

MixS as RDF: Fields as Object Properties

2020-11-09

Context: modeling of fields (e.g. altitude)

Ticket: <https://github.com/GenomicsStandardsConsortium/mixs-rdf/issues/9>

Possibilities

- Data Property: Range is a literal (e.g. xsd string)
- Class
- Annotation Property: Range is either literal or object
- Object Property: Range is an object

Note:

- This will not impact people using CSVs for metadata (it's all just columns)

NMDC rendering of mixes (via biolinkml)

The screenshot displays a web browser window with the URL `https://microbiomedata/schema`. The browser tabs include 'attributa', 'alt', 'Active ontology', 'Entities', 'Individuals by class', 'DL Query', and 'Individual Hierarchy Tab'. The main content area shows the 'alt' class definition from the schema `https://microbiomedata/schema/mixs/alt`. The 'Annotations: alt' section lists several properties: `rdfs:label` (value: 'alt'), `skos:definition` (value: '"Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric layers and clouds. It is used to measure the height of an object which is above the earth's surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the sampled position in the air"'), `pattern` (value: '{float} {unit}'), and `skos:inScheme` (value: 'environment'). The 'Description: alt' section shows relationships: 'Equivalent To' (none), 'SubProperty Of' (value: 'attribute'), 'Inverse Of' (none), 'Domains (intersection)' (value: 'study or biosample'), and 'Ranges (intersection)' (value: 'quantity value'). A left-hand sidebar displays a tree view of the ontology, with 'alt' highlighted under the 'attribute' class.

<https://github.com/microbiomedata/nmdc-metadata/tree/master/schema>

NMDC rendering of mixes (via biolinkml)

The screenshot displays a web browser window with the URL `https://microbiomedata/schema`. The browser's address bar shows the current page is `alt`. The interface includes a search bar and several tabs: `active ontology`, `Entities`, `Individuals by class`, `DL Query`, and `Individual Hierarchy Tab`. On the left, there are navigation panels for `Datatypes`, `Individuals`, `Data properties`, `Annotation properties`, `Classes`, and `Object properties`. The `Object property hierarchy` is expanded to show `owl:topObjectProperty` and `attribute`. The main content area shows the `alt` class with its `Annotations` and `Usage` tabs. The `Annotations` tab is active, showing `rdfs:label` as `alt` and `skos:definition` as "Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric layers and clouds. It is used to measure the height of an object which is above the earth's surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the sampled position of the air". The `Usage` tab shows `pattern` as `{float} {unit}` and `skos:inScheme` as `environment`. Below the main content, there is a `Description: alt` section with a list of property types: `Functional`, `Inverse function`, `Transitive`, `Symmetric`, `Asymmetric`, `Reflexive`, and `Irreflexive`. The `Equivalent To` section shows `attribute`. The `SubProperty Of` section shows `attribute`. The `Inverse Of` section is empty. The `Domains (intersection)` section shows `study or biosample`. The `Ranges (intersection)` section shows `'quantity value'`. A large yellow arrow points from the right towards the `skos:definition` text, with the text "Regex for string" written inside the arrow.

NMDC rendering of mixes (via biolinkml)

The screenshot shows a web browser window with the URL `https://microbiomedata/schema`. The browser tabs include 'Active ontology', 'Entities', 'Individuals by class', 'DL Query', and 'Individual Hierarchy Tab'. The main content area displays the 'alt' class definition from the schema `https://microbiomedata/schema/mixs/alt`. The 'Annotations: alt' section lists several properties: `rdfs:label` (value: alt), `skos:definition` (value: "Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric layers and clouds. It is used to measure the height of an object which is above the earth's surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the sampled position in the air"), `pattern` (value: {float} {unit}), and `skos:inScheme` (value: environment). The 'Description: alt' section shows a list of property types: Functional, Inverse functional, Transitive, Symmetric, Asymmetric, Reflexive, and Irreflexive. The 'Domains (intersection)' section shows a domain of `study or biosample`. The 'Ranges (intersection)' section shows a range of `'quantity value'`. A yellow arrow points from the 'Ranges' section to the text 'Object Property; range is Quantity Value'.

Object Property; range is Quantity Value

Example range class: Quantity Value

The screenshot displays a web-based ontology editor interface. On the left, a class hierarchy is shown under 'owl:Thing', with 'quantity value' highlighted. The main area is divided into two panels: 'Annotations: quantity value' and 'Description: quantity value'.

Class hierarchy: quantity value

- owl:Thing
 - activity
 - agent
 - attribute value
 - boolean value
 - controlled term value
 - geolocation value
 - integer value
 - person value
 - quantity value**
 - text value
 - timestamp value
 - url value
 - boolean
 - bytes
 - class_definition
 - database
 - date
 - datetime
 - decimal degree
 - double
 - float
 - integer
 - language code
 - named thing
 - ncname
 - nodeidentifier
 - objectidentifier
 - slot_definition
 - string

Annotations: quantity value

- Annotations +
- rdfs:label quantity value
- skos:definition A simple quantity, e.g. 2cm
- skos:exactMatch <http://schema.org/QuantityValue>

Description: quantity value

- Equivalent To +
- SubClass Of +
- 'attribute value'
- 'has numeric value' max 1 double
- 'has raw value' max 1 string
- 'has unit' max 1 unit

Example range class: Quantity Value

The screenshot displays an ontology editor interface. On the left, a class hierarchy tree is shown under 'owl:Thing', with 'quantity value' highlighted. The main area is divided into two panes: 'Annotations: quantity value' and 'Description: quantity value'. The annotations pane lists three annotations: 'rdfs:label' with the value 'quantity value', 'skos:definition' with the value 'A simple quantity, e.g. 2 cm', and 'skos:exactMatch' with the value '<http://schema.org/QuantityValue>'. The description pane shows 'Equivalent To' and 'SubClass Of' sections. The 'SubClass Of' section lists three subclasses: ''attribute value'', ''has numeric value' max 1 double', and ''has raw value' max 1 string'. A yellow arrow points from the text 'E.g. "2 cm" -- literal string representation' to the 'skos:definition' annotation.

Class hierarchy: quantity value

Annotations: quantity value

Annotations

- rdfs:label
quantity value
- skos:definition
A simple quantity, e.g. 2 cm
- skos:exactMatch
<http://schema.org/QuantityValue>

Class description Taxon constraints

Description: quantity value

Equivalent To

SubClass Of

- 'attribute value'
- 'has numeric value' max 1 double
- 'has raw value' max 1 string
- 'has unit' max 1 unit

E.g. "2 cm" -- literal string representation

Example range class: Quantity Value

The screenshot displays an ontology editor interface. On the left, a class hierarchy is shown under 'owl:Thing', with 'quantity value' highlighted. The main area is divided into two panels: 'Annotations: quantity value' and 'Description: quantity value'.

Annotations: quantity value

- rdfs:label**: quantity value
- skos:definition**: A simple quantity, e.g. 2 cm
- skos:exactMatch**: <http://schema.org/QuantityValue>

Description: quantity value

Equivalent To: +

SubClass Of: +

- 'attribute value'
- 'has numeric value' max 1 double
- 'has raw value' max 1 string
- 'has unit' max 1 unit

Structured;
broken into 3 sub-parts

Example range class: Quantity Value

The screenshot displays a web-based ontology editor interface. On the left, a class hierarchy tree is shown under 'owl:Thing', with 'quantity value' highlighted in blue. The main area is divided into two panels: 'Annotations: quantity value' and 'Description: quantity value'. The 'Annotations' panel lists three annotations: 'rdfs:label' with the value 'quantity value', 'skos:definition' with the value 'A simple quantity, e.g. 2 cm', and 'skos:exactMatch' with the value '<http://schema.org/QuantityValue>'. The 'Description' panel shows 'Equivalent To' and 'SubClass Of' sections. The 'SubClass Of' section lists three subclasses: ''attribute value'', ''has numeric value' max 1 double', and ''has raw value' max 1 string'. Two yellow arrows point to the 'SubClass Of' list: one points to ''has numeric value' max 1 double' with the text 'E.g. "2"^^xsd:double', and another points to ''has unit' max 1 unit' with the text 'E.g. cm (UO)'. The interface also includes tabs for 'Classes', 'Object properties', 'Data properties', 'Annotations', and 'Usage', and a 'Class hierarchy: quantity value' header.

Class hierarchy: quantity value

Annotations: quantity value

Annotations

- rdfs:label
quantity value
- skos:definition
A simple quantity, e.g. 2 cm
- skos:exactMatch
<http://schema.org/QuantityValue>

Class description Taxon constraints

Description: quantity value

Equivalent To

SubClass Of

- 'attribute value'
- 'has numeric value' max 1 double
- 'has raw value' max 1 string
- 'has unit' max 1 unit

E.g. "2"^^xsd:double

E.g. cm (UO)

Other Object Property Ranges

The screenshot displays the Protege ontology editor interface. The main window shows the 'env_local_scale' object property configuration. The 'Object property hierarchy' on the left lists various properties, with 'env_local_scale' selected. The 'Annotations' tab is active, showing the 'skos:definition' for the property: "In this field, report the entity or entities which have significant causal influences on your sample are of smaller spatial grain than your entry for (multiple terms): termLabel [termID]|termLabel taken from various vegetation layers in a forest (01000337)|litter layer [ENVO:01000338]|ur request new terms on the ENVO tracker, ide". The 'Ranges' section is expanded to show the range configuration for 'env_local_scale', which is set to 'controlled term value'. The 'Class description' tab is also visible, showing the description: 'controlled term value'. The 'Equivalent To' section is empty. The 'SubClass Of' section shows that 'env_local_scale' is a subclass of 'term max 1 'ontology class''. The 'General class axioms' section is empty. The 'SubClass Of (Anonymous Ancestor)' section shows that 'env_local_scale' is a subclass of 'was generated by' max 1 activity.

schema (https://microbiomedata/schema)

env_local_scale — https://microbiomedata/schema

Annotations: env_local_scale

Annotations +

- rdfs:label
- env_local_scale
- skos:definition

Description: env_local_scale

Characteristics

- Functional
- Inverse function
- Transitive
- Symmetric
- Asymmetric
- Reflexive

Inverse Of +

Domains (intersection) +

Ranges (intersection) +

- 'controlled term value'

Disjoint With +

Annotations: controlled term value

Annotations +

- rdfs:label
- controlled term value
- skos:definition
- A controlled term or class from an ontology

Class description Taxon constraints

Description: controlled term value

Equivalent To +

SubClass Of +

- 'attribute value'
- term max 1 'ontology class'

General class axioms +

SubClass Of (Anonymous Ancestor)

- 'was generated by' max 1 activity

Example instance data

The screenshot displays a web-based ontology editor interface. At the top, there are tabs for 'Active ontology', 'Entities', 'Individuals by class', 'DL Query', and 'Individual Hierarchy Tab'. The main content area is divided into several panels:

- Class hierarchy: biosample**: A tree view on the left showing various classes like 'timestamp value', 'url value', 'boolean', 'bytes', 'class_definition', 'database', 'date', 'datetime', 'decimal degree', 'double', 'float', 'integer', 'language code', 'named thing', and 'biosample' (highlighted).
- Annotations: SAM1234**: A panel showing the annotations for the instance 'SAM1234'. It lists 'rdfs:label' with the value 'SAM1234' and a language of 'en'.
- Description: SAM1234**: A panel showing the types of the instance, which is 'biosample'.
- Property assertions: SAM1234**: A panel showing the property assertions for the instance. It lists 'alt' with the value '87afe5342e'.
- Direct instances: SAM1234**: A panel showing the direct instances of the class 'biosample', which are 'SAM1234' and 'SAM2345'.

Example instance data

The screenshot displays the Protege ontology editor interface. On the left, a class hierarchy for 'biosample' is shown, with 'biosample' selected. The main area is divided into several panels:

- Annotations: SAM1234**: Shows an annotation 'rdfs:label [language: en] SAM1234'.
- Description: SAM1234**: Shows the type 'biosample'.
- Property assertions: SAM1234**: Shows an object property assertion 'alt 87afe5342e'.
- Direct instances: SAM1234**: Lists instances 'SAM1234' and 'SAM2345'.
- Direct instances: 87afe5342e**: Shows the instance '87afe5342e' is of type 'quantity value'.
- Description: 87afe5342e**: Shows the type 'quantity value'.
- Property assertions: 87afe5342e**: Shows object property assertions 'has unit cm' and data property assertions 'has raw value' '2 cm' and 'value' '2.0'.

A red line highlights the relationship between the 'alt 87afe5342e' property assertion in the SAM1234 panel and the '87afe5342e' instance in the SAM1234 Direct instances panel.

Yes I know 2cm is a silly value for alt...
Didn't have time to change example

Example instance data - as JSON-LD

Active ontology x Entities x Individuals by class x DL Query x Individual Hierarchy Tab x

Class hierarchy: biosample

- timestamp value
- url value
- boolean
- bytes
- class_definition
- database
- date
- datetime
- decimal degree
- double
- float
- integer
- language code
- named thing
- biosample**

Annotations Usage

Annotations: SAM1234

Annotations +

- rdfs:label [language: en]
- SAM1234

Description: SAM1234

Types +

- biosample

Same Individual As +

Different Individuals +

Direct instances: SAM1234

For: biosample

- SAM1234
- SAM2345

```
## JSON-LD
```

```
{ "@context": { ... },
```

```
{ "id": "SAM1234",
```

```
  "alt": {
```

```
    "Has_raw_value": "2 cm",
```

```
    "Value" : 2.0,
```

```
    "Has unit": "cm"
```

```
  },
```

```
  ...
```

```
}
```

Direct instances: 87afe5342e

For: quantity value'

- 87afe5342e

Types +

quantity value'

Same Individual As +

Different Individuals +

Object property assertions +

'has unit' cm

Data property assertions +

'has raw value' "2 cm"

value "2.0"^^xsd:double

ShEx

```
<Biosample> CLOSED {
  ( $<Biosample_tes> ( &<NamedThing_tes> ;
    rdf:type [ <NamedThing> ] ? ;
    <https://microbiomedata/schema/mixs/env_package> @<TextValue> ? ;
    <lat_lon> @<GeolocationValue> ;
    <https://microbiomedata/schema/mixs/geo_loc_name> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/collection_date> @<TimestampValue> ? ;
    <env_broad_scale> @<ControlledTermValue> ;
    <env_local_scale> @<ControlledTermValue> ;
    <env_medium> @<ControlledTermValue> ;
    <ecosystem> @<AttributeValue> ? ;
    <ecosystem_category> @<AttributeValue> ? ;
    <ecosystem_type> @<AttributeValue> ? ;
    <ecosystem_subtype> @<AttributeValue> ? ;
    <specific_ecosystem> @<AttributeValue> ? ;
    <depth> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/tot_org_carb> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/alt> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/elev> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/salinity> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/diss_oxygen> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/nitrate> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/chlorophyll> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/cur_land_use> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/cur_vegetation> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/cur_vegetation_meth> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/previous_land_use> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/previous_land_use_meth> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/crop_rotation> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/agrochem_addition> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/tillage> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/fire> @<TimestampValue> ? ;
    <https://microbiomedata/schema/mixs/flooding> @<TimestampValue> ? ;
    <https://microbiomedata/schema/mixs/extreme_event> @<TimestampValue> ? ;
    <https://microbiomedata/schema/mixs/horizon> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/horizon_meth> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/sieving> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/water_content> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/water_content_soil_meth> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/samp_vol_we_dna_ext> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/pool_dna_extract> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/store_cond> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/link_climate_info> @<TextValue> ? ;
    <https://microbiomedata/schema/mixs/season_temp> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/annual_temp> @<QuantityValue> ? ;
    <https://microbiomedata/schema/mixs/season_precpt> @<QuantityValue> ? ;
    .....
```

JSON-Schema

```
    },
    "Biosample": {
      "additionalProperties": false,
      "description": "A material sample. It may be environmental (encompassing many organisms) or isolate or tissue. An environmental sample containing genetic material from multiple individuals as a biosample.",
      "properties": {
        "agrochem_addition": {
          "$ref": "#/definitions/QuantityValue",
          "description": "\"Addition of fertilizers, pesticides, etc. - amount and time of applications\""
        },
        "al_sat": {
          "$ref": "#/definitions/QuantityValue",
          "description": "Aluminum saturation (esp. For tropical soils)"
        },
        "al_sat_meth": {
          "$ref": "#/definitions/TextValue",
          "description": "Reference or method used in determining Al saturation"
        },
        "alt": {
          "$ref": "#/definitions/QuantityValue",
          "description": "\"Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric used to measure the height of an object which is above the earth's surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the surface\""
        },
        "alternate_identifiers": {
          "description": "The same biosample may have distinct identifiers in different databases (e.g. GOLD and EMSL)",
          "items": {
            "type": "string"
          },
          "type": "array"
        },
        "annual_precpt": {
          "$ref": "#/definitions/QuantityValue",
          "description": "\"The average of all annual precipitation values known, or an estimated equivalent value derived by such methods as regional indexes or Isohyetal maps.\""
        },
        "annual_temp": {
          "$ref": "#/definitions/QuantityValue",
          "description": "Mean annual temperature"
        },
        "chlorophyll": {
          "$ref": "#/definitions/QuantityValue",
          "description": "Concentration of chlorophyll"
        },
        "collection_date": {
          "$ref": "#/definitions/TimestampValue",
          "description": "\"The time of sampling, either as an instance (single point in time) or interval. In case no exact time is available, the date/time can be right truncated i.e. all 2008-01-23T19:23:10+00:00; 2008-01-23T19:23:10; 2008-01-23; 2008-01; 2008; Except: 2008-01; 2008 all are ISO8601 compliant\""
        },
        "crop_rotation": {
          "$ref": "#/definitions/TextValue",
          "description": "\"Whether or not crop is rotated, and if yes, rotation schedule\""
        },
        "cur_land_use": {
          "$ref": "#/definitions/TextValue",
          "description": "Present state of sample site"
        },
        "cur_vegetation": {
          "$ref": "#/definitions/TextValue",
          "description": "\"Vegetation classification from one or more standard classification systems. or agricultural crop\""
        }
      }
    }
  }
}
```

Summary

- Using Object Properties allows us to have complex objects as values
 - Have our cake and eat it: one value can be the raw string value
 - E.g. “2 cm”, “ENVO:nnnnn mangrove biome”
 - Extensible to including provenance
- Will not affect majority of MlxS users