

# Bringing Semantics to Citizen Science Data – A Semantic Extension of Open Data Kit 1

Markus D. Steinberg<sup>1</sup>, Sirko Schindler<sup>2</sup>, Friederike Klan<sup>3</sup>

<sup>1</sup>*Friedrich-Schiller-Universität, Jena*

<sup>2</sup>*Data Management Technologies Group, DLR Institute of Data Science, Jena*

<sup>3</sup>*Citizen Science Group, DLR Institute of Data Science, Jena*

EGU General Assembly 2019  
Vienna, 2019-04-12

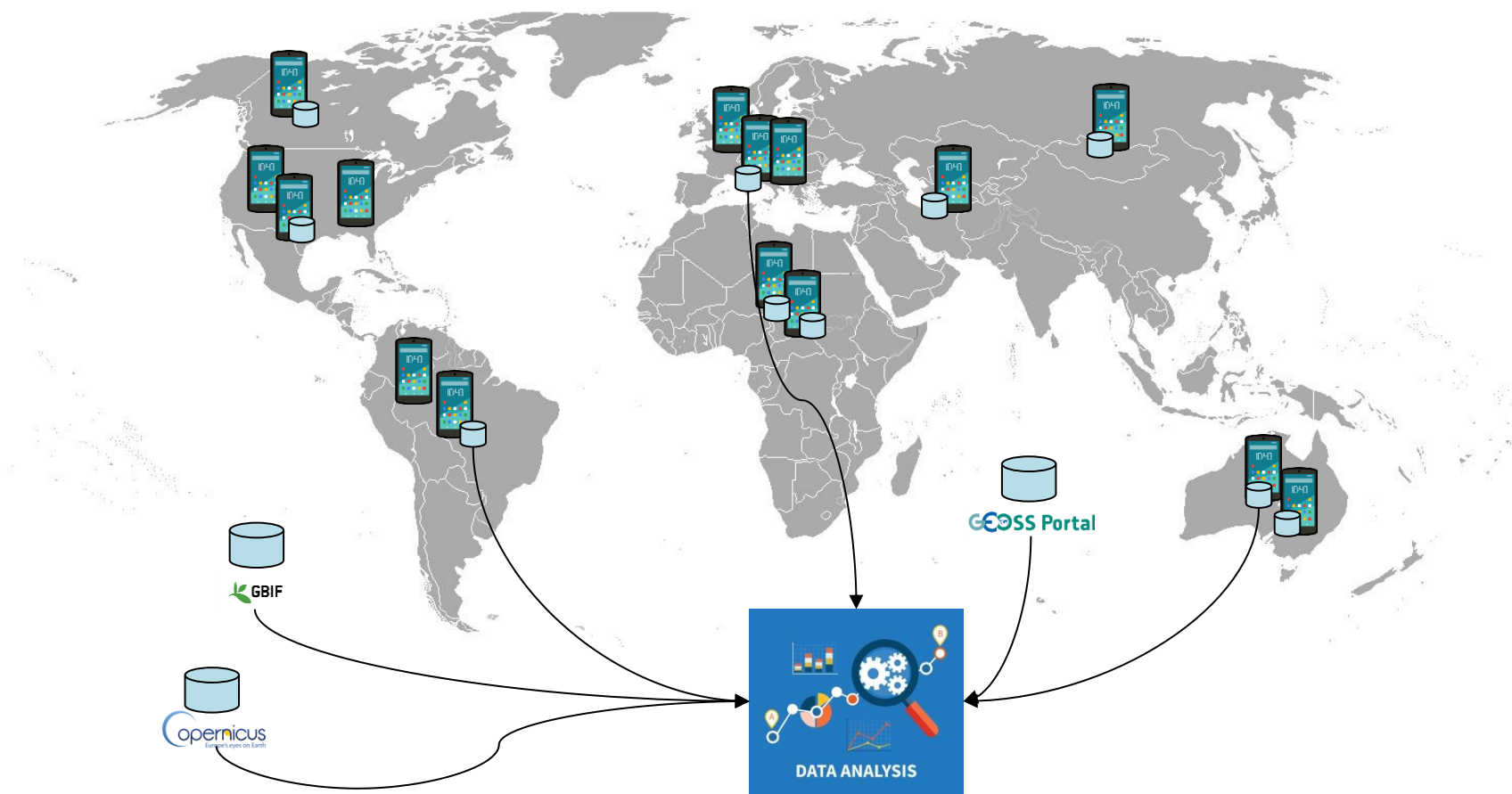
EGU2019-19123



Knowledge for Tomorrow

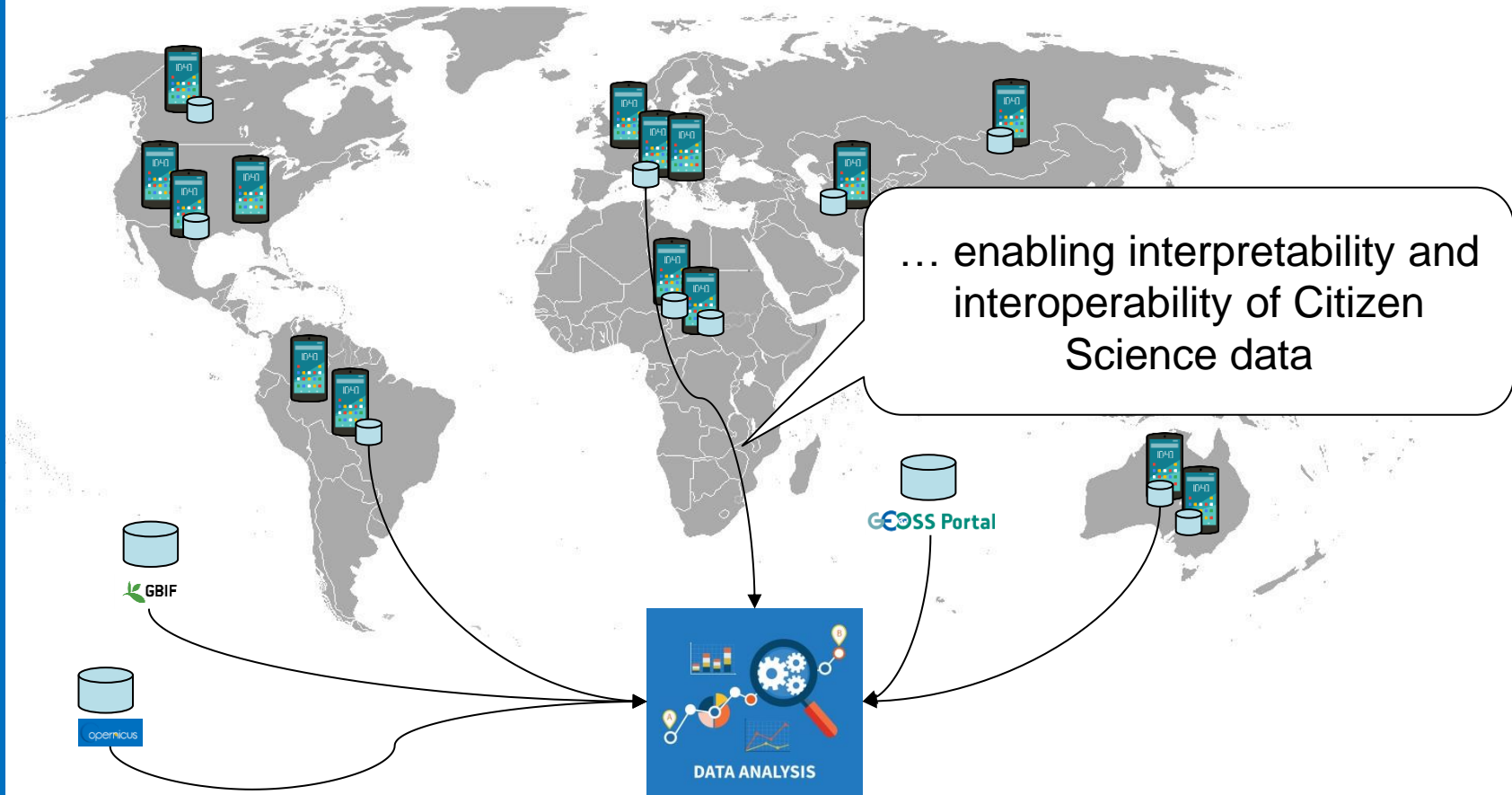


# Unlocking the Value of Citizen Contributions by ...

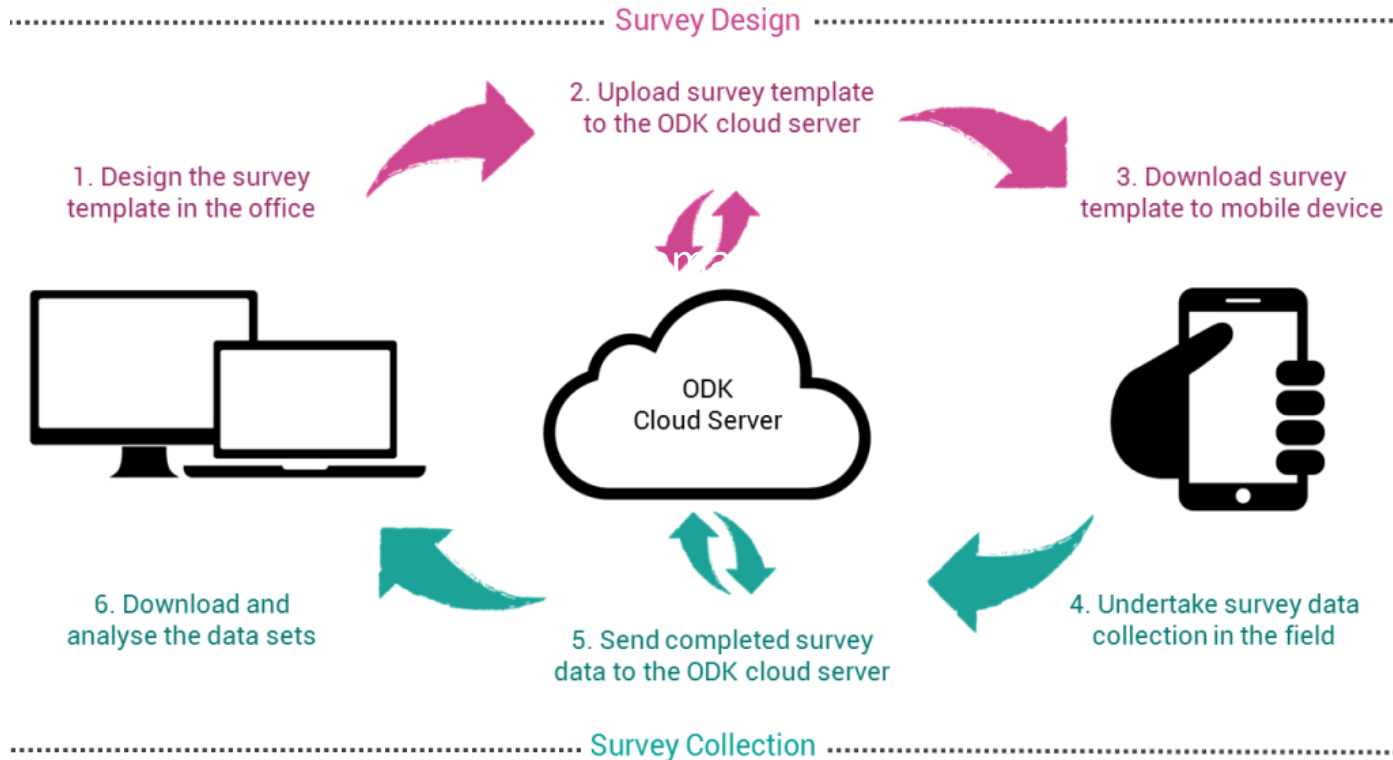


# Unlocking the Value of Citizen Contributions by ...

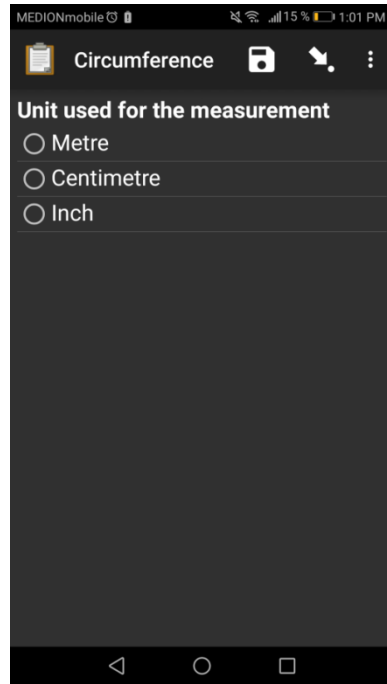
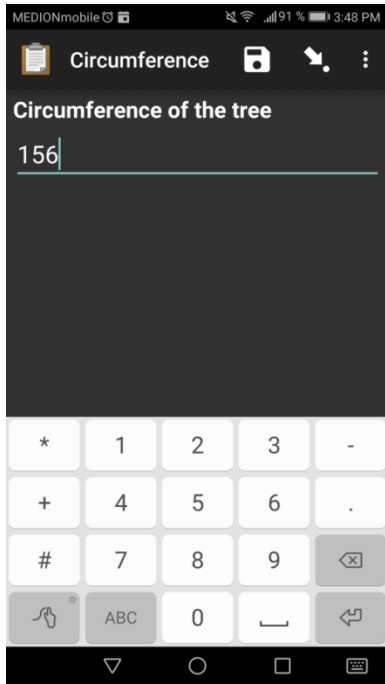
Reusable Software & Best Practices



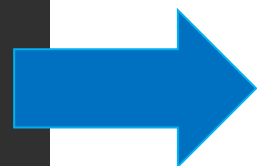
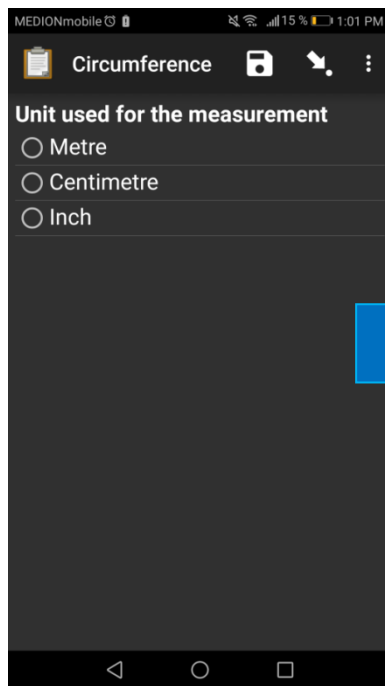
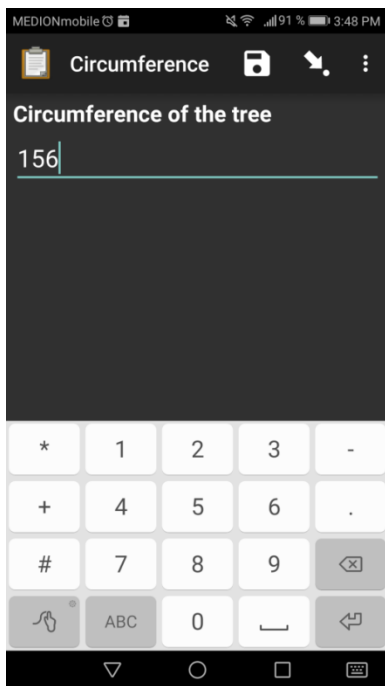
# Mobile Data Collection (with ODK)



# Enabling CS Data Interoperability & Reuse

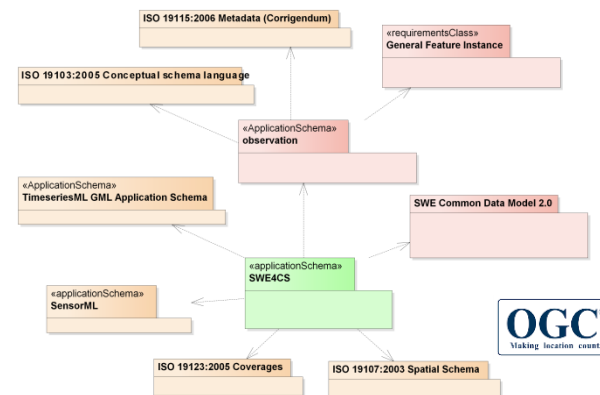
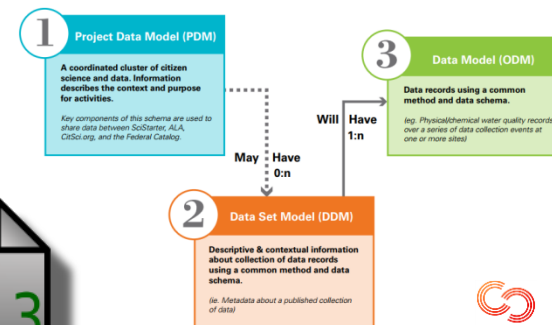


# Enabling CS Data Interoperability & Reuse



## Introducing PPSR CORE: A Common Data Model (CDM) with three main schemas

### Overview



# (1) Semantic Annotation at Design Time

The screenshot shows a web-based data entry tool interface. At the top, there is a header with the text "Circumference" and a "rename" link, followed by menu items: "File", "Edit", "View", and "Help". The main area contains three stacked input fields, each with a close button (X) on the right:

- Field 1: "Please enter your name" with a sub-label "name" and an "Abc" icon.
- Field 2: "Circumference of the tree" with a sub-label "circumference" and a "123" icon.
- Field 3: "Unit used for the measurement" with a sub-label "unit" and a "000" icon.

At the bottom of the main area, there is a toolbar with the following options: "+ Add new", "Text", "Numeric", "Date/Time", "Time", "Location", "Media", "Barcode", "Choose One", "Select Multiple", "Metadata", and "Group".

On the right side, there is a configuration panel for the selected field. It starts with "Signed in as markus3. Sign out." and includes several checkboxes: "Read Only", "Required", and "Valid Range". Under "Valid Range", there are input fields for "Minimum" and "Maximum", each with an "Inclusive" checkbox. Below these are fields for "Invalid Text" and "English". There are also dropdown menus for "Style" (set to "Textbox") and "Kind" (set to "Decimal").

The "Advanced" section is collapsed. The "Semantic Properties" section is expanded, showing:

- Creator: "\_col\_name"
- Unit: "\_col\_unit"
- Characteristic: "\_onto\_http://example.org/circur"

At the bottom of the configuration panel, there is an "Information: Characteristic" section with the text: "Information about the Characteristic is helpful for the RDF-Export of ODK Aggregate".



# (1) Semantic Annotation at Design Time

Circumference rename
File Edit View Help

Abc Please enter your name  
name

123 Circumference of the tree  
circumference

001 Unit used for the measurement  
unit

```

graph TD
    OC[oboe:ObservationCollection] -- memberOf --> O[oboe:Observation]
    O -- hasMember --> OC
    O -- ofEntity --> E1[oboe:Entity]
    O -- measurementFor --> M[oboe:Measurement]
    M -- hasMeasurement --> O
    M -- ofCharacteristic --> C[oboe:Characteristic]
    M -- hasValue --> E2[oboe:Entity]
    M -- usesProtocol --> P[oboe:Protocol]
    
```

Signed in as markus3. Sign out.

Read Only

Required

Valid Range

Minimum

Inclusive

Maximum

Inclusive

Invalid Text

English

Style

Textbox

Kind

Decimal

---

Advanced

Semantic Properties

Creator

Unit

Characteristic

---

Information: Characteristic

Information about the Characteristic is helpful for the RDF-Export of ODK Aggregate





# (1) Semantic Annotation at Design Time

The screenshot shows a form design tool interface. At the top, there is a menu bar with 'File', 'Edit', 'View', and 'Help'. Below the menu, there are three question cards: 'Please enter your name' (with 'name' as a label), 'Circumference of the tree' (with 'circumference' as a label), and 'Unit used for the measurement' (with 'unit' as a label). The 'Circumference of the tree' card is highlighted with a yellow border. On the right side, there is a configuration panel for the selected question. It includes a 'Signed in as markus3. Sign out.' header, a list of validation options (Read Only, Required, Valid Range), a 'Style' dropdown set to 'Textbox', and a 'Kind' dropdown set to 'Decimal'. Below these are 'Advanced' and 'Semantic Properties' sections. The 'Semantic Properties' section contains fields for 'Creator' (with the value 'n|'), 'Unit', and 'Characteristic'. A blue arrow points from the 'Characteristic' field in the 'Semantic Properties' section to the 'Characteristic' field in the 'Advanced' section of the configuration panel. At the bottom left, there is a 'DLR' logo. At the bottom right, there is a satellite map of Europe.

Circumference rename | File Edit View Help

Please enter your name  
name

Circumference of the tree  
circumference

Unit used for the measurement  
unit

Signed in as markus3. Sign out.

Read Only  
 Required  
 Valid Range  
Minimum  
 Inclusive  
Maximum  
 Inclusive

Invalid Text  
English

Style  
Textbox

Kind  
Decimal

Advanced

Semantic Properties

Creator  
n|

Unit  
name (Reference to question)

Characteristic

Information: Characteristic  
Information about the Characteristic is helpful for the RDF-Export of ODK Aggregate

+ Add new Text Numeric Date/Time Time Location

DLR

# (1) Semantic Annotation at Design Time

## Warning

The RDF export on this server uses some semantic properties that your ODK Build is currently not providing. You can select which to add to your surveys. You can also select the RDF-export templates you are interested in and the related properties will be selected for you.

Select all   Deselect all

- Creator
- Unit
- Characteristic

Extensible Observation Ontology

Add selected


The following information is **required** by one or more templates and is currently missing:


The following information is **optional** for one or more templates and is currently missing:


Upload anyway   Cancel upload



# (1) Semantic Annotation at Design Time


 Please enter your name  
 name  
 required


 Circumference of the tree  
 circumference


 Unit used for the measurement  
 unit

Properties

Data Name  
unit

Label  
English  
Unit used for the measurement

Hint  
English

Default Value

Read Only

Required

Options bulk edit

Option 1 X  
English  
Underlying Value  
\_onto\_om\_Metre

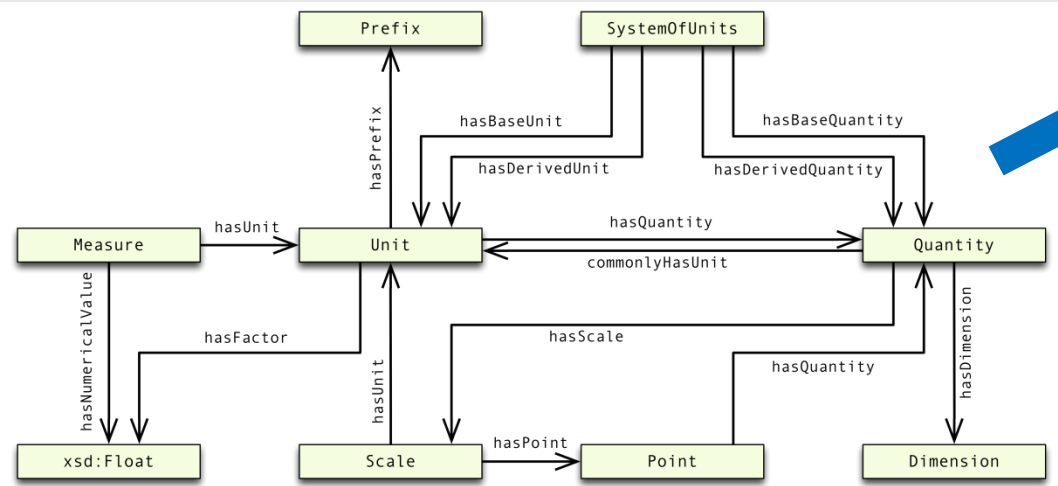
Option 2 X  
English  
Underlying Value  
\_onto\_om\_Centimetre

Option 3 X  
English  
Underlying Value  
\_onto\_om\_Inch-international

Information: Choose One

Allows the form filler to choose one from a set of predetermined choices.

To create a freeform "Other" option, add one here, then follow it up with a Text question with the Relevance property set to reference this question, for example:  
`$(xpath/to/this/choose/one) = 'other'`



```

graph TD
    Measure -- hasUnit --> Unit
    Unit -- hasNumericalValue --> Float[xsd:Float]
    Unit -- hasFactor --> Scale
    Unit -- hasPrefix --> Prefix
    Unit -- hasBaseUnit --> SystemOfUnits
    Unit -- hasDerivedUnit --> SystemOfUnits
    SystemOfUnits -- hasBaseQuantity --> Quantity
    SystemOfUnits -- hasDerivedQuantity --> Quantity
    Unit -- hasQuantity --> Quantity
    Quantity -- commonlyHasUnit --> Unit
    Scale -- hasScale --> Quantity
    Scale -- hasPoint --> Point
    Quantity -- hasQuantity --> Point
    Quantity -- hasDimension --> Dimension
  
```



# (1) Semantic Annotation at Design Time

The image illustrates the design-time semantic annotation process for a form field. It shows a form with three fields: "Please enter your name", "Circumference of the tree", and "Unit used for the measurement". The "Unit used for the measurement" field is highlighted in orange. A "Properties" panel on the right shows settings for the "unit" data name, including "Label", "Hint", "Default Value", and "Options". A blue arrow points from the "Options" list in the Properties panel to the "Options Editor" window, which shows a table of units with their underlying values and URIs. Another blue arrow points from the "Options Editor" to the "Options" list in the main form design tool.

**Properties Panel:**

- Data Name: unit
- Label: English, Unit used for the measurement
- Hint: English
- Default Value:
- Read Only
- Required
- Options:
  - Option 1: English, Metre, Underlying Value: \_onto\_om\_Metre
  - Option 2: English, Centimetre, Underlying Value: \_onto\_om\_Centimetre
  - Option 3: English, Inch, Underlying Value: \_onto\_om\_Inch-international

**Options Editor:**

Options	Underlying Value	English
	_onto_http_----www.ont.cchain	
	_onto_http_----www.ont.colcero	
	_onto_http_----www.ont.fathom (US survey)	
	_onto_http_----www.ont.fermi	
	_onto_http_----www.ont.foot (international)	
	_onto_http_----www.ont.foot (US survey)	
	_onto_http_----www.ont.furlong (international)	
	_onto_http_----www.ont.dinch (international)	
	_onto_http_----www.ont.metre	
	_onto_http_----www.ont.micrometre	
	_onto_http_----www.ont.millimetre	
	_onto_http_----www.ont.centimetre	
	_onto_http_----www.ont.decimetre	
	_onto_http_----www.ont.decametre	
	_onto_http_----www.ont.hectometre	
	_onto_http_----www.ont.kilometre	

**Options List (Main Form Design Tool):**

- Option 1: English, Metre, Underlying Value: \_onto\_http\_----www.ontology-of-
- Option 2: English, Centimetre, Underlying Value: \_onto\_http\_----www.ontology-of-
- Option 3: English, Inch, Underlying Value: \_onto\_http\_----www.ontology-of-

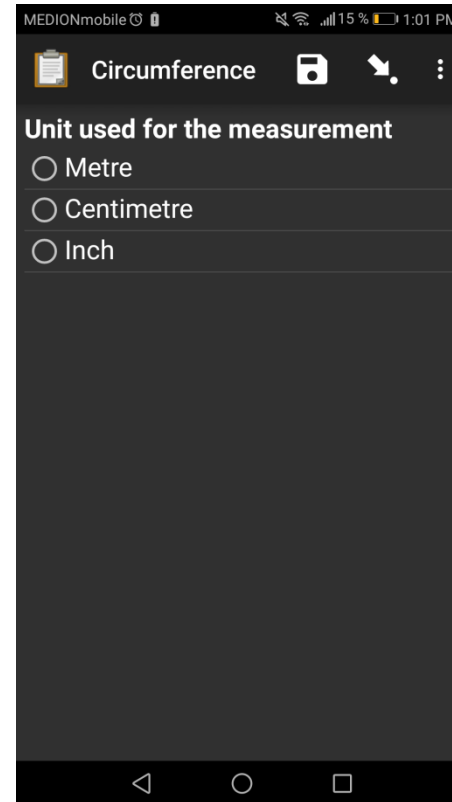
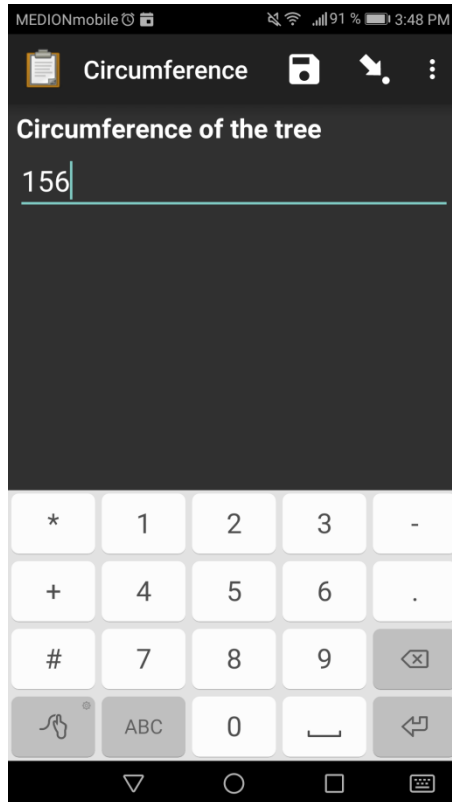
**Information: Choose One**

Allows the form filler to choose one from a set of predetermined choices.

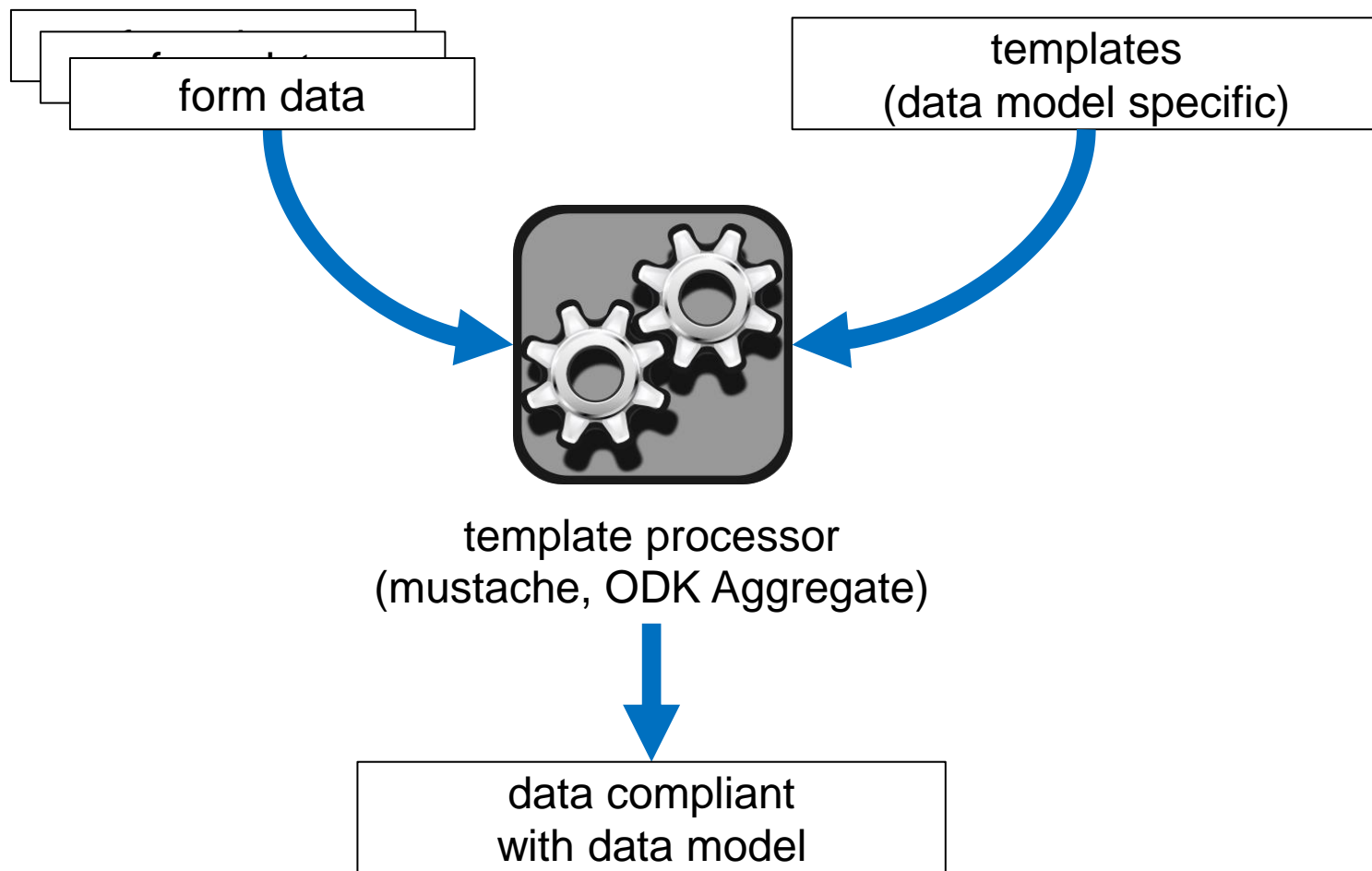
To create a freeform "Other" option, add one here, then follow it up with a Text question with the Relevance property set to reference this question; for example:  
`$(/xpath/to/this/choose/one) = 'other'`



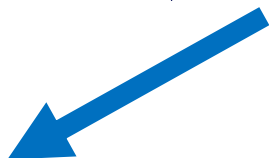
## (2) Transparent during Data Collection



## (3) Flexible Data Export via Templates



## (3) Flexible Data Export via Templates



```

{{!Construct an oboe-core:Measurement for the cell}}
{{cellEntityIdentifier}} rdf:type oboe-core:Measurement .

{{!Relate the Measurement to the Observation}}
{{columnModel.columnEntityIdentifier}} oboe-core:hasMeasurement {{cellEntityIdentifier}} .

{{!Construct an oboe-core:MeasuredValue for the cell}}
:measuredValue_{{rowModel.rowId}}_{{columnModel.columnHeader}}
  rdf:type oboe-core:MeasuredValue .

{{!Relate the Measurement to the MeasuredValue}}
{{cellEntityIdentifier}}
  oboe-core:hasValue :measuredValue_{{rowModel.rowId}}_{{columnModel.columnHeader}} .

```

```

{{^semantics.Unit.isLiteral}}
  {{cellEntityIdentifier}} om:hasUnit <{{semantics.Unit.value}}> .
{{/semantics.Unit.isLiteral}}

```

```

:measurement_1_circumference rdf:type oboe-core:Measurement .
:column_circumference oboe-core:hasMeasurement
:measurement_1_circumference .
:measuredValue_1_circumference rdf:type oboe-core:MeasuredValue .
:measurement_1_circumference oboe-core:hasValue
:measuredValue_1_circumference .

:measurement_1_circumference dcterms:creator
:creator_circumference_1 .
:creator_circumference_1 rdf:type foaf:Person .
:creator_circumference_1 foaf:name "Markus" .

:measurement_1_circumference
  oboe-core:ofCharacteristic <http://example.org/circumference>
.

:measurement_1_circumference
  om:hasUnit <http://www.ontology-of-units-of-
measure.org/resource/om-2/centimetre> .

:measuredValue_1_circumference oboe-core:hasCode
"156.0000000000"^^xsd:decimal .

```



## Contact

Markus Daniel Steinberg [markus.daniel.steinberg@uni-jena.de](mailto:markus.daniel.steinberg@uni-jena.de)

Sirko Schindler [sirko.schindler@dlr.de](mailto:sirko.schindler@dlr.de)

Friederike Klan [friederike.klan@dlr.de](mailto:friederike.klan@dlr.de)

## Publications

M. D. Steinberg, S. Schindler, F. Klan: [Software Solutions for Form-Based, Mobile Data Collection - A Comparative Evaluation](#),  
18th Conference on Datenbanksysteme für Business, Technologie und Web  
(BTW 2019), Rostock, 2019, <https://arxiv.org/abs/1901.11053>

M.D. Steinberg, S. Schindler, F. Klan: [Bringing Semantics to Citizen Data Collection - A Semantic Extension of Open Data Kit 1](#), Geophysical Research  
Abstracts, Vol. 21, EGU2019-19123, EGU General Assembly 2019

## Software

<https://github.com/MaSteinberg/aggregate>

<https://github.com/MaSteinberg/build>





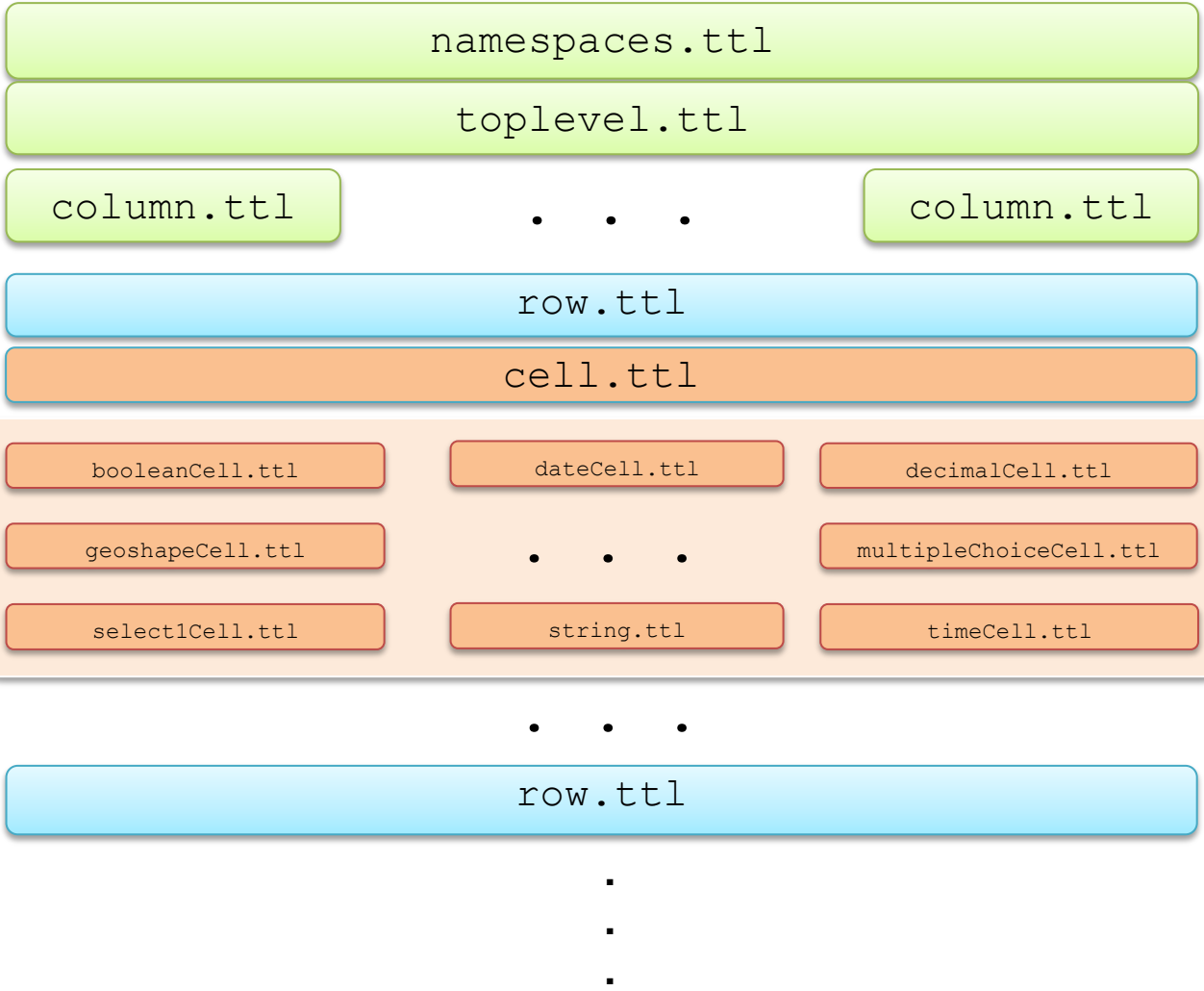
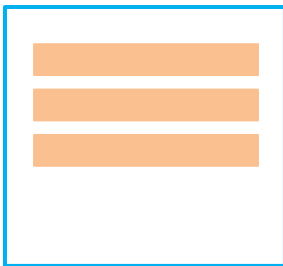
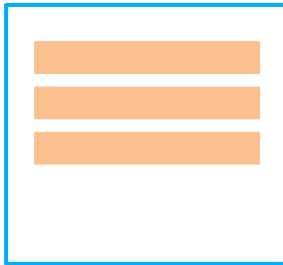
# Image Sources

- p.2/3:
  - <https://commons.wikimedia.org/wiki/File:BlankMap-World-v2.png>
  - <http://www.geoportal.org/documents/20181/24501/geoss-portal-beta.png/5b52dbab-a3c8-4962-8708-010ccff65c94?t=1480584717161>
  - <https://tr1.cbsistatic.com/hub/i/r/2015/09/30/f66fef72-5d57-48cc-bd8c-7ec0941e73ef/resize/770x/ac0363c3e414d51bafb1dd9b1c8a604a/dataanalysisistockrobuart.jpg>
  - <https://www.eea.europa.eu/about-us/who/copernicus-1/copernicus-logo/image>
  - <https://images.ctfassets.net/uo17ejk9rkwi/4rmEF9F4ZGiCwQk2WSMcce/b47146eacaf7b0dfc166656678c7ffe6/GBIF-2015.png>
- p.4/15:
  - <https://opendatakit.org/>
- p.6:
  - <https://openclipart.org>
  - A. Bowser, P. Brenton, R. Stevenson, G. Newman, S. Schade, L. Bastin, A. Parker, J. Oliver: Citizen Science Association Data & Metadata Working Group: Report from CSA 2017 and Future Outlook, Wilson Center 2017
  - Ingo Simonis, Rob Atkinson: OGC Discussion Paper Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data, OGC 16-129, <http://www.opengis.net/doc/DP/16-129>, 2016
  - [http://www.opengeospatial.org/pub/www/files/OGC\\_Logo\\_2D\\_Blue\\_x\\_0\\_0.png](http://www.opengeospatial.org/pub/www/files/OGC_Logo_2D_Blue_x_0_0.png)
- p.8:
  - <https://www.w3.org/TR/vocab-ssn>, Fig.22
- p.11:
  - <https://github.com/HajoRijgersberg/OM/blob/master/images/OM2.0-UML-diagram.png>
- p.14:
  - <https://www.clipartfree.net>



# (3) Flexible Data Export via Templates

time



## (3) Flexible Data Export via Templates

### namespaces.ttl.mustache

```
@BASE <{{base}}> .
@PREFIX : <{{base}}> .
{{#namespaces}}
    @PREFIX {{prefix}}: <{{uri}}> .
{{/namespaces}}
@PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@PREFIX owl: <http://www.w3.org/2002/07/owl#> .
@PREFIX oboe-core: <http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#> .
...
```

### toplevel.ttl.mustache

```
{{toplevelEntityIdentifier}} rdf:type oboe-core:ObservationCollection .
```



## (3) Flexible Data Export via Templates

### column.ttl.mustache

```
{{columnEntityIdentifier}} rdf:type oboe-core:Observation .
{{topLevelModel.toplevelEntityIdentifier}} oboe-core:hasMember {{columnEntityIdentifier}} .
```

### cell.ttl.mustache

```
{{cellEntityIdentifier}}
  rdf:type          oboe-core:Measurement ;
  oboe-core:hasValue :measuredValue_{{rowModel.rowId}}_{{columnModel.columnHeader}} .
  om:hasUnit        <{{{semantics.Unit.value}}}> .
```

### decimalCell.ttl.mustache

```
:measuredValue_{{rowModel.rowId}}_{{columnModel.columnHeader}}
  oboe-core:hasCode "{{{cellValue}}}"^^xsd:decimal .
```



# Exported Data

```
:measurement_1_circumference rdf:type oboe-core:Measurement .
:column_circumference oboe-core:hasMeasurement :measurement_1_circumference .
:measuredValue_1_circumference rdf:type oboe-core:MeasuredValue .
:measurement_1_circumference oboe-core:hasValue :measuredValue_1_circumference .

:measurement_1_circumference dcterms:creator :creator_circumference_1 .
:creator_circumference_1 rdf:type foaf:Person .
:creator_circumference_1 foaf:name "Markus" .

:measurement_1_circumference oboe-core:ofCharacteristic <http://example.org/circumference> .
:measurement_1_circumference
  om:hasUnit <http://www.ontology-of-units-of-measure.org/resource/om-2/centimetre> .

:measuredValue_1_circumference oboe-core:hasCode "156.0000000000"^^xsd:decimal .
```



# Template Configuration

The following snippets show a parts of a template configuration:

## availableProperties:

Creator:

Endpoint:

Query:

Unit:

Endpoint: <http://192.168.0.8:7200/repositories/om>

Query: |-

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
PREFIX : <http://ecoinformatics.org/oboe/oboe.1.2/oboe.owl#>
```

```
PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>
```

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT DISTINCT ?uri ?displayName
```

```
WHERE {
```

```
  ?uri rdf:type om:Unit .
```

```
  OPTIONAL{
```

```
    ?uri rdfs:label ?displayName .
```

```
    FILTER (lang(?displayName) = 'en')
```

```
  }
```

```
}
```

```
displayName: Extensible Observation Ontology
```

```
templateProperties:
```

```
  optionalProperties:
```

```
    - Creator
```

```
  requiredProperties:
```

```
    - Characteristic
```

```
    - Unit
```

