JSON-LD
from semi-structured data

A presentation of options

Press Space for next page →

Boundary Conditions

Context

- starting with CSV or Forms based input exported to CSV or YAML
- python as the base language where possible

Principle(s)

- avoid lock in: Planned migration to other approaches

Goal(s)

- Sustainable self-hosted approaches
- Focus on established languages and libraries

Ecosystem

This is a simple and not exhaustive review of tools and approaches for ETLs to RDF

- W3C Wiki for ConverterToRDF
- Github search for convert CSV to RDF
- LinkML general purpose modeling language
- SchemaSheets focus on wide data in spreadsheets, uses LinkML
- OpenRefine
- Any23 now in Apache Attic

Focus on these

- RML based rml.io a superset of W3C R2RML
- Morph-KGC
- Code with templates (Jinja2)

Profiles 1/2

Science on Schema.org (SOSO)

This is also what the UNESCO ODIS Ocean InfoHub Project is using.

```json
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "Dataset",
  "name": "Removal of organic carbon by natural bacterioplankton communities ...",
  ...
  "variableMeasured": [
    {
      "@type": "PropertyValue",
      "name": "latdd",
      "alternateName":"latitude, decimal degrees",
      "propertyID":"http://purl.obolibrary.org/obo/NCIT_C68642",
      "description": "Latitude where water samples were collected ...",
    },
    ...
  ]
}
```
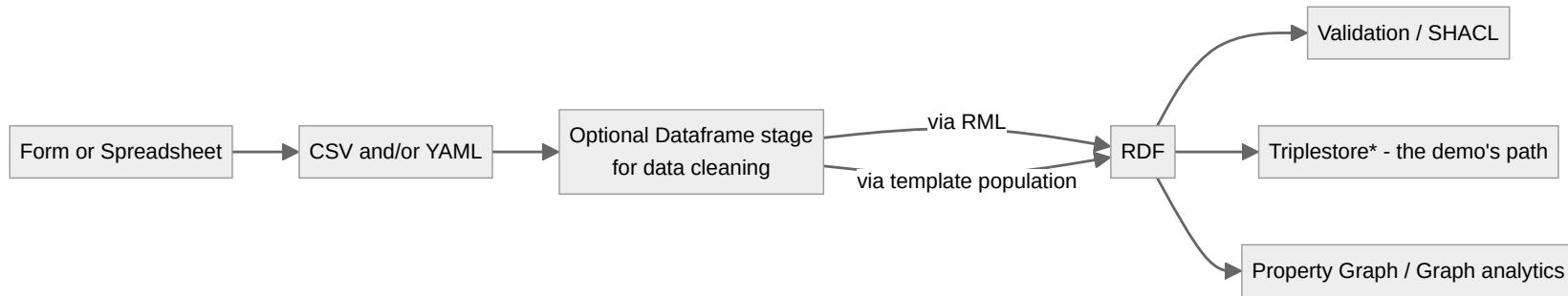
Profiles 2/2

SOSO pattern used in OIH Essential Ocean Variables approach.

CODATA CDIF is lookin at https://schema.org/StatisticalVariable

```
"variableMeasured": [
    {
        "@type": "PropertyValue",
        "name": "Sea State",
        "description": "Sea State is the characterization of wave and swell, typically in terms of height, wavelength,
        "propertyID": [
            "http://purl.obolibrary.org/obo/ENVO_01001374",
            "https://gcmd.earthdata.nasa.gov/KeywordViewer/scheme/all/11aca777-8a01-42ce-b076-b3059c3d8cae?gtm_keyword=
        ]
    },
    {
        "@type": "PropertyValue",
        "name": "Ocean Surface Stress",
        "description": "The 2-dimensional horizontal vector Ocean Surface Stress is the rate per unit area at which mom
        "propertyID": "http://purl.obolibrary.org/obo/ENVO_01001844",
        "value": 0.2,
        "unitCode": "http://qudt.org/vocab/unit/N-PER-M2"
    }
],
```

Workflow

Just a quick overview of the workflow

```
Form or Spreadsheet → CSV and/or YAML → Optional Dataframe stage for data cleaning
```

via RML

via template population

→ RDF → Validation / SHACL

→ RDF → Triplestore* - the demo's path

→ RDF → Property Graph / Graph analytics

Notebook

Let's transition to the notebook