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

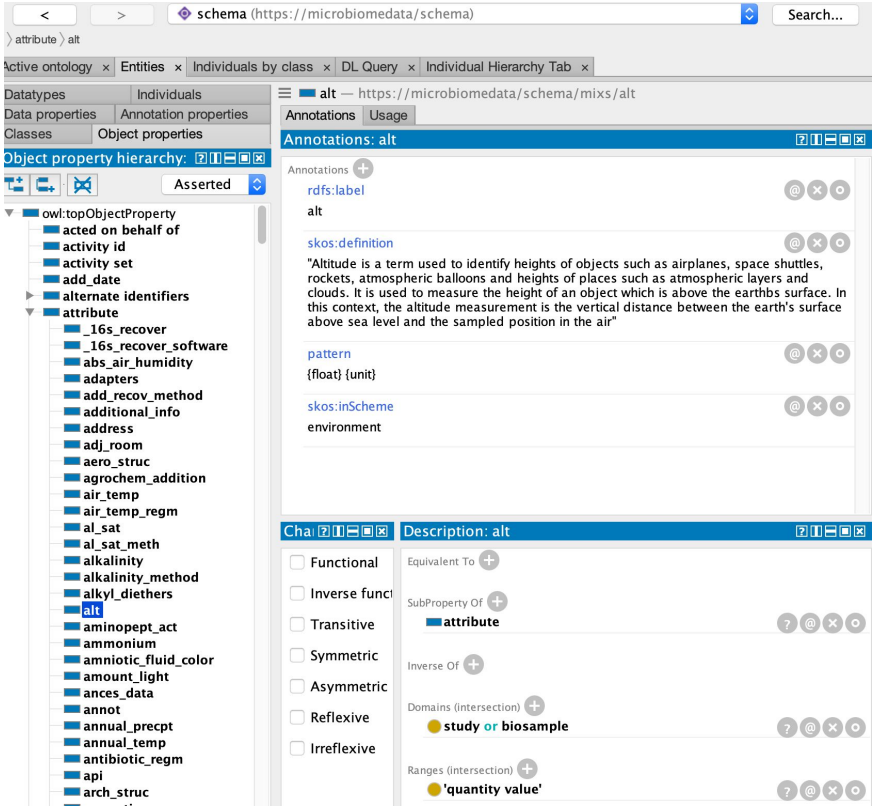
Possibilities

- 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:

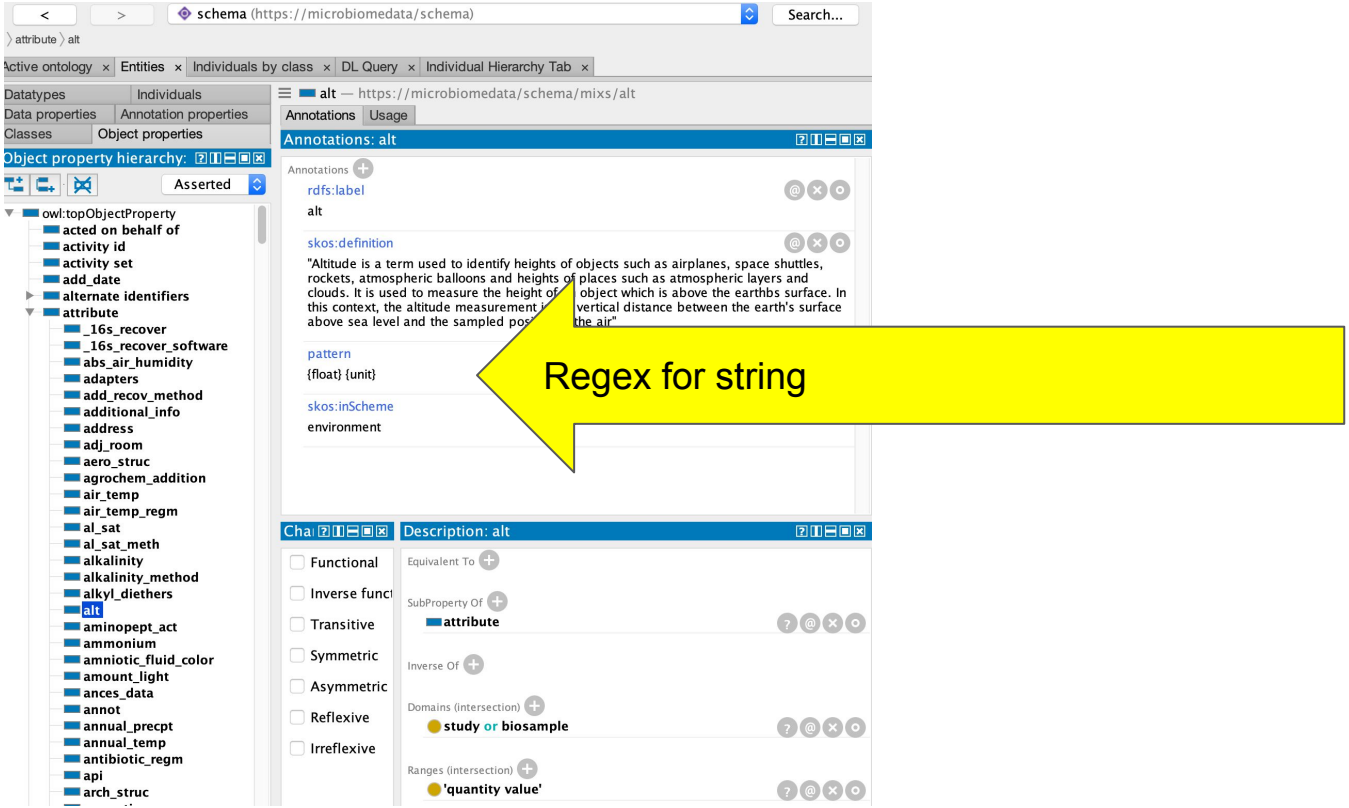
- This will not impact people using CSVs for metadata (it’s all just columns)

NMDC rendering of mixs (via biolinkml)

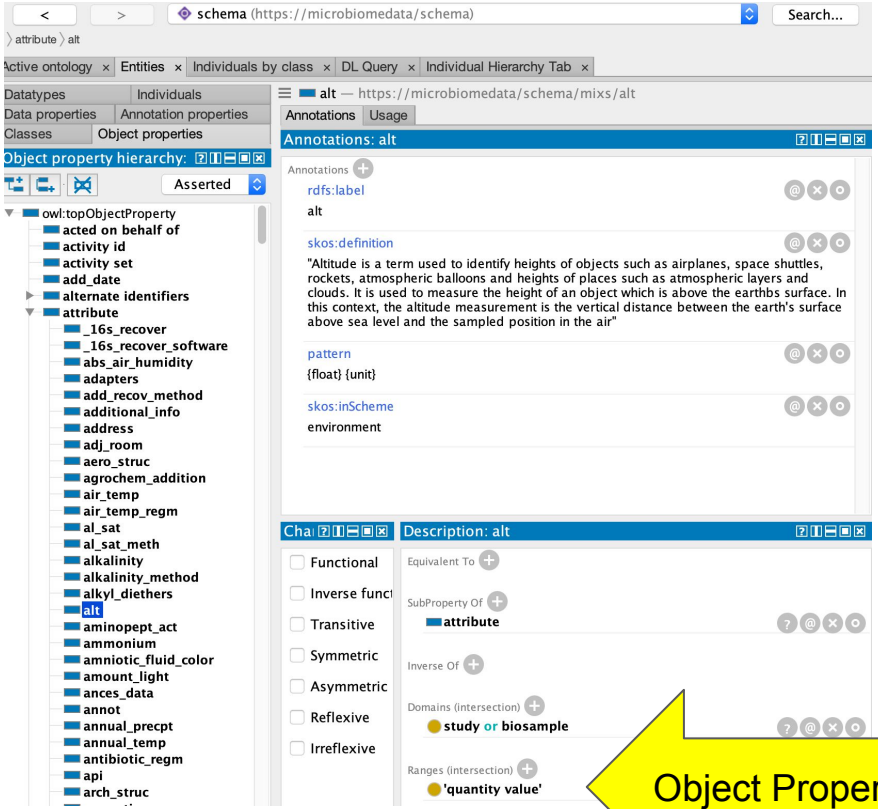


https://github.com/microbiomedata/nmdc-metadata/tree/master/schema

NMDC rendering of mixs (via biolinkml)

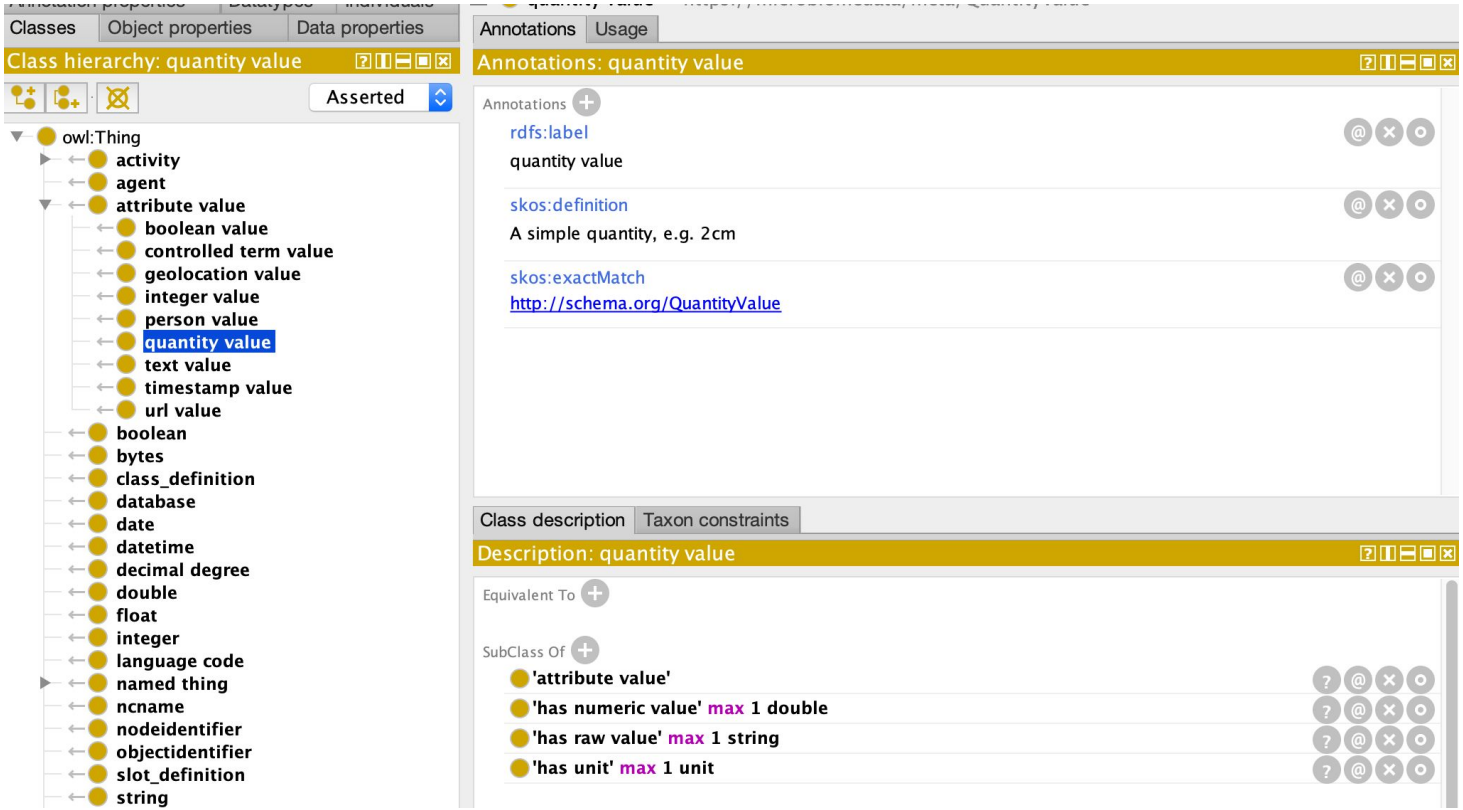


NMDC rendering of mixs (via biolinkml)

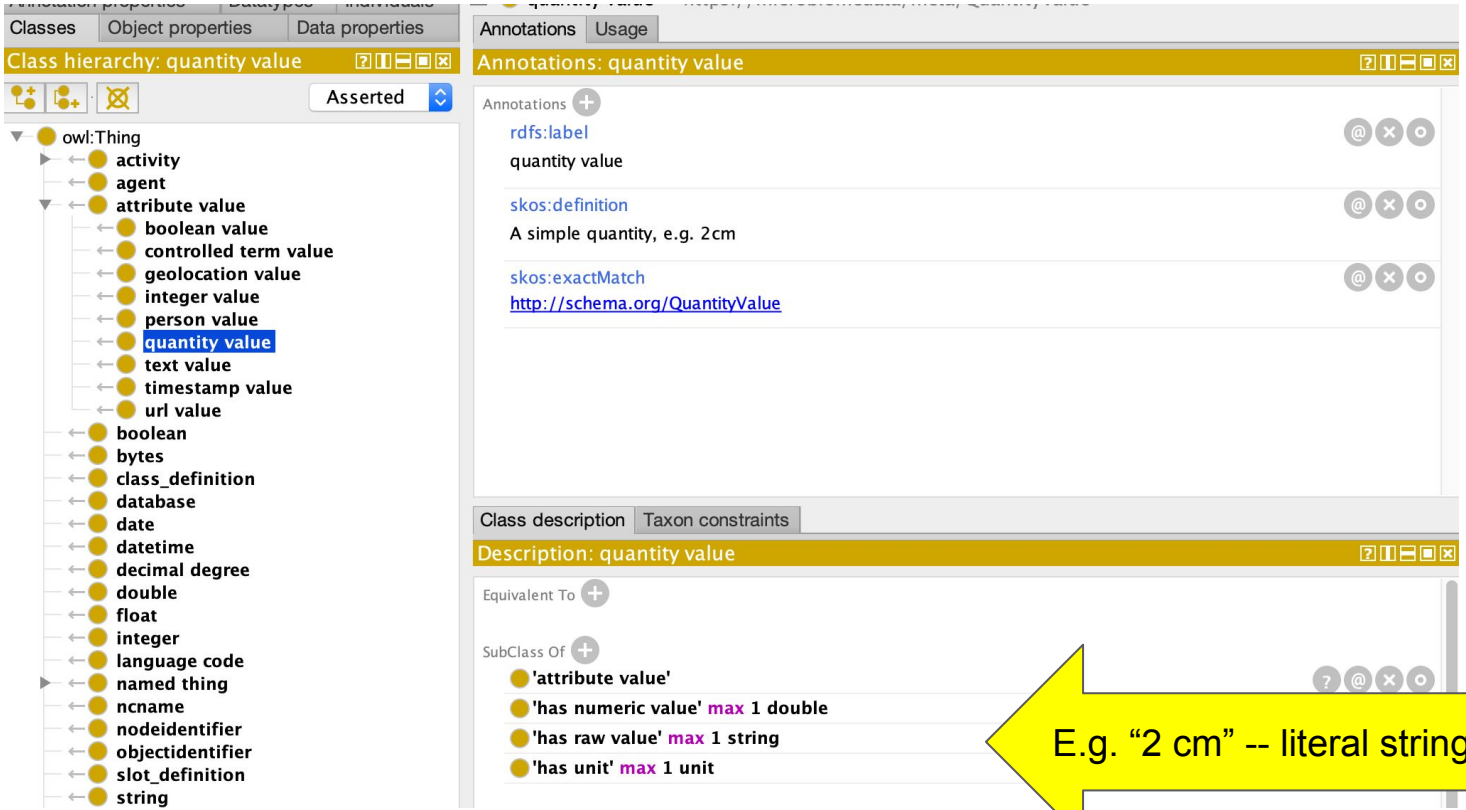


Object Property; range is Quantity Value

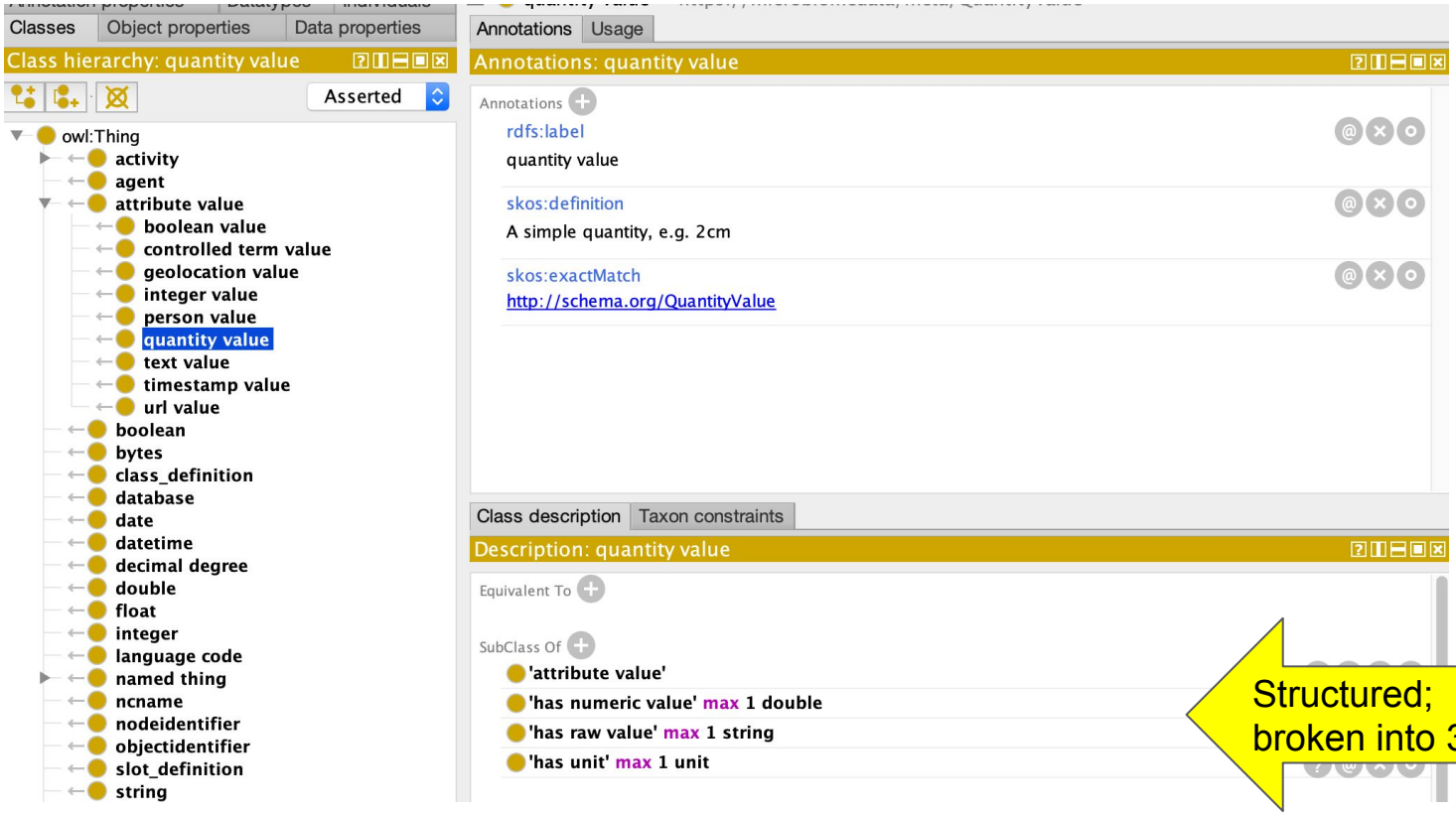
Example range class: Quantity Value



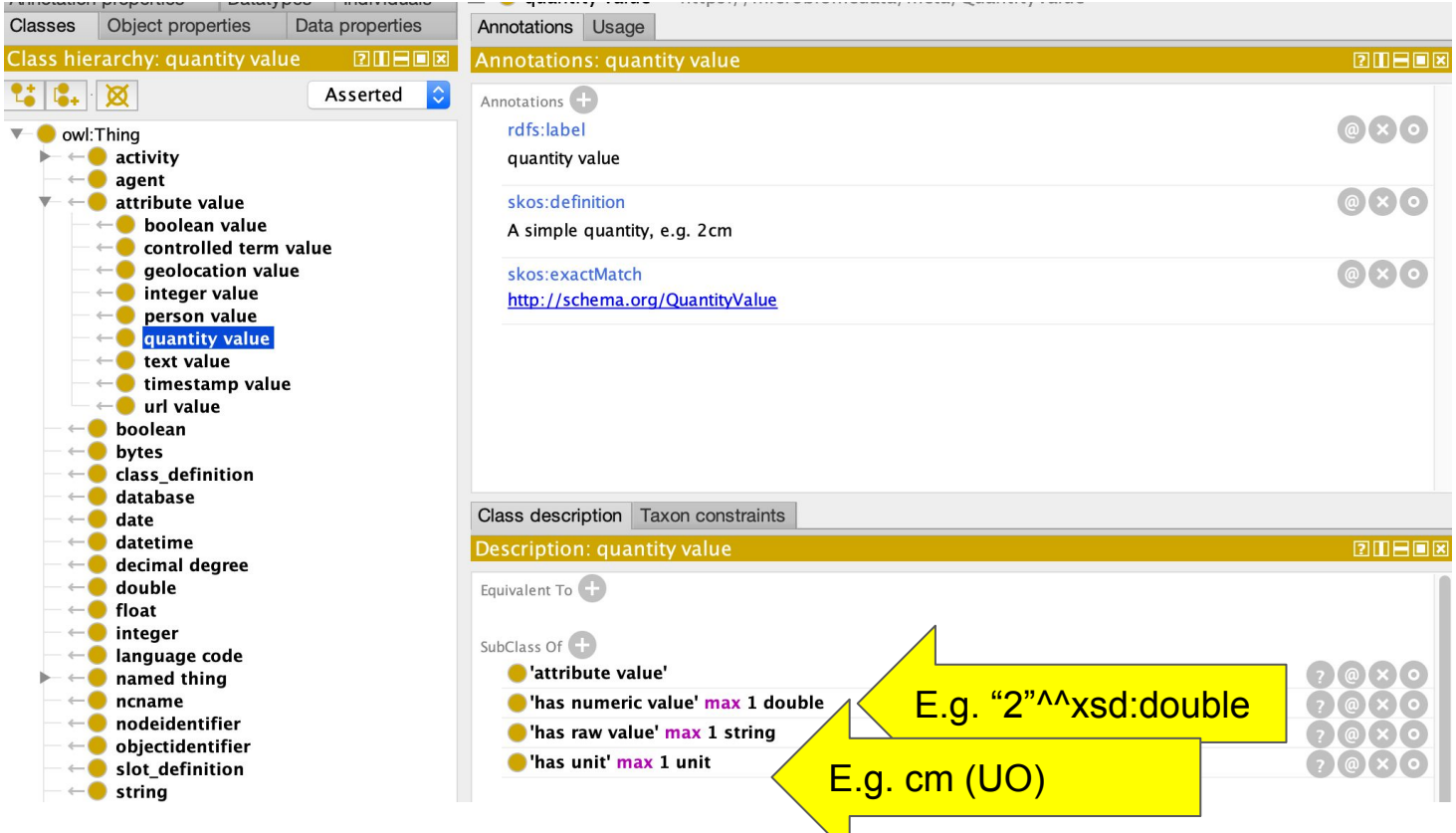
Example range class: Quantity Value



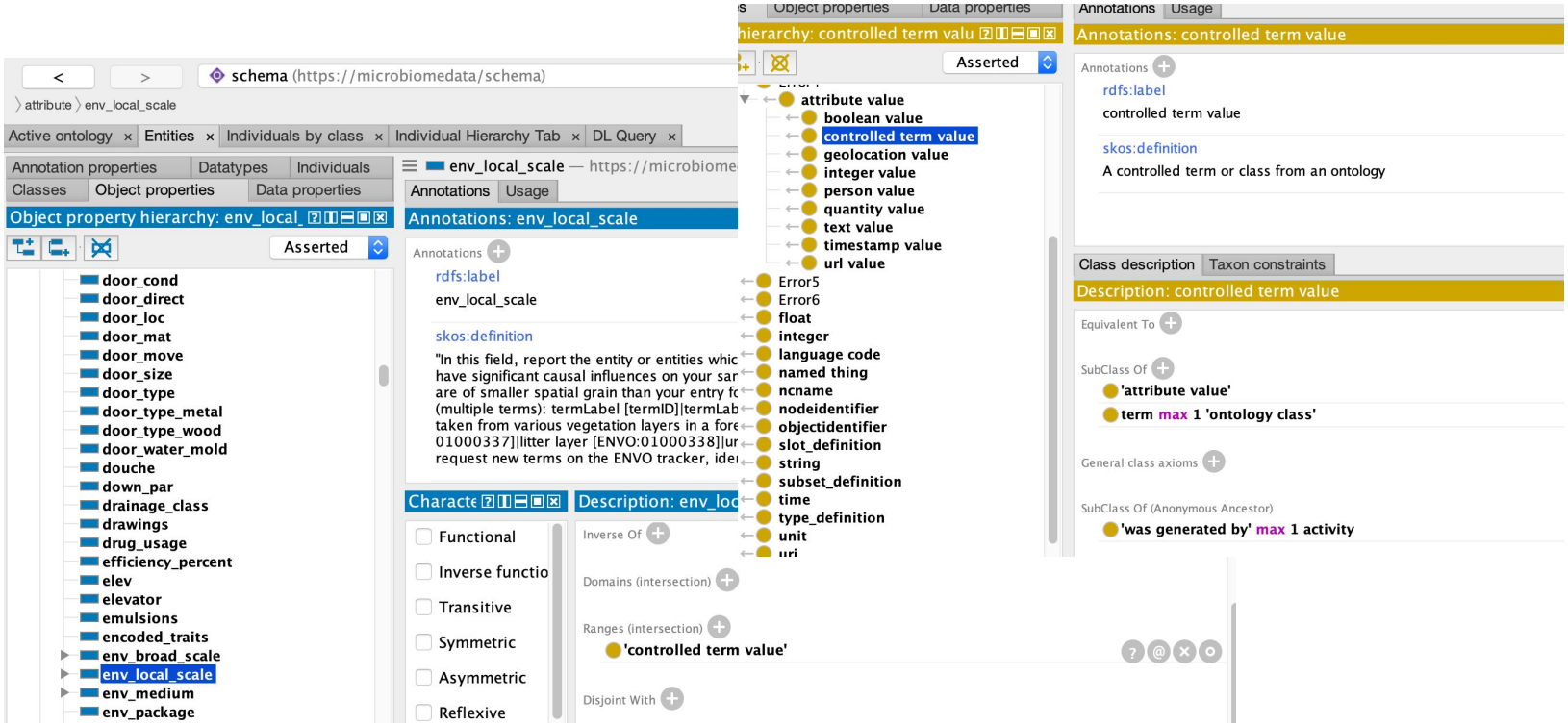
Example range class: Quantity Value



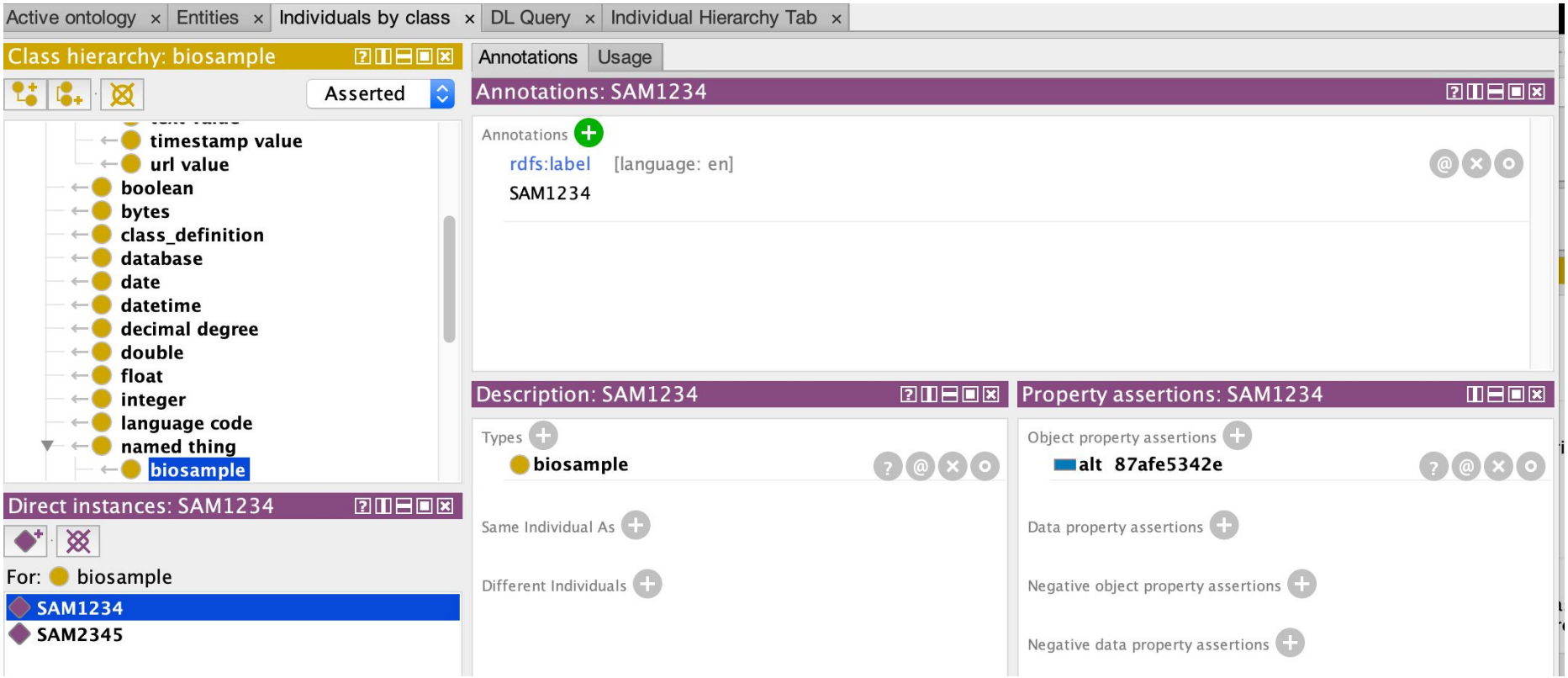
Example range class: Quantity Value



Other Object Property Ranges

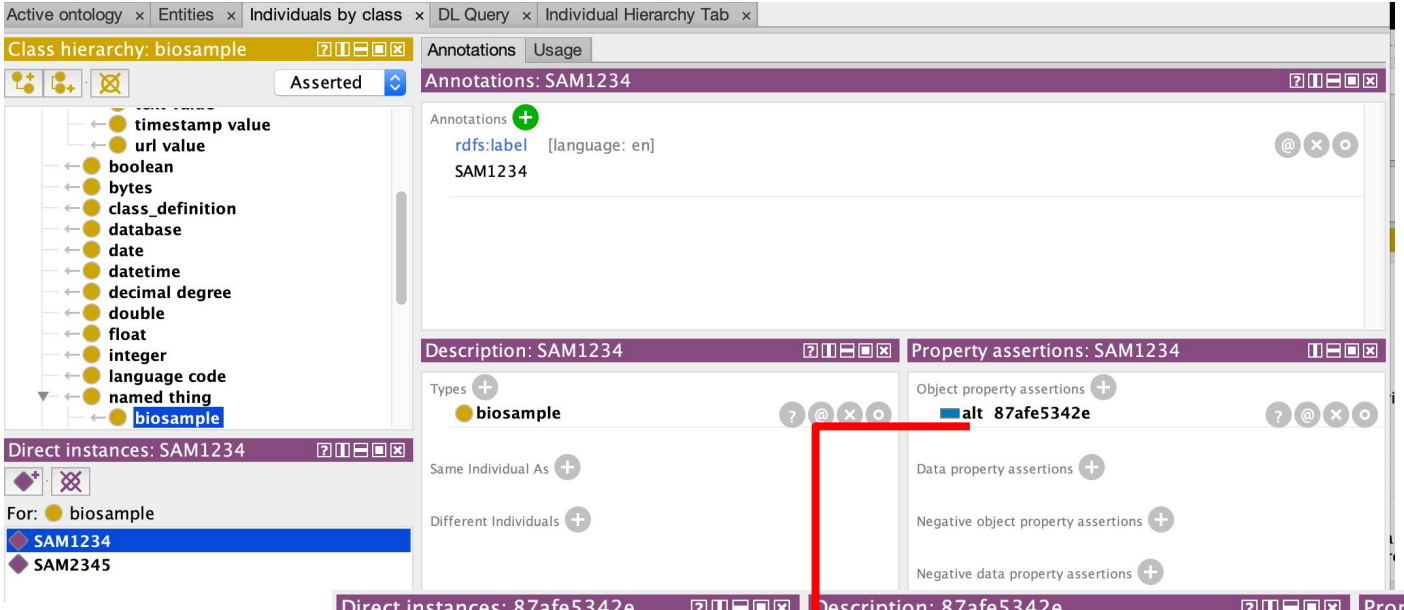


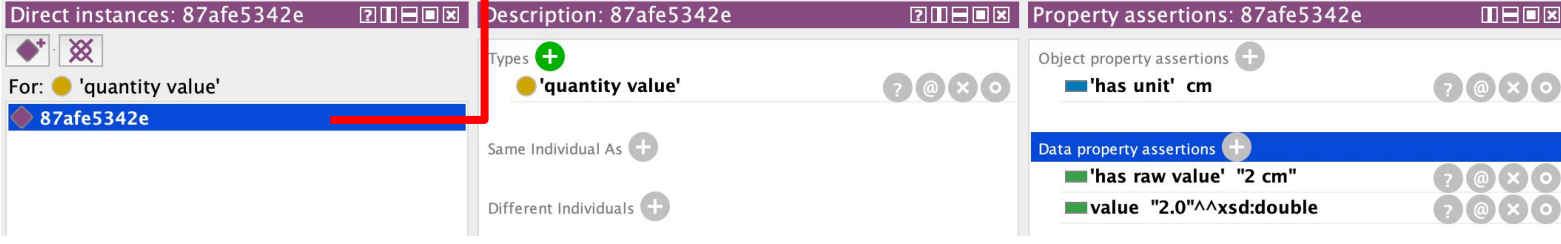
Example instance data



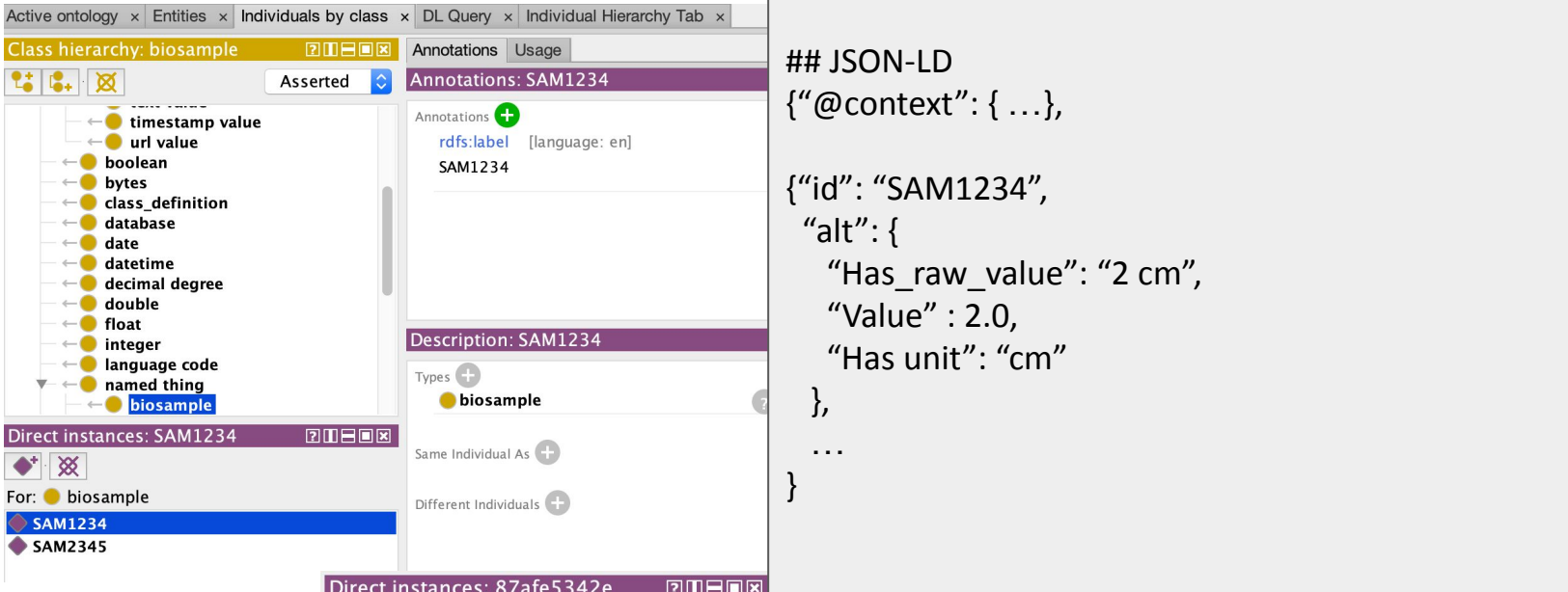
Example instance data

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





Example instance data - as JSON-LD



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_extracts> @<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 indivi
to 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
sed to measure the height of an object which is above the earthbs surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the
e\""
        },
        "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 MIxS users