": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
"member of": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d"
}]
}
Compaction: aliasing units

Context

```json
{
    "analyte aggregate document": {"@id": "af-r:AFR_0002319"},
    "analyte document": {"@id": "af-r:AFR_0002320"},
    "qNMR purity result": {"@id": "af-r:AFR_0002140"},
    "has facet": {
        "@id": "af-x:AFX_0002803",
        "@type": "@id"
    },
    "has numeric value": {"@id": "qudt:numericValue"},
    "has unit": {
        "@id": "qudt:unit",
        "@type": "@vocab"
    },
    "member of": {
        "@id": "obo:RO_0002350",
        "@type": "@id"
    },
    "%": {"@id": "unit:Percent"}
}
```

Data

```json
{"@graph": [
    {
        "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
        "@type": "qNMR purity result",
        "has numeric value": 58.40250856,
        "has unit": "%"
    },
    {
        "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
        "@type": "analyte aggregate document"
    },
    {
        "@id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
        "@type": "analyte document",
        "has facet": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
        "has numeric value": 58.40250856,
        "member of": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d"
    }
]}
```
Framing: coerce to ADM-specified hierarchy

Frame

```json
{
  "@context": {
    "@import": "https://example.com/afo/REC/2021/06/context.jsonld",
    "has member": {
      "@reverse": "member of"
    }
  },
  "@type": "analyte aggregate document",
  "has member": {
    "@type": "analyte document"
  }
}
```

Data

```json
{
  "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
  "@type": "analyte aggregate document",
  "has member": {
    "@id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
    "@type": "analyte document",
    "has facet": {
      "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
      "@type": "qNMR purity result",
      "has numeric value": 58.40250856,
      "has unit": "%"
    }
  }
}
```
Compaction: type-indexed properties

Frame

```json
{
  "@context": {
    "@import": "https://example.com/afo/REC/2021/06/context.jsonld",
    "has facet": {
      "@id": "af-x:AFX_0002803",
      "@type": "@id",
      "@container": "@type"
    },
    "has member": {
      "@reverse": "member of"
    }
  },
  "@type": "analyte aggregate document",
  "has member": {
    "@type": "analyte document"
  }
}
```

Data

```json
{
  "@id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
  "@type": "analyte aggregate document",
  "has member": {
    "@id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
    "@type": "analyte document",
    "has facet": {
      "qNMR purity result": {
        "@id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
        "has numeric value": 58.40250856,
        "has unit": "%"
      }
    }
  },
  "member of": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d"
}
```
Compaction: aliases of JSON-LD keywords

Frame

```json
{
    "@context": {
        "@import": "https://example.com/afo/REC/2021/06/context.jsonld",
        "has member": {
            "@reverse": "member of"
        },
        "id": "@id",
        "type": "@type"
    },
    "@type": "analyte aggregate document",
    "has member": {
        "@type": "analyte document"
    }
}
```

Data

```json
{
    "id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
    "type": "analyte aggregate document",
    "has member": {
        "id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
        "type": "analyte document",
        "has facet": {
            "qNMR purity result": {
                "id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
                "has numeric value": 58.40250856,
                "has unit": "%"
            }
        }
    },
    "member of": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d"
}
```
Beyond JSON-LD API: remove redundancies

Original

```json
{
    "id": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d",
    "type": "analyte aggregate document",
    "has member": {
        "id": "urn:uuid:96d649e9-29d7-43ca-8b0a-28c92d9c78a5",
        "type": "analyte document",
        "has facet": {
            "qNMR purity result": {
                "id": "urn:uuid:0ed8b506-5cc1-469f-a2f5-062d29d49f36",
                "has numeric value": 58.40250856,
                "has unit": "%"
            }
        }
    },
    "member of": "urn:uuid:4e685472-d7b6-4a47-8a75-2bbfff8dc39d"
}
```

Post-processed

```json
{
    "type": "analyte aggregate document",
    "has member": {
        "type": "analyte document",
        "has facet": {
            "type": "qNMR purity result",
            "has numeric value": 58.40250856,
            "has unit": "%"
        }
    }
}
```
Limits of JSON-LD for AFO and ADM

JSON-LD terms have to be unique, so some AFO terms have to be **disambiguated** by domain, such as “table (AFE)” and “table (IAO)”.

**Unsupported slashes** present in unit symbols, such as “cm^3/min”, can be replaced with supported characters, such as “cm^3 per min”.

**JSON-native data types** are available only for xsd:boolean, xsd:integer, and xsd:double, but AFO/ADM use others, such as xsd:int.

Cannot **omit “@id”** if not used for embedded blank nodes.

Cannot **redefine reserved JSON-LD keywords**, such as for type-indexing of “@graph”.
Summary

Why?
Leverage developer familiarity and tool support for JSON for easier consumption of ADF data without compromising the semantic interoperability provided by standards.

How?
AFO → JSON-LD context
ADMs → JSON-LD frames
ADF API: ADF Data Description → RDF
JSON-LD API: RDF → JSON (and back)

Generic approach applicable to both tabular and graph ADMs.