Critical Apparatus Ontology
–
Adding Semantics
to XML/TEI Documents

Diego Canali for TALOS – AI4SSH
ERA Chair in Artificial Intelligence for Humanities and Social Sciences
Rethymno, 6 November 2023

Critical Apparatus Ontology

Ontology Specification Draft

Release 08.07.2019

IRI: https://w3id.org/cao
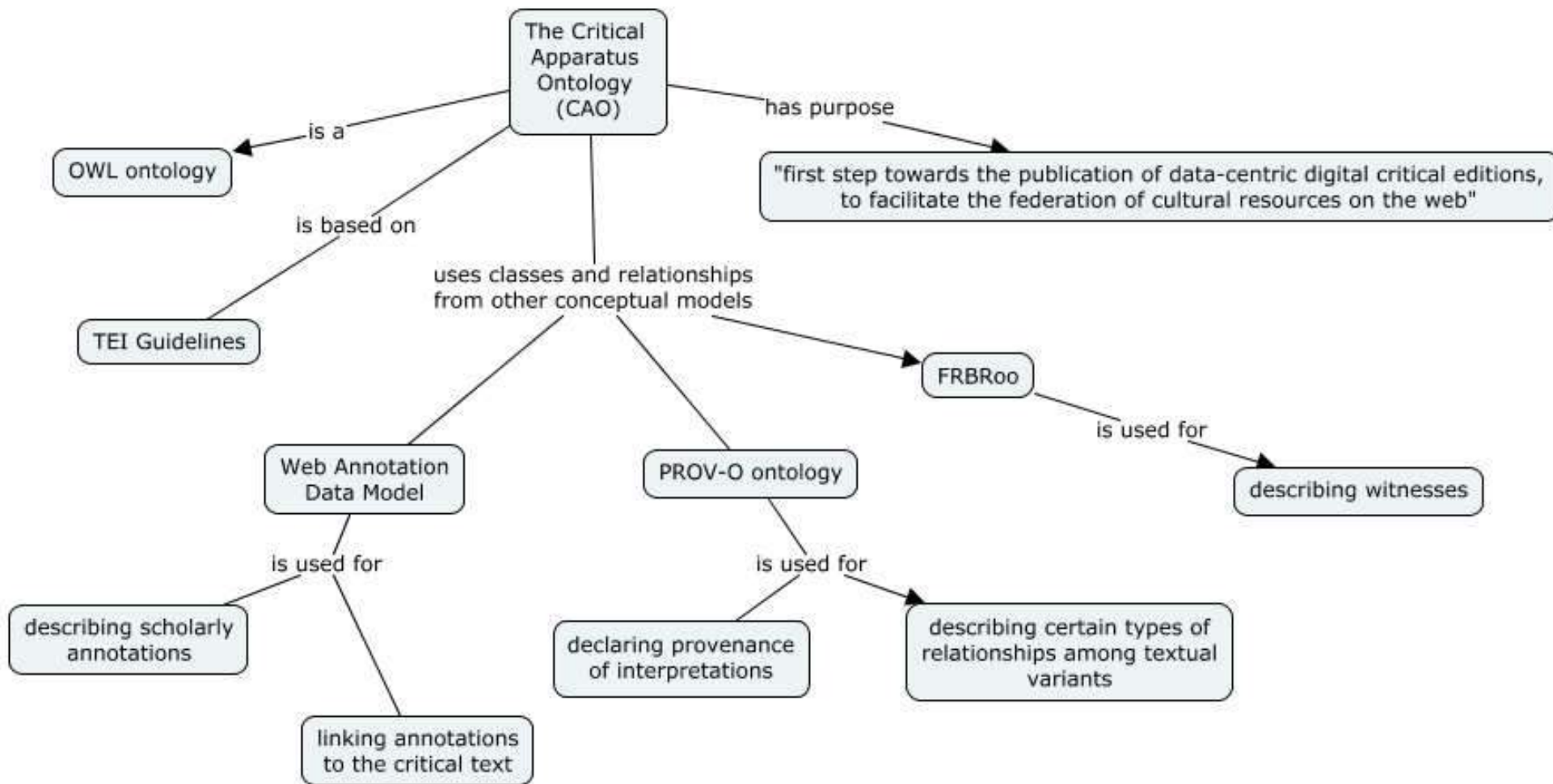
Current version: 0.9

Authors:Francesca Giovannetti

Imported Ontologies and models:

- *hico*
- *frbr*
- *naryparticipation.owl*
- *participation.owl*
- *situation.owl*
- *timeinterval.owl*
- *cpannotationschema.owl*
- *oa#*

1. Introduction

2. Description

- The Critical Apparatus Ontology (CAO) is an OWL 2 DL ontology for representing the apparatus of textual variants and philological arguments that accompanies a digital critical edition.

- It allows to express **RDF data** within an **.xml document** through **formal syntaxes** such as:
    - **RDF/XML**: https://www.w3.org/TR/rdf-syntax-grammar/
    - **Turtle**: https://www.w3.org/TR/rdf12-turtle/#abstract
    - **JSON-DL**: https://www.w3.org/TR/json-ld11/

- In this presentation, along with **F. Giovannetti's** practice, I will use **Turtle** syntax to provide examples.
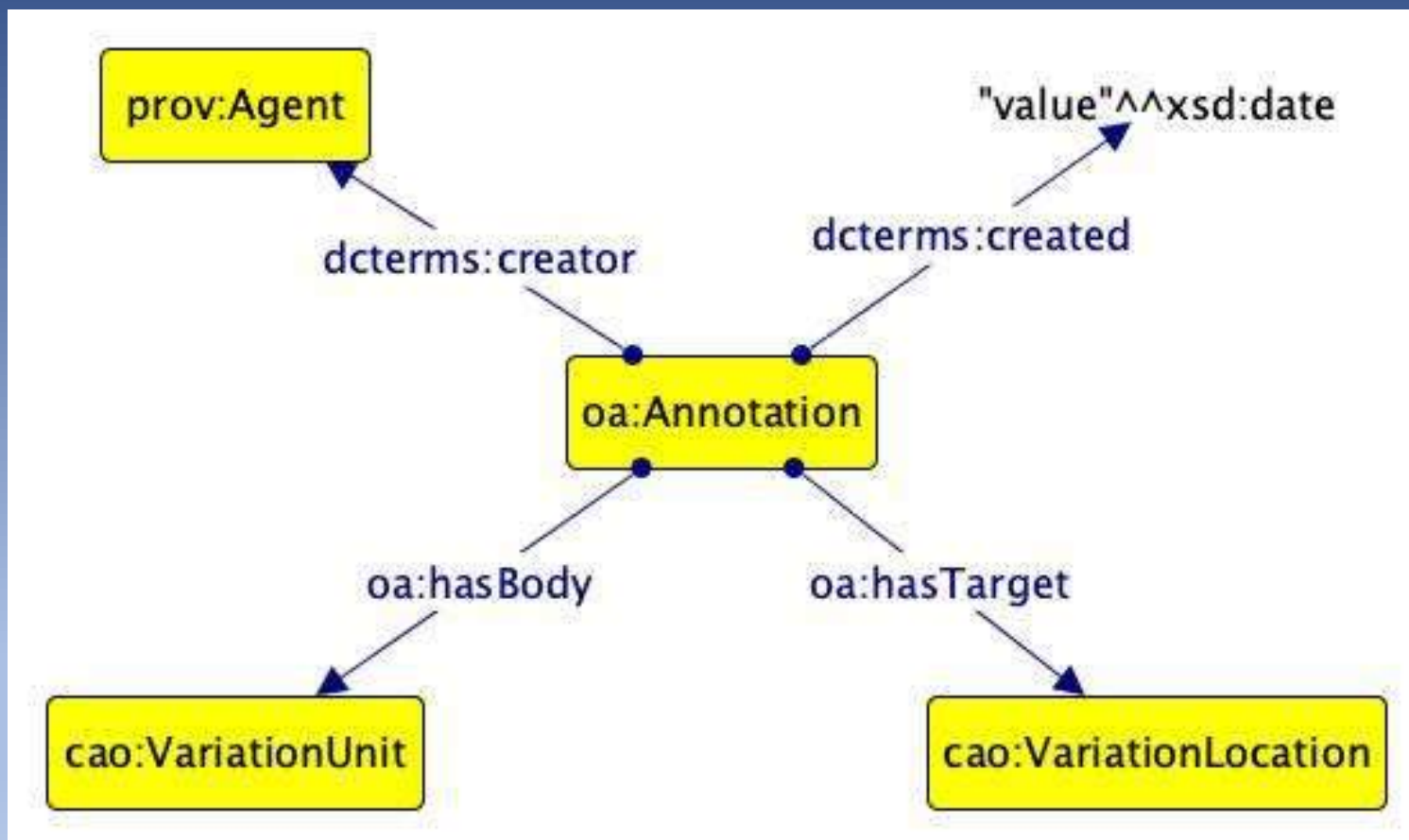
2. Description

- How to **represent the critical apparatus** using the CAO model:

    - From the creation of a **new apparatus entry**

    - To the **description** of the **readings** and their **relationships**

    - To the **formalization** of the **process** of **scholarly interpretation.**

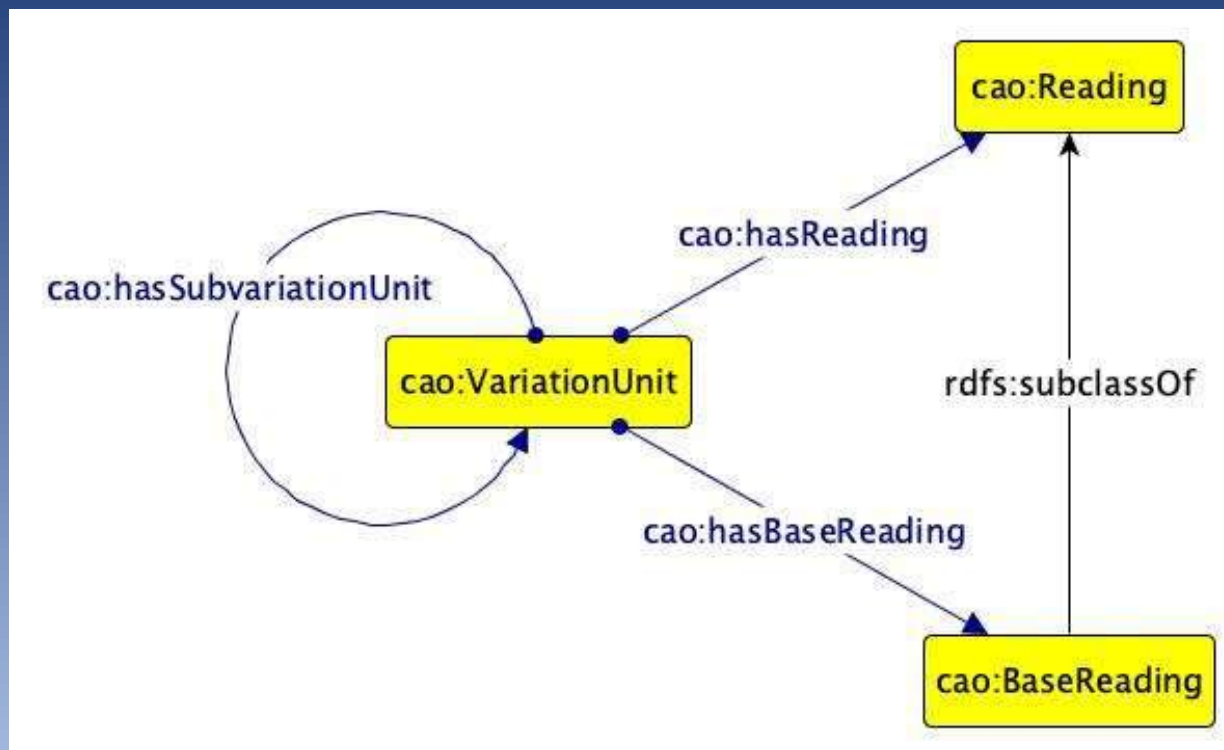# 2.1. Creation of a new apparatus entry

Web Annotation Ontology

(https://www.w3.org/TR/annotation-vocab/)

- **Body**
- **Target**
- **Creator**
- **Date of Creation**

"provenance"

2.2. the **body** of the annotation: the **apparatus entry**

Domain	Property	Range
cao:VariationUnit	cao:hasReading	cao:Reading
cao:VariationUnit	cao:hasBaseReading	cao:BaseReading
cao:VariationUnit	cao:hasSubvariationUnit	cao:VariationUnit
cao:BaseReading	rdfs:subclassOf	cao:Reading

2.1. Creation of a new apparatus entry

Example

```
@prefix ex: <http://example.org/> .
@prefix cao: <https://w3id.org/cao/> .
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix frbroo: <http://iflastandards.info/ns/fr/frbr/frbroo/> .
@prefix oa: <http://www.w3.org/ns/oa#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# the scholarly annotation
<annotation/anno01> a oa:Annotation ;
  # the apparatus entry
    oa:hasBody <app/app01> ;
    # the variation point within the edition TEI document
    oa:hasTarget <varloc/vl01> ;
    # the scholar's identifier
    dcterms:creator <resp/fgiov> ;
    # the date on which the apparatus entry was created
    dcterms:created "2019-03-09"^^xsd:date .
```
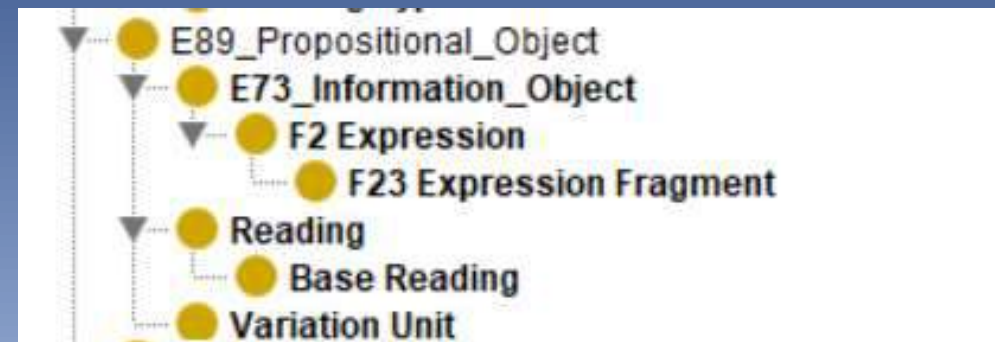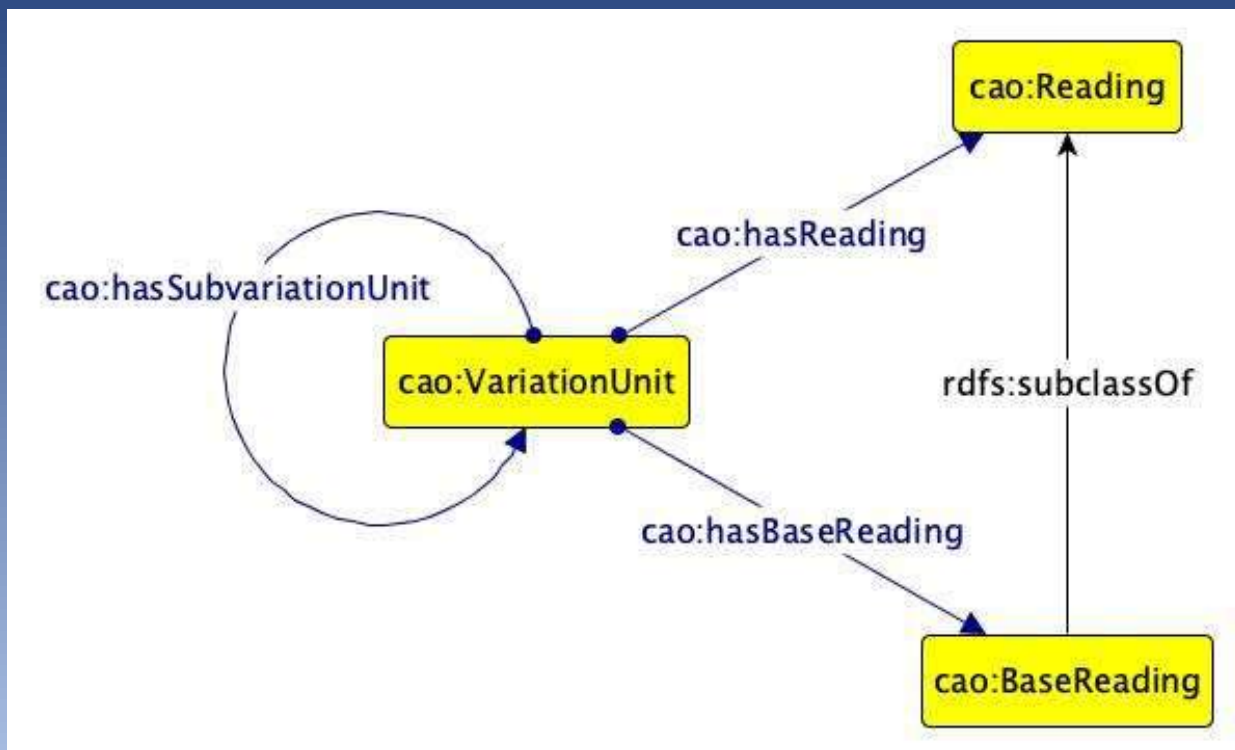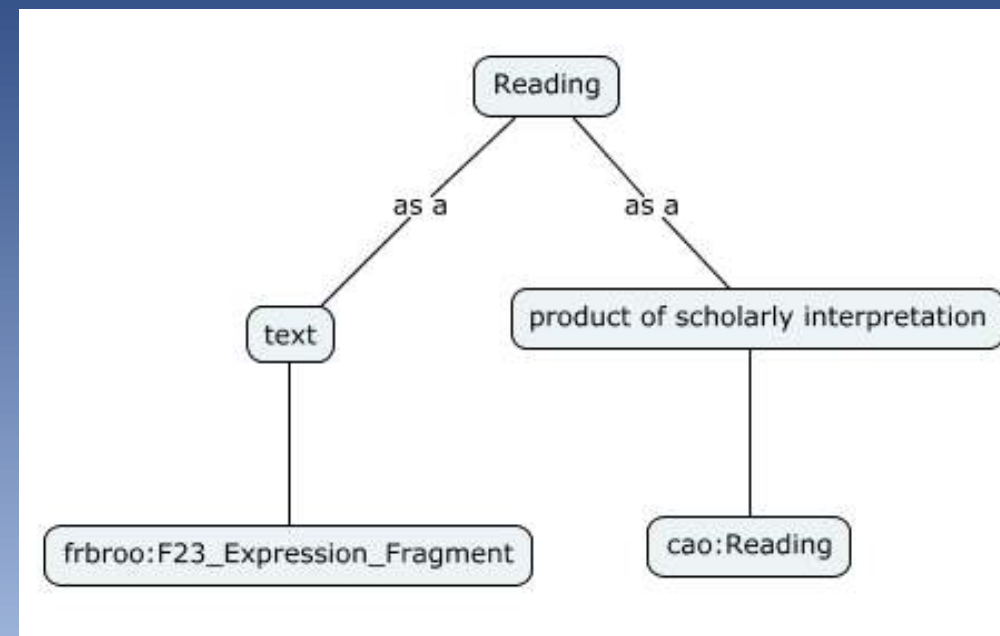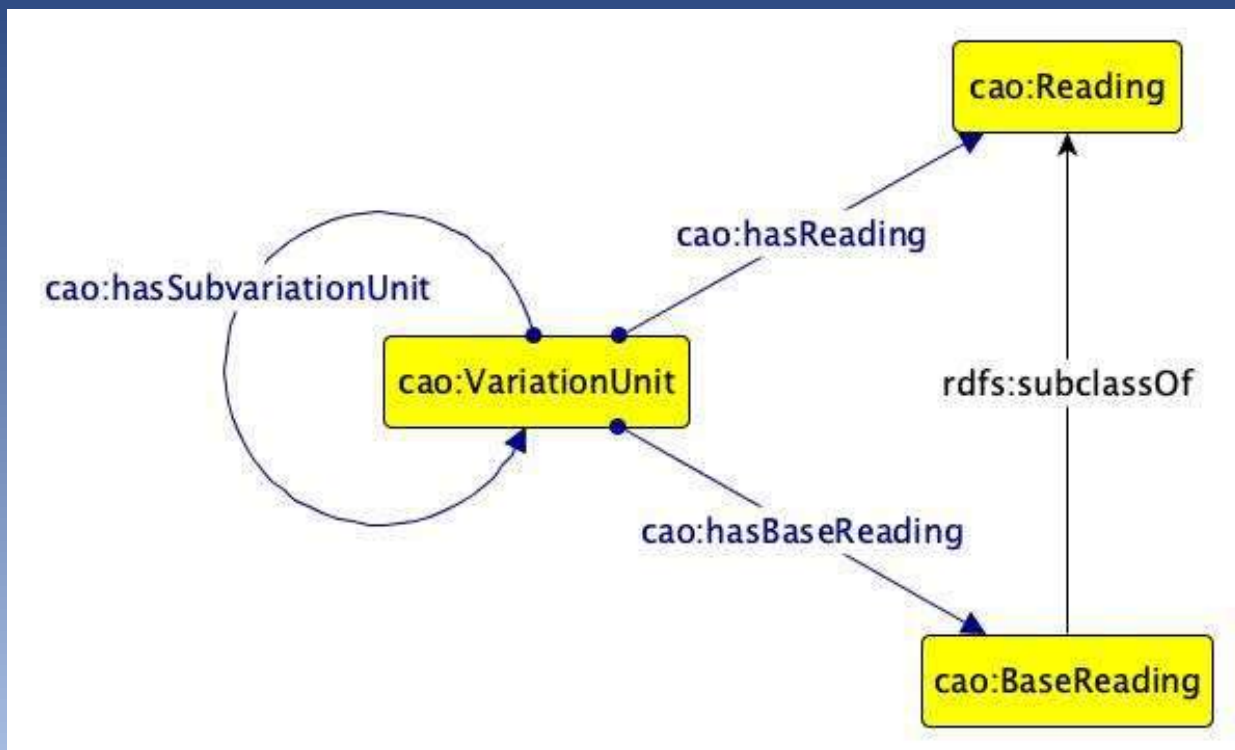
2.2. the body of the annotation: the apparatus entry

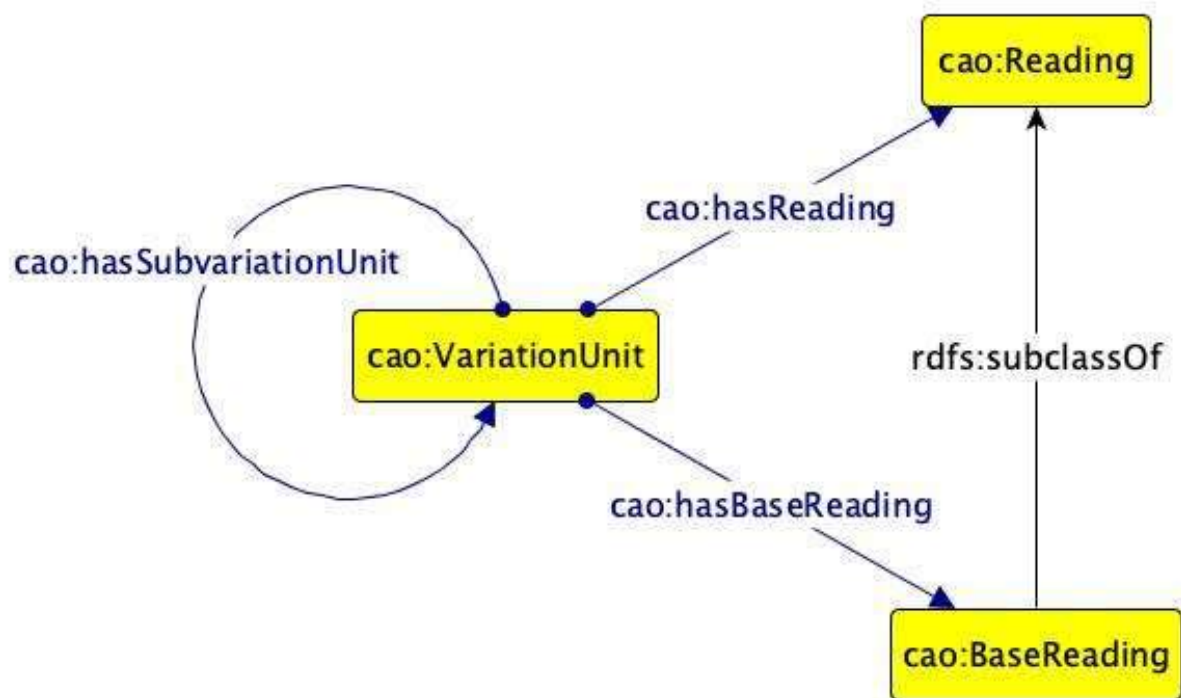| Domain	Property	Range
cao:VariationUnit	cao:hasReading	cao:Reading
cao:VariationUnit	cao:hasBaseReading	cao:BaseReading
cao:VariationUnit	cao:hasSubvariationUnit	cao:VariationUnit
cao:BaseReading	rdfs:subclassOf	cao:Reading

2.2. the body of the annotation: the apparatus entry

Domain	Property	Range
cao:VariationUnit	cao:hasReading	cao:Reading
cao:VariationUnit	cao:hasBaseReading	cao:BaseReading
cao:VariationUnit	cao:hasSubvariationUnit	cao:VariationUnit
cao:BaseReading	rdfs:subclassOf	cao:Reading

2.2. the __body__ of the annotation: the __apparatus entry__


```
# a unit of variation
<app/app01> a cao:VariationUnit ;
    # the base reading
    cao:hasBaseReading <rdg/anno01-rdg01> ;
    # the variant readings
    cao:hasReading <rdg/anno01-rdg02> ,
        <rdg/anno01-rdg03> .

<rdg/anno01-rdg01> a cao:BaseReading ;
    rdf:value "diffundi"^^xsd:string .

<rdg/anno01-rdg02> a cao:Reading ;
    rdf:value "diffudit"^^xsd:string .

<rdg/anno01-rdg03> a cao:Reading ;
    rdf:value "diffundit"^^xsd:string .

<app/a10> a cao:VariationUnit ;
    cao:hasSubvariationUnit <app/a10-1> ,
        <app/a10-2> .
```
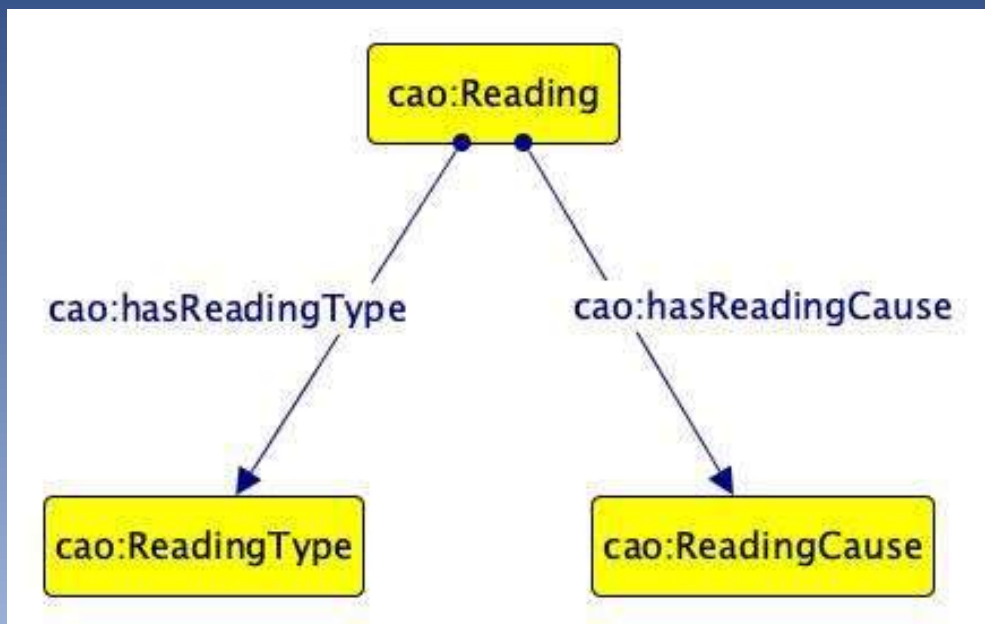
Domain	Property	Range
cao:VariationUnit	cao:hasReading	cao:Reading
cao:VariationUnit	cao:hasBaseReading	cao:BaseReading
cao:VariationUnit	cao:hasSubvariationUnit	cao:VariationUnit
cao:BaseReading	rdfs:subclassOf	cao:Reading

2.3. the <u>body</u> of the annotation: <u>classifying the readings</u> by <u>type</u> and <u>cause</u>
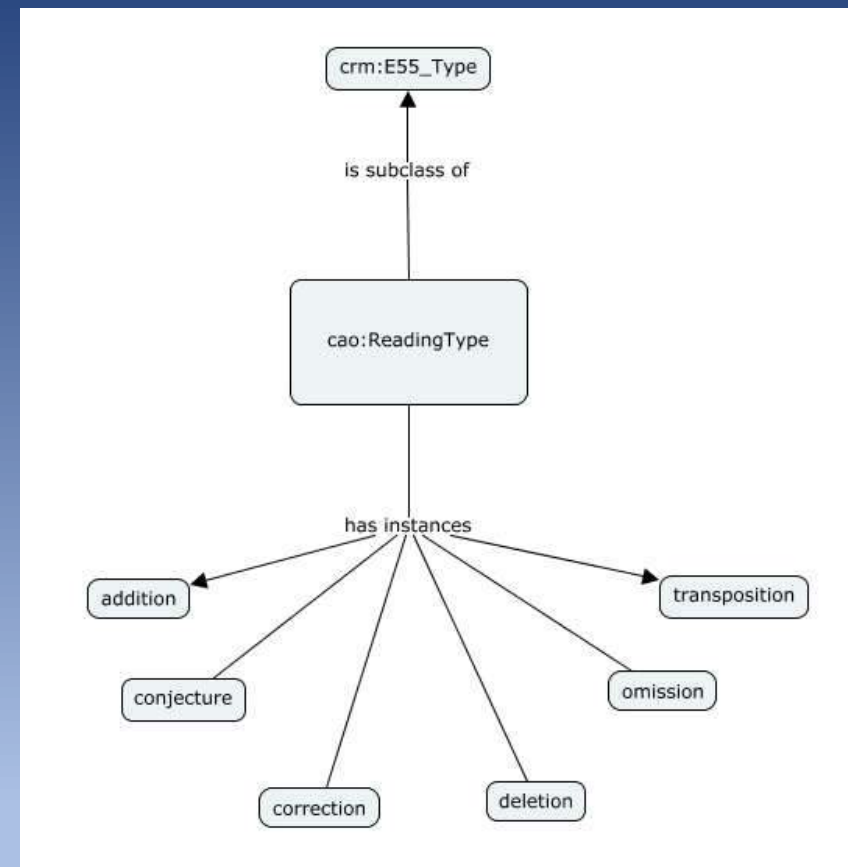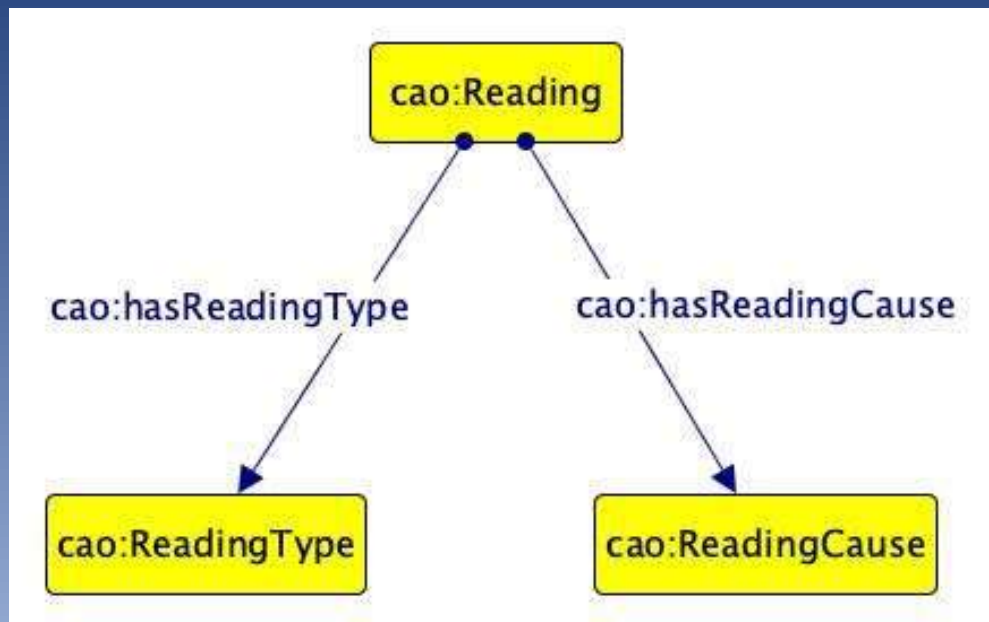

Same function of the TEI
@type
@cause

cao:ReadingType and **cao:ReadingCause** are direct subclasses of **crm:E55_Type**

cao:Reading	cao:hasReadingType	cao:ReadingType
cao:Reading	cao:hasReadingCause	cao:ReadingCause

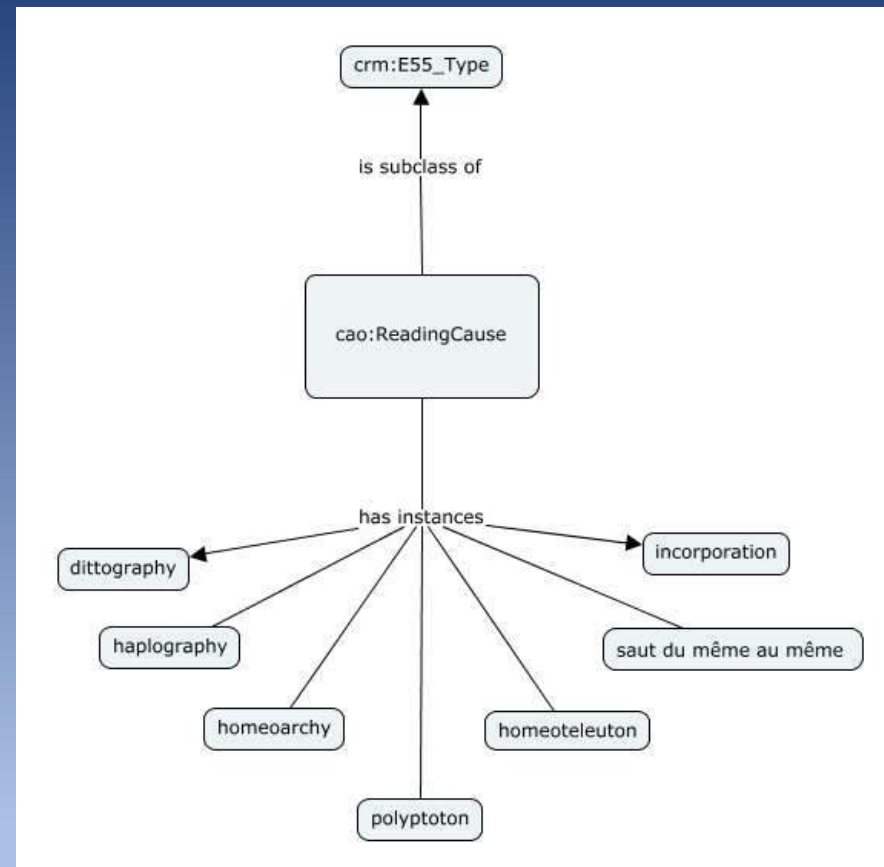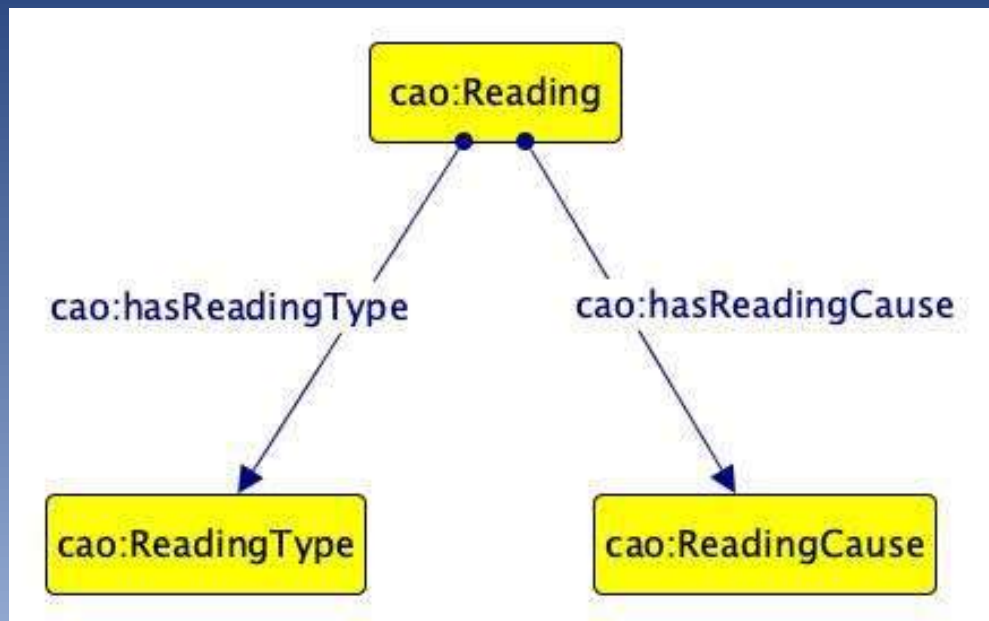# 2.3. the <u>body</u> of the annotation: <u>classifying the readings</u> by <u>type</u> and <u>cause</u>

Instances of **cao:ReadingType** are **individuals**

Each of them is matched with the corresponding entity of the **LAWD** (**Linked Ancient World Data ontology https://github.com/lawdi/LAWD**) reading classification through the property **owl:sameAs.**

```
<rdg/anno01-rdg03> a cao:Reading ;
  cao:hasReadingType cao:omission ;
  cao:hasReadingCause cao:homeoteleuton ;
  rdf:value ""^^xsd:string .
```
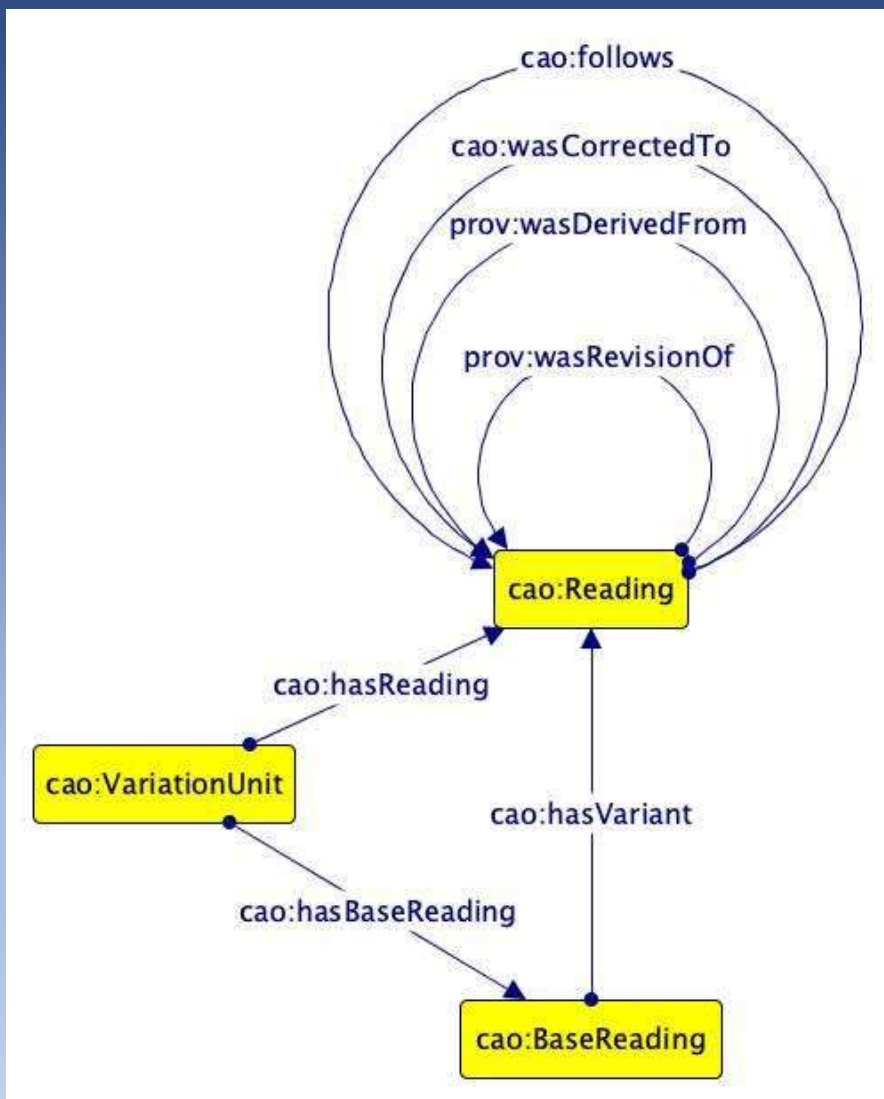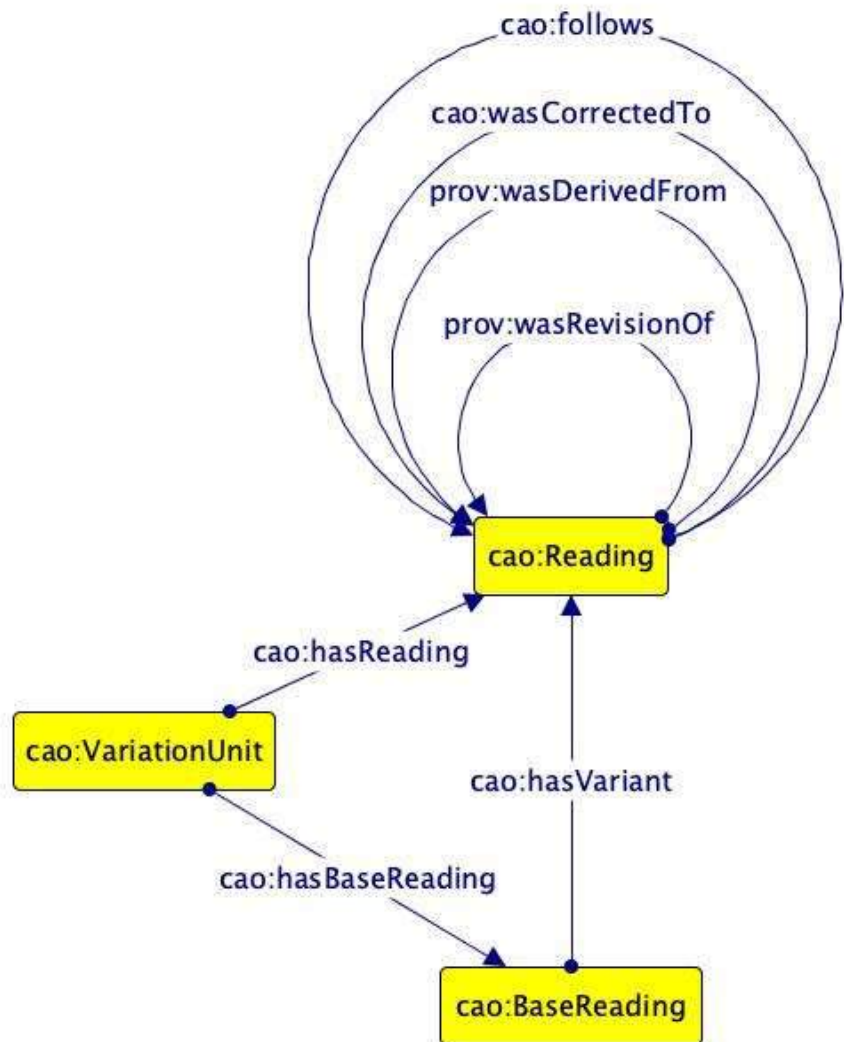
Domain	Property	Range
cao:VariationUnit	cao:hasBaseReading	cao:BaseReading
cao:VariationUnit	cao:hasReading	cao:Reading
cao:BaseReading	cao:hasVariant	cao:Reading
cao:Reading	prov:wasRevisionOf	cao:Reading
cao:Reading	prov:wasDerivedFrom	cao:Reading
cao:Reading	cao:wasCorrectedTo	cao:Reading
cao:Reading	cao:follows	cao:Reading

```
<app/app08> a cao:VariationUnit ;
  cao:hasReading <rdg/anno08-rdg01> ,
    <rdg/anno08-rdg02> ,
    <rdg/anno08-rdg03> .

<rdg/anno08-rdg01> a cao:Reading ;
  cao:wasCorrectedTo <rdg/anno08-rdg02> ;
  cao:hasReadingType cao:deletion ;
  rdf:value "this"^^xsd:string .

<rdg/anno08-rdg02> a cao:Reading ;
  cao:wasCorrectedTo <rdg/anno08-rdg03> ;
  cao:hasReadingType cao:deletion ;
  rdf:value "such a"^^xsd:string .

<rdg/anno08-rdg03> a cao:Reading ;
  cao:hasReadingType cao:addition ;
  rdf:value "a"^^xsd:string .
```
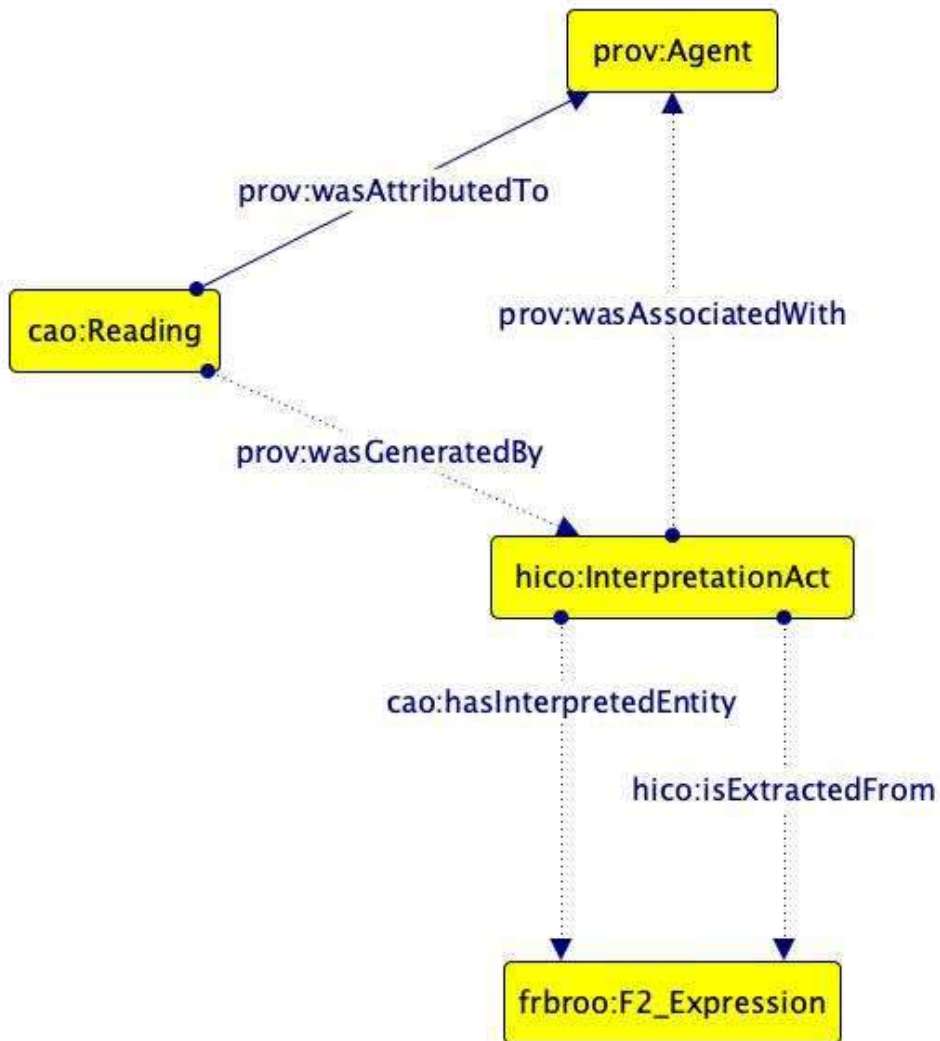
Domain	Property	Range
cao:Reading	prov:wasAttributedTO	prov:Agent
cao:Reading	prov:wasGeneratedBy	hico:InterpretationAct
hico:InterpretationAct	prov:wasAssociatedWith	prov:Agent
hico:InterpretationAct	cao:hasInterpretedEntity	frbroo:F2_Expression
hico:InterpretationAct	hico:isExtractedFrom	frbroo:F2_Expression
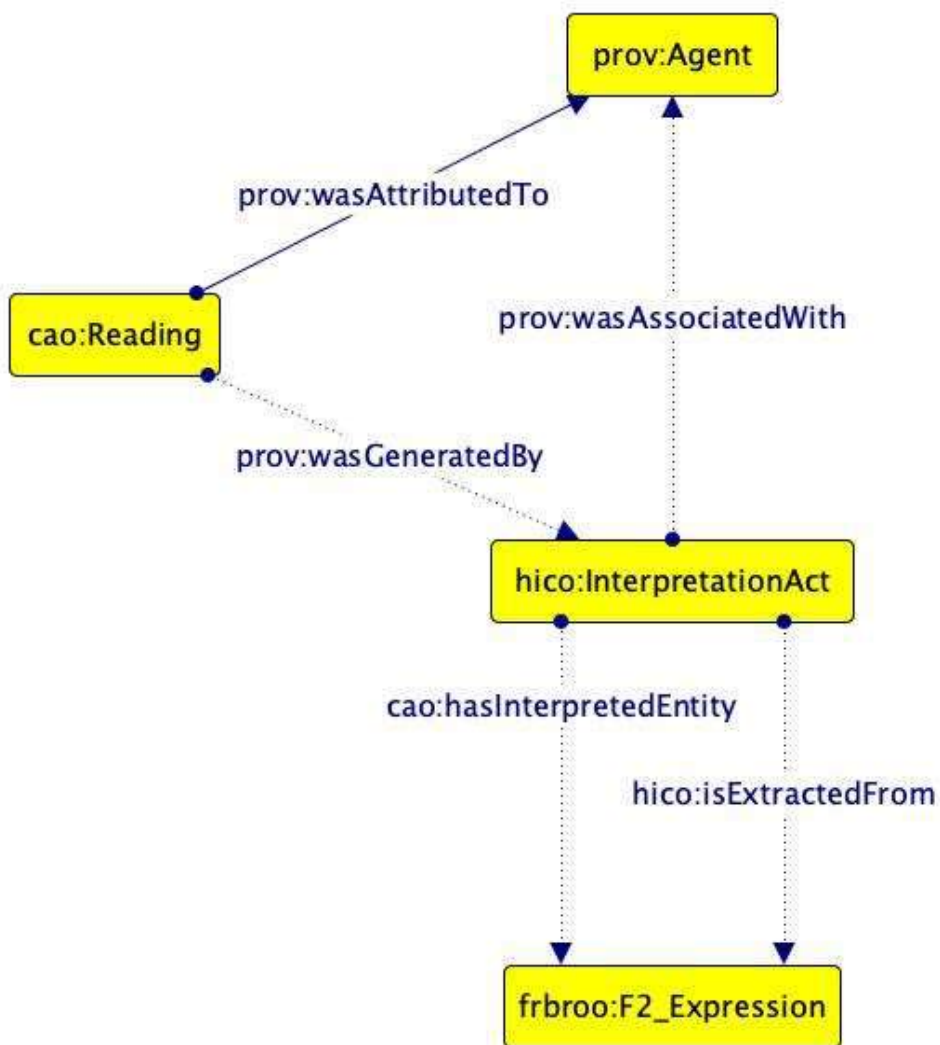
Citing other scholars' readings

In TEI, **<u>@source</u>** is used to indicate **<u>responsibility</u>** to say that a witness supports a particular reading.
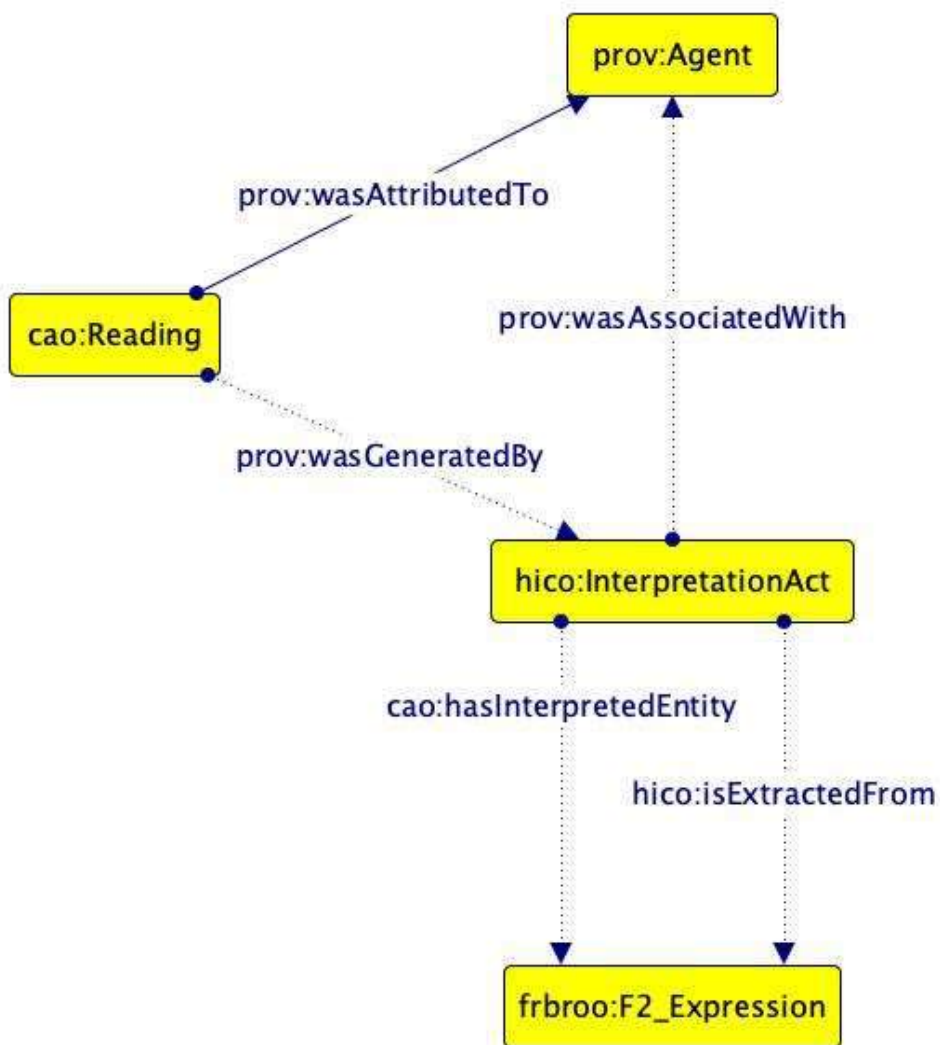
It contains a reference to an <u>external source</u> (a **published edition**).

In **CAO**, a **cited reading** is related to its original source by the property **<u>prov:hadPrimarySource</u>**.

```
<rdg/anno01-rdg02> a cao:Reading ;
    rdf:value "diffudit"^^xsd:string ;
    prov:hadPrimarySource <http://example.org> .
```

Attributing a reading to a specific hand

TEI uses **<u>@hand</u>**;

CAO the property **cao:hasAttributedHand**.

Its domain is **cao:Reading**, its range is **prov:Agent**

```
<rdg/anno01-rdg02> a cao:Reading ;
    rdf:value "diffudit"^^xsd:string ;
    cao:hasAttributedHand <resp/m1> .
```
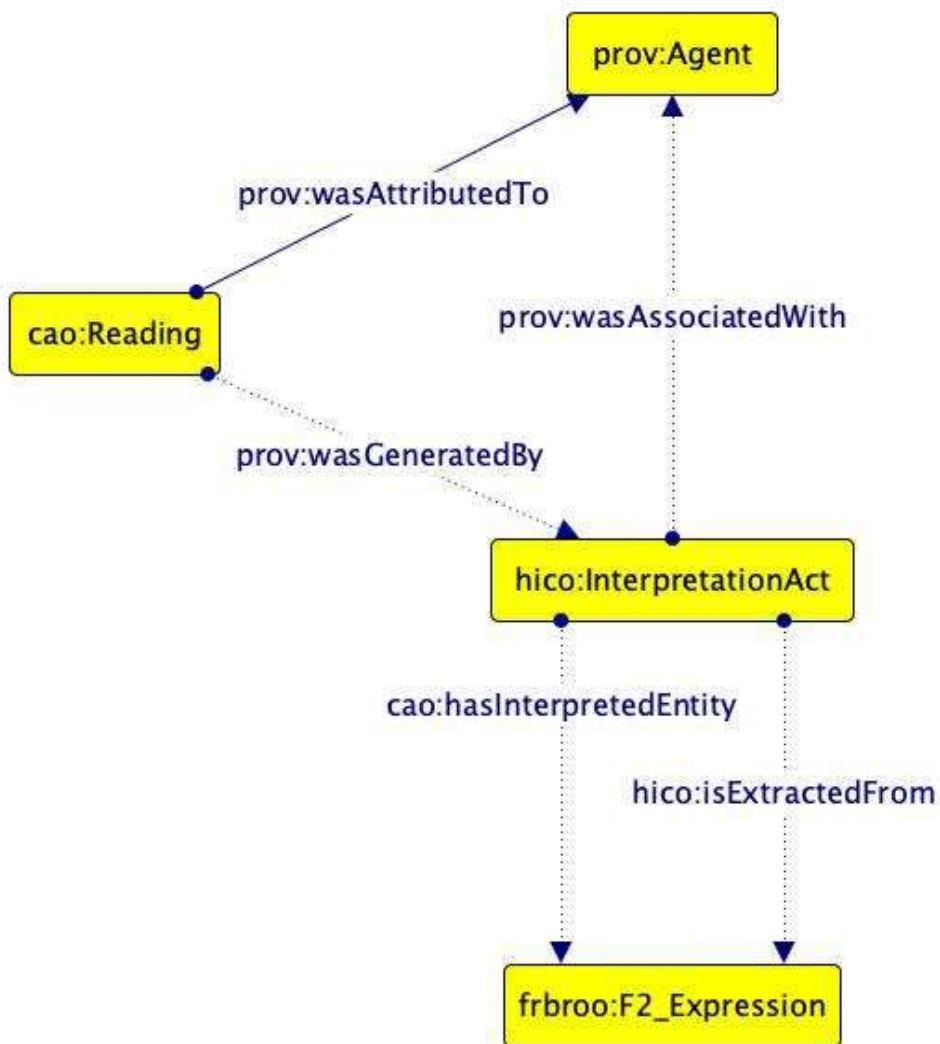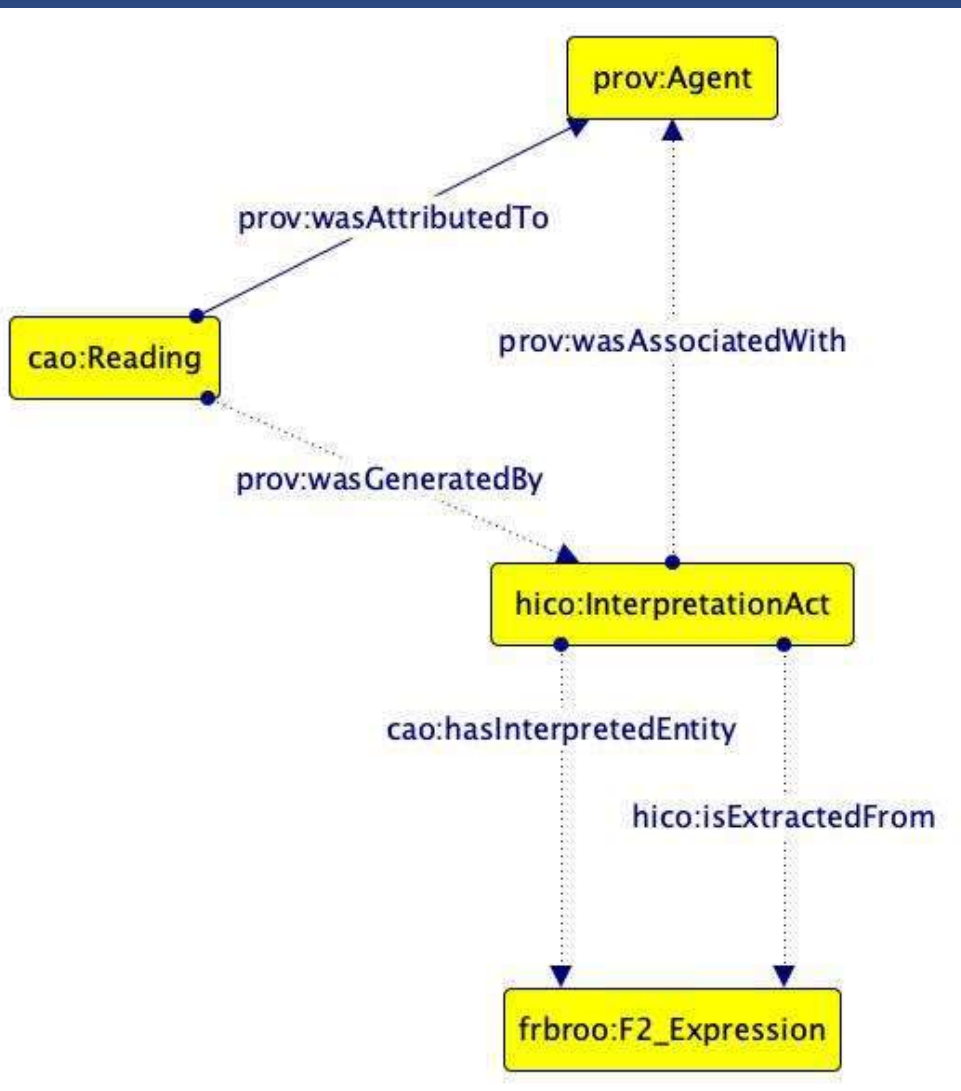
Representing the interpretation act which generated a reading

cao:Readings --> acts of scholarly interpretation: we use Classes and Properties from **HiCO** (Historical Context Ontology) (https://marilenadaquino.github.io/hico/).

Historical Context Ontology:

"It addresses features characterising <u>hermeneutical activities performed by scholars while generating new information</u> (i.e. an interpretation act). It allows to represent and reason on reliability of argumentations around attributions, by evaluating features such as motivations, types of cited sources or criteria, dates, relations with other claims (e.g. agreement/disagreement)".

"***HiCO*** extends the ***PROV Ontology*** with terms for describing aspects of the hermeneutical activity, and reuses existing ontologies, such as ***CiTO Ontology*** for linking attributions to related sources."
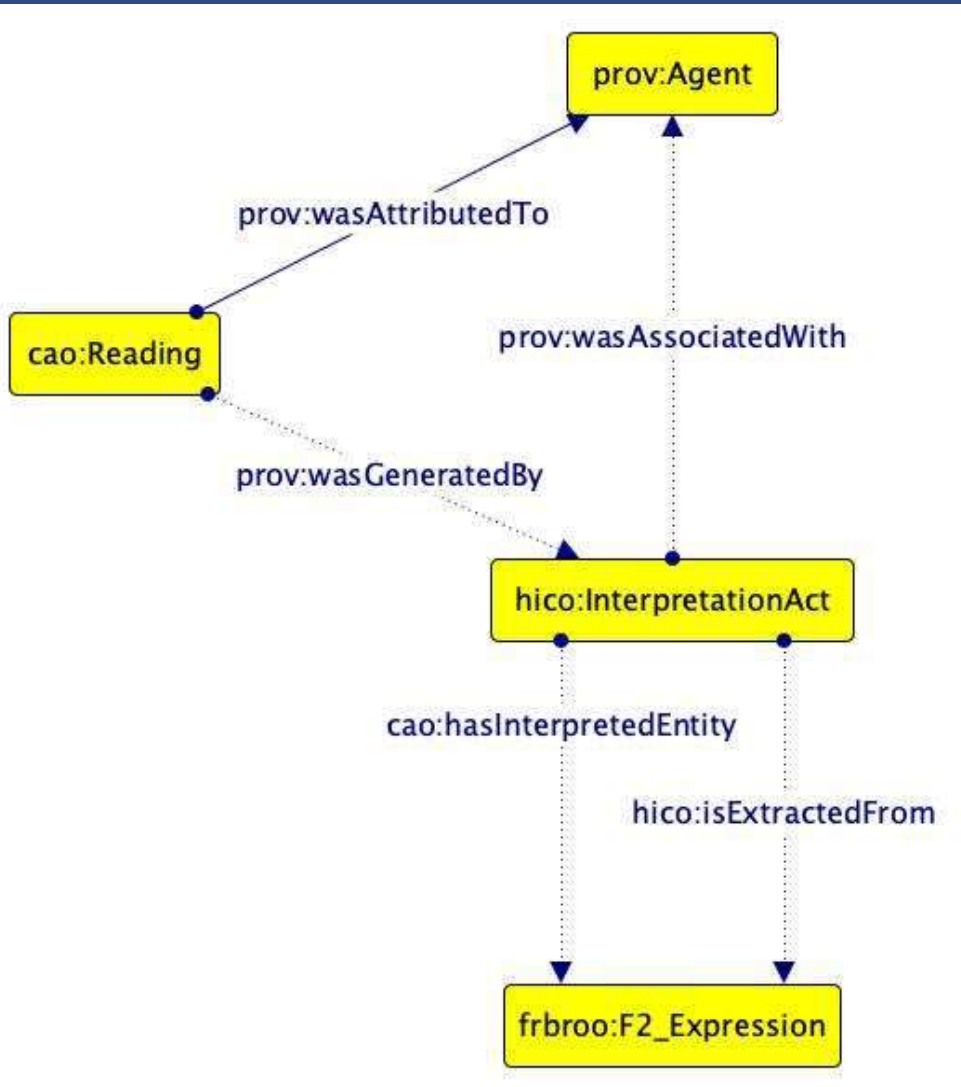
*From **https://marilenadaquino.github.io/hico/***

*See also PROV-O **https://www.w3.org/TR/prov-o/***

*And CiTO **https://sparontologies.github.io/cito/current/cito.html***

Representing the interpretation act which generated a reading

<u>hico:InterpretationAct</u> --> a process of scholarly interpretation, related to the responsible **agent** (**prov:Agent**) through the property **prov:was AssociatedWith**.

<u>cao:Reading</u> as result of an interpretation act. An interpretation act and a reading are associated through the property **prov:wasGeneratedBy**
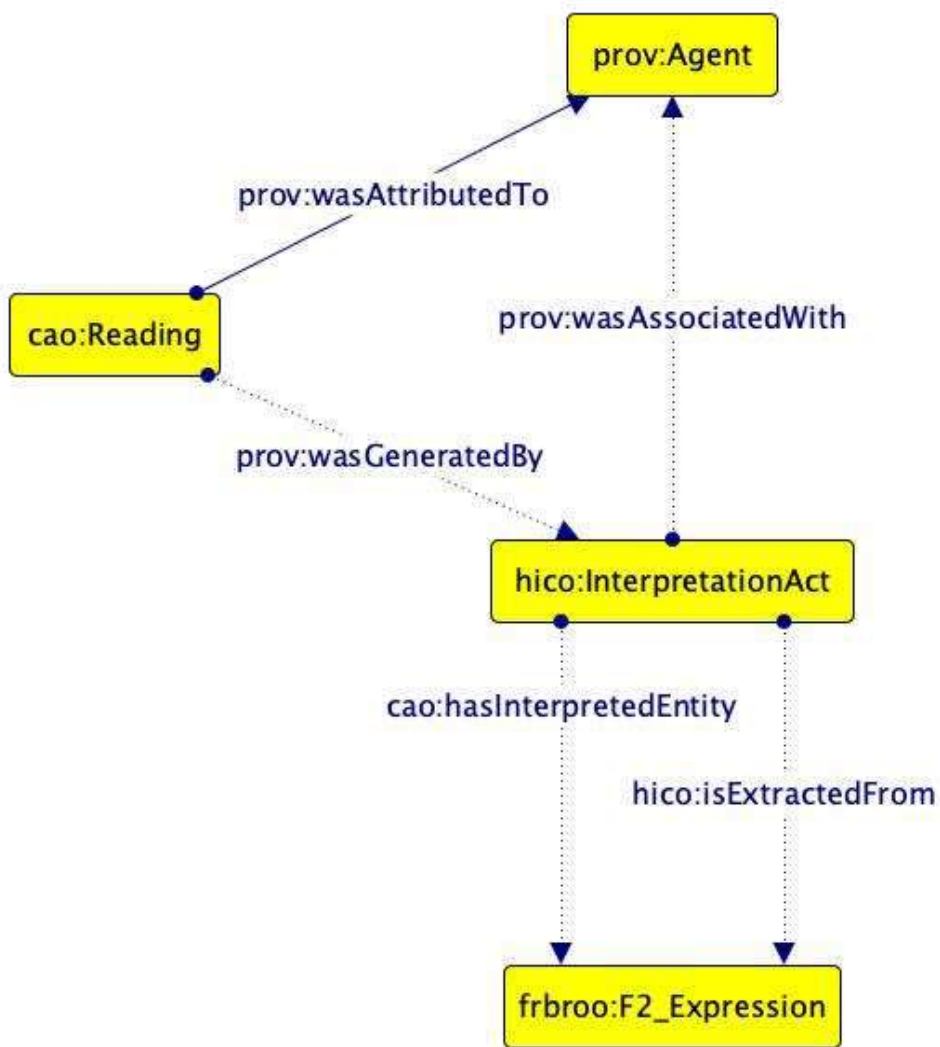
hico:InterpretationAct	prov:wasAssociatedWith	prov:Agent
cao:Reading	prov:wasGeneratedBy	hico:InterpretationAct

Representing the interpretation act which generated a reading

It's desirable to connect **cao:Reading** with the **prov:Agent** through the property **prov:AttributedTo**
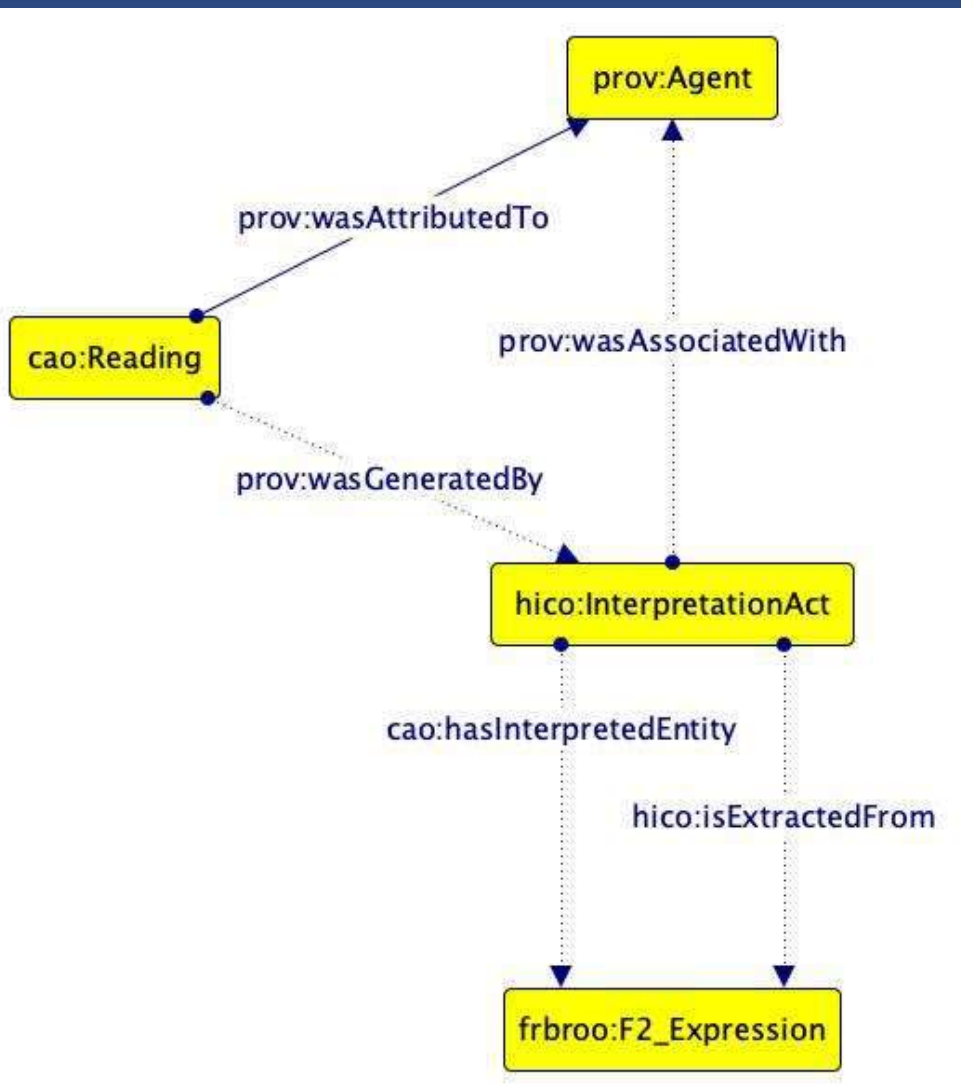
To sum up:

hico:InterpretationAct	prov:wasAssociatedWith	prov:Agent
cao:Reading	prov:wasGeneratedBy	hico:InterpretationAct
cao:Reading	prov:wasAttributedTo	prov:Agent

Reusing another scholar's reading

cao:Reading	prov:wasAttributedTo	prov:Agent
hico:InterpretationAct	hico:isExtractedFrom	frbroo:F2_Expression

Here Expression is a ***published edition***

Conjecturing a new reading

cao:Reading	prov:wasAttributedTo	prov:Agent
prov:Agent	cao:hasInterpretedEntity	frbroo:F2_Expression

Here Expression is a ***fragment of a witness***
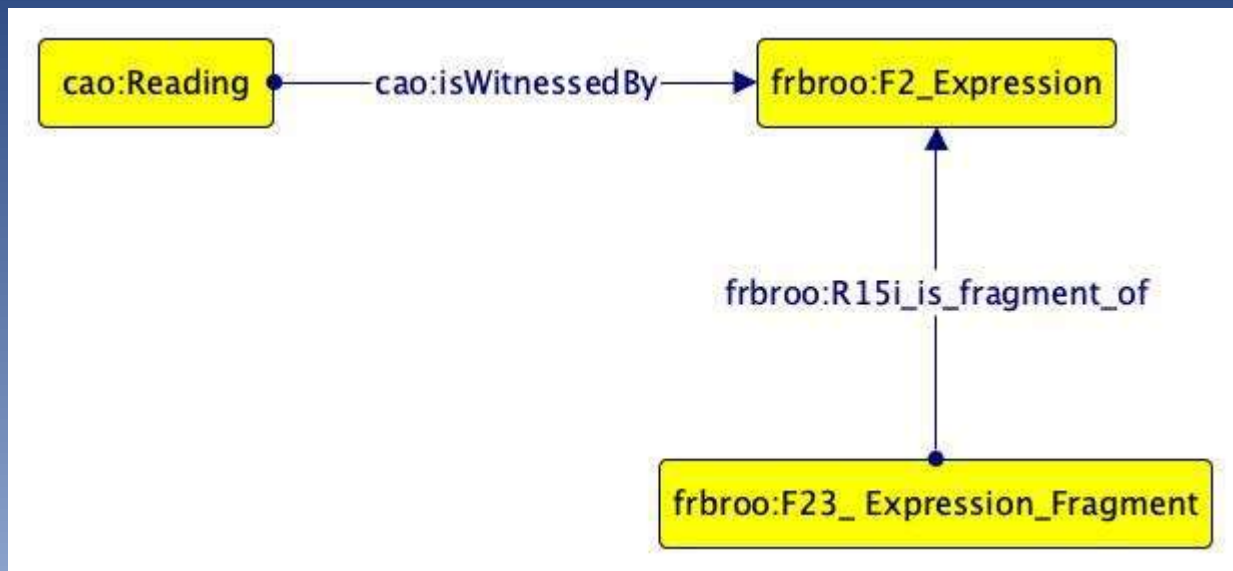
FRBR model:(https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr_2008.pdf)

WEMI

   Work

   Expression

   Manifestation

   Item
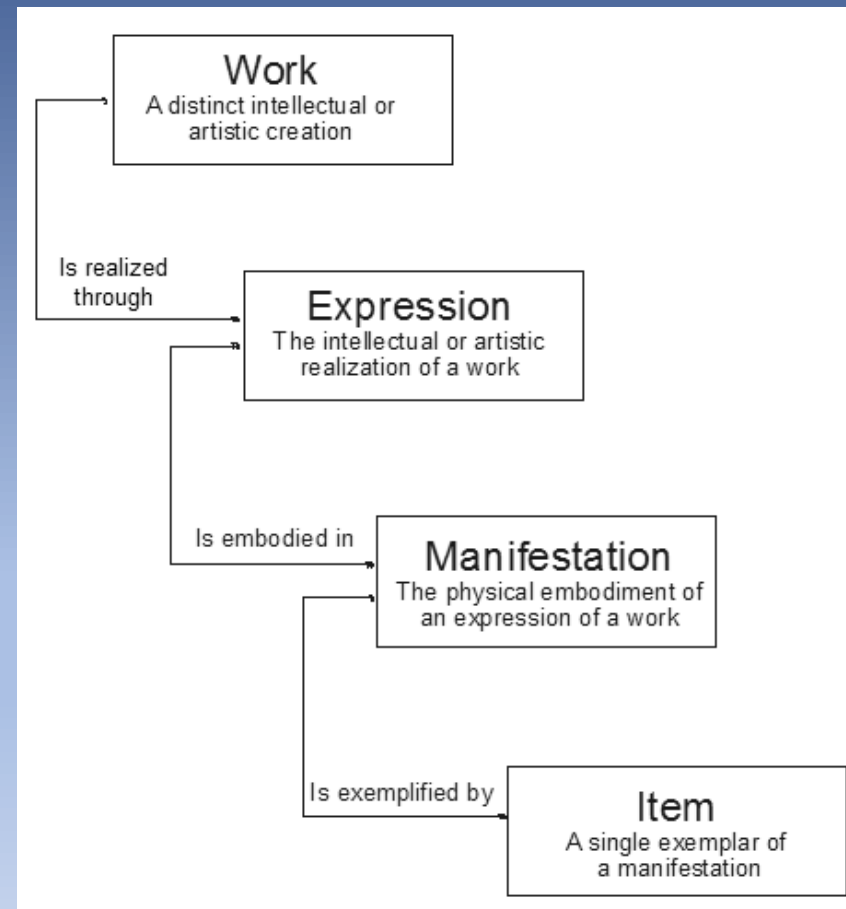
Texts reaches us in **multiple versions**.

Each textual version represents a **different realization** of the **same work.**

An **FRBR expression**, i.e. a text, is modelled using the **FRBRoo class F2_Expression**, which defines "*the specific intellectual or artistic form that a work takes each time it is realized*" ([IFLA 2009](#)).

```
<rdg/anno01-rdg01> a cao:BaseReading ;
    rdf:value "diffundi"^^xsd:string ;
    cao:isWitnessedBy <wit-frag/anno01-wfrag01> ,
        <wit-frag/anno01-wfrag02> .

<rdg/anno01-rdg02> a cao:Reading ;
    rdf:value "diffudit"^^xsd:string ;
    cao:isWitnessedBy <wit-frag/anno01-wfrag03> .

<rdg/anno01-rdg03> a cao:Reading ;
    rdf:value "diffundit"^^xsd:string ;
    cao:isWitnessedBy <wit-frag/anno01-wfrag04> .

<wit-frag/anno01-wfrag01> a frbroo:F23_Expression_Fragment ;
    frbroo:R15i_is_fragment_of <wit-expr/wexp01> .

<wit-frag/anno01-wfrag02> a frbroo:F23_Expression_Fragment ;
    frbroo:R15i_is_fragment_of <wit-expr/wexp02> .
```
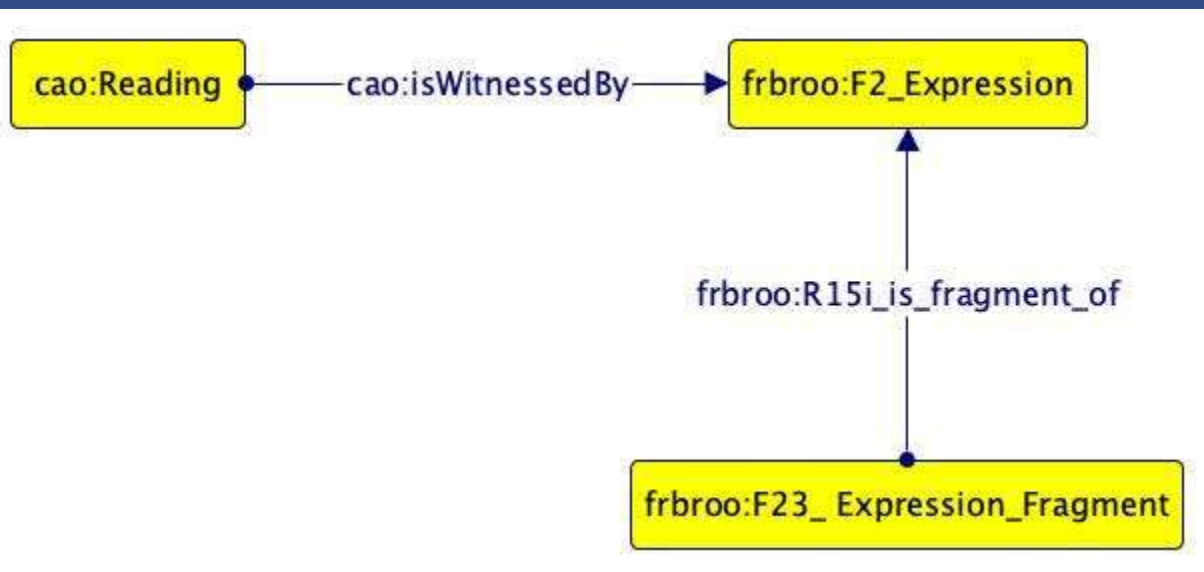
A reading is a scholarly claim, the outcome of an act of interpretation, which is supported by a particular **fragment of a witness**, a **frbroo_F23_Expression_Fragment**.

The CAO property **cao:isWitnessedBy** relates a reading to an expression, i.e. to the text of the witness (either the <u>full text</u> or the <u>specific textual fragment</u> that supports the reading within a witness).

The property **frbroo:R15i_is_fragment of** connects an <u>expression fragment</u> to the <u>expression</u> to which it belongs.

Expressions cannot exist without a __carrier.__

A __carrier__ is, in FRBR terms, a __*manifestation*__.

FRBRoo distinguishes between two classes of manifestation:

- *Manifestation Singleton*: belongs to the realm of physical things: it is a unique, physical object such as a manuscript, and cannot be replicated.

- *Manifestation Product Type*: an abstract notion, it comprehends publications and printed editions, which can exist in multiple physical copies.

Domain	Property	Range
frbroo:F2_Expression	crm:P128i_is_carried_by	frbroo:F3_Manifestation_Product_Type
frbroo:F2_Expression	frbroo:R4_carriers_provided_by	frbroo:F4_Manifestation_Singleton

2.6. the <u>body</u> of the annotation: from the reading to the witness

To sum up:

A reading (**cao:Reading**) is a scholarly interpretation of a witness fragment (**frbroo:F23_Expression_Fragment**),

that belongs to a witness (**frbroo:F2_Expression**).
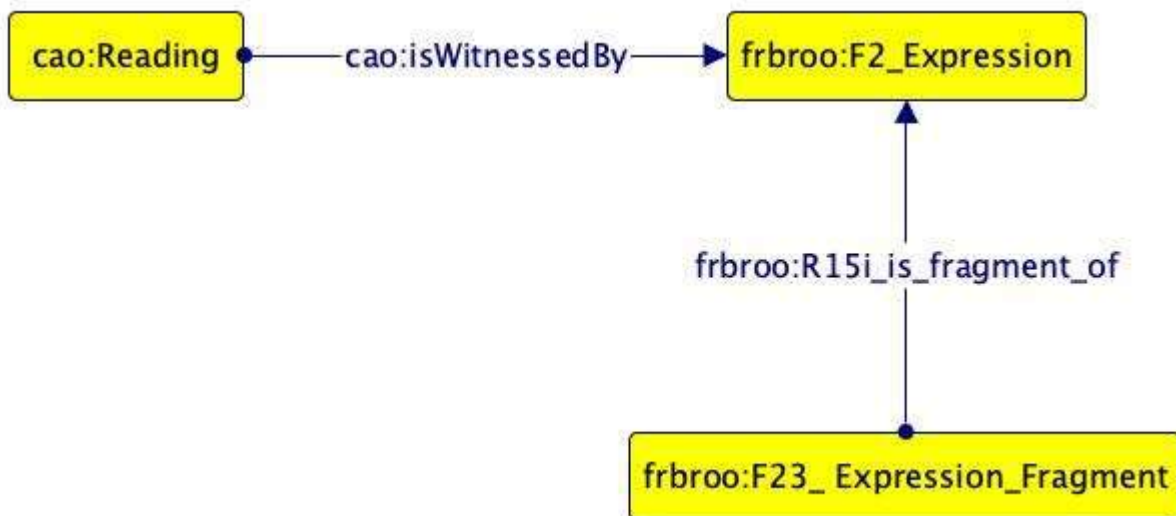
frbroo:F2_Expression is carried by

a physical document (**frbroo:F4_Manifestation_Singleton**)

or an edition (**frbroo:F3_Manifestation_Product_Type**).
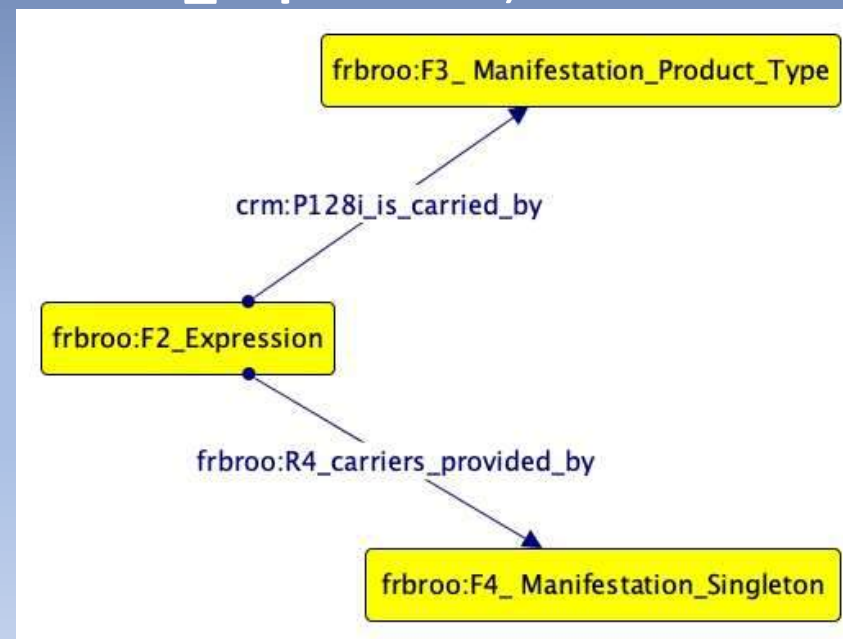
2.7. the body of the annotation: philological notes

```
<rdg/a6-r1> a cao:BaseReading ;
    crm:P3_has_note <note/a6-n1> ;
    rdf:value "['Beowulfe']"^^xsd:string .

<note/a6-n1> a crm:E62_String ;
    rdf:value  "['[...]  Folio  179,  with  the  last  page  (Fol.  198b)
[...]']"^^xsd:string ;
    crm:P2_has_type <https://www.example.org/type/philological> .
```

The property **crm:P3_has_note** links a reading or apparatus entry to a scholarly note.

The range of the property **crm:P3_has_note** is an **rdfs:Literal**.

It is possible to use **crm:P2_has_type** on the note to further characterize the type of annotation (a possible value is, e.g., 'philological').

(https://cidoc-crm.org/html/cidoc_crm_v7.1.1.html)

2.8. A special case of variation: transposition

Transposition occurs when text is transferred from one location to another.

As transposition involves a relationship between a transposed text and (at least) two distinct locations, its representation requires to **link** the annotation to **multiple targets.**
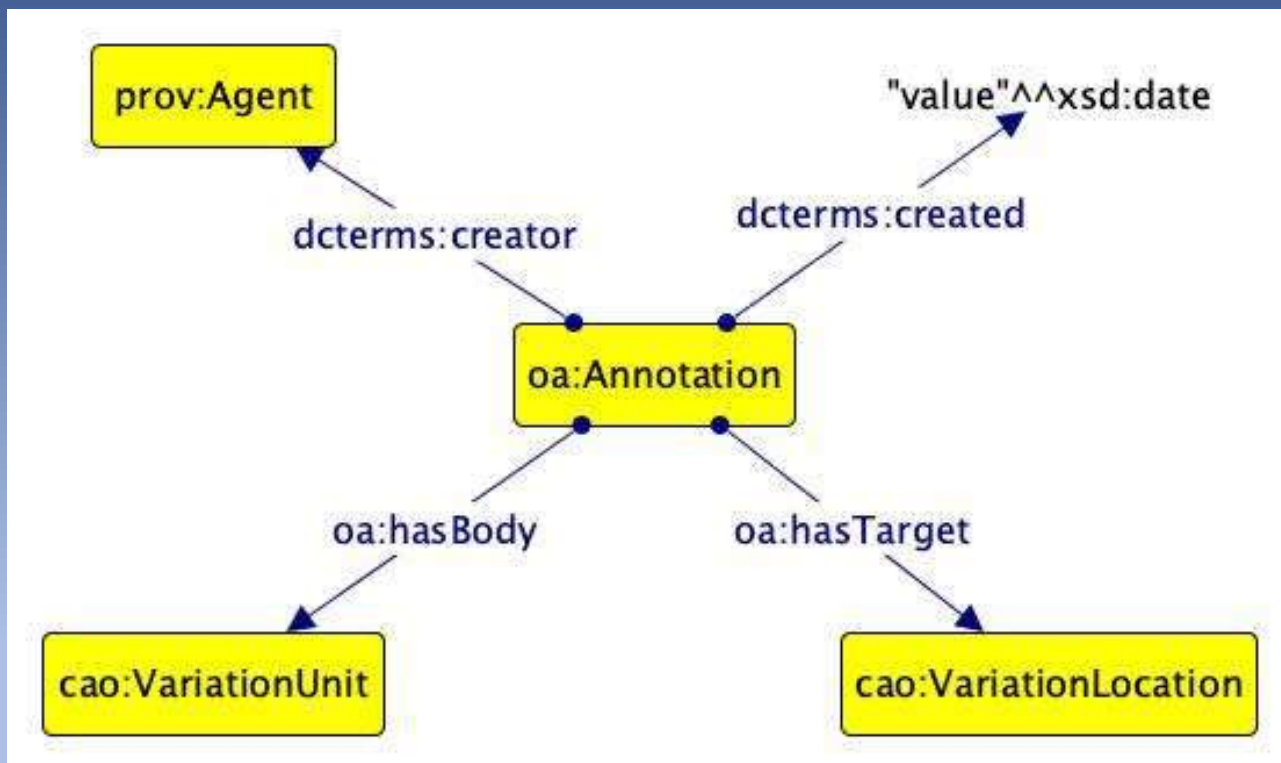
The transposed text is then related to the original and new location via the properties **cao:wasTransposedFrom** and **cao:wasTransposedTo.**

```
<annotation/anno02> a oa:Annotation ;
    oa:hasBody <app/app02> ;
    oa:hasTarget <varloc/vl02> , <varloc/vl03> ;
    dcterms:creator <person/fgiovan> ;
    dcterms:created "2019-03-09"^^xsd:date .
```

```
<rdg/anno02-rdg02> a cao: Reading ;
    cao:hasReadingType cao:transposition ;
    cao:wasTransposedFrom <varloc/vl02> ;
    cao:wasTransposedTo <varloc/vl03> .
```

- **Web Annotation Ontology**
- (<u>https://www.w3.org/TR/annotation-vocab/</u>)
  - **Body**
  - **Target**
  - **Creator**
  - **Date of Creation**

"the locus of variation in the critical text"

Methods for linking the apparatus to the critical text:

1. *location-referenced* method

2. *double-end-point-attached* method

3. *parallel segmentation* method

Parallel segmentation method:

Not compatible with an external apparatus: it does not impose the existence of a base text and for this it allows only an *inline markup*

But it allows the specification of the boundaries of the **location** of the variant in detail, through the use of empty elements equipped with unique identifiers.

Location-referenced method:

It can be used even when the critical apparatus is **external**.

Giving the exact location of the variant is not a requirement:

the readers reconstruct the exact location autonomously, once they have identified the relevant line as they would do with a print edition.

Double-end-point-attached method:

It turns out to be the most suitable for linking the RDF critical apparatus to the text

It's used for linking the apparatus to a specific base text within the TEI document containing the edition, through the Web Annotation class

oa:RangeSelector,

defining the beginning and the end through the subclasses

oa:hasStartSelector

oa:hasEndSelector

using *XPath* expressions.

```
... freta rapidisque <anchor xml:id="varloc-s01"/>diffundi<anchor
xml:id="varloc-e01"/> ...

<varloc/vl01> a cao:VariationLocation ;
    oa:hasSource <https://example.org/example.xml> ;
    oa:hasSelector [
    a oa:RangeSelector ;
        oa:hasStartSelector [
        a oa:XPathSelector ;
        rdf:value "//l/anchor[@xml:id='varloc-s01']" ] ;
        oa:hasEndSelector [
            a oa:XPathSelector ;
            rdf:value "//l/anchor[@xml:id='varloc-e01']" ] ] .
```

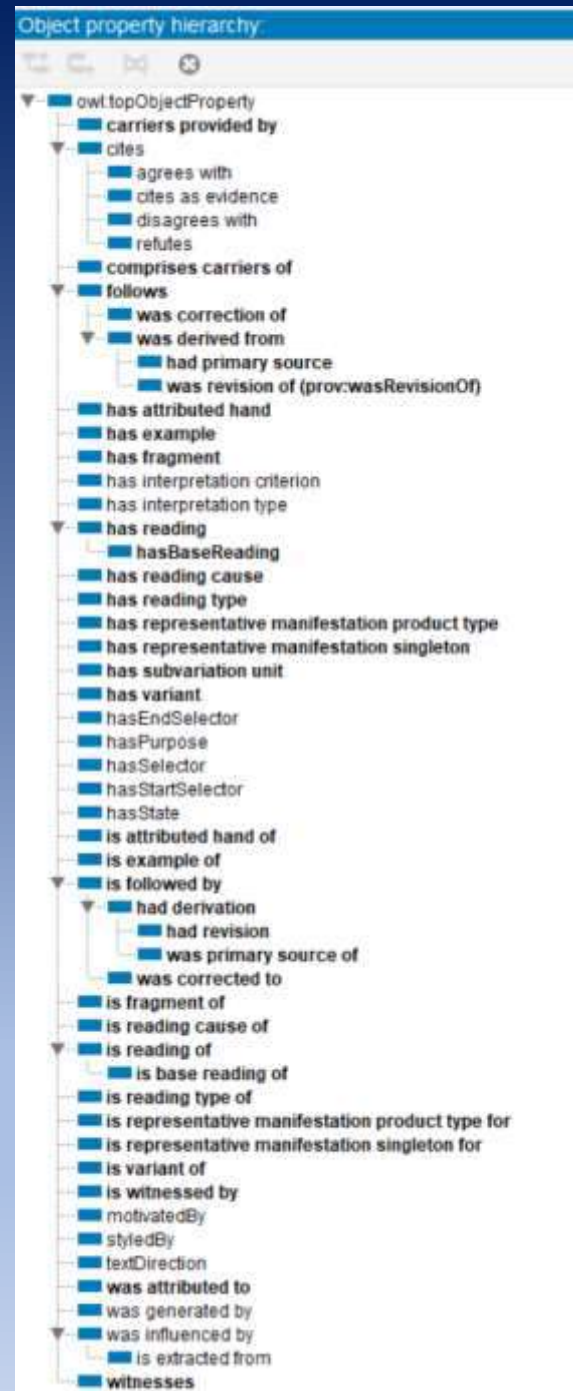# 3. CAO Entities – An Overview

Class Hierarchy
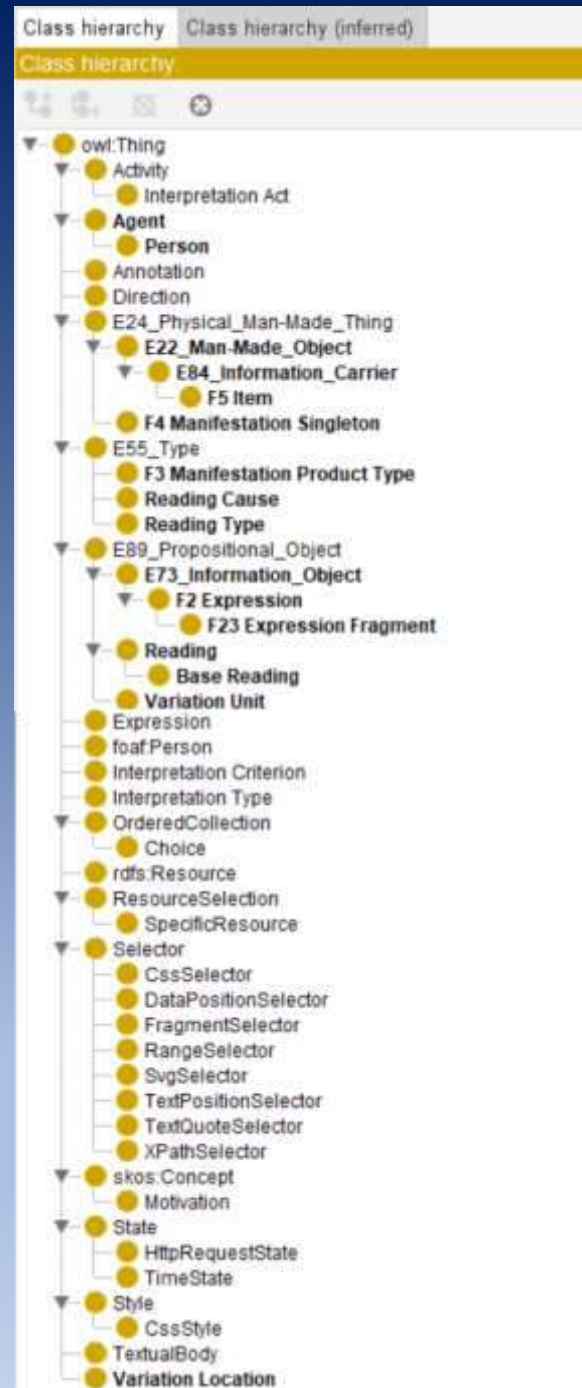
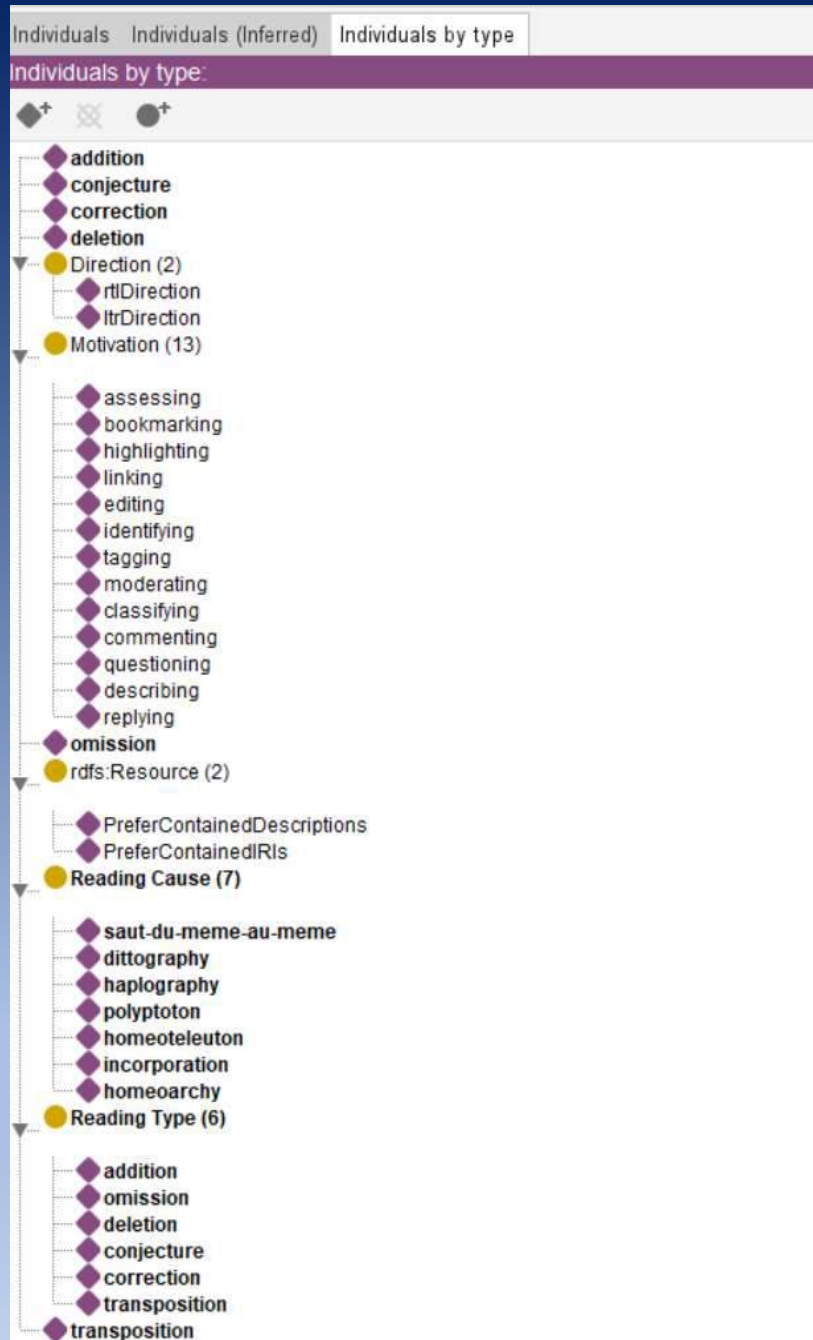# 3. CAO Entities – An Overview

Object Properties Hierarchy

3. CAO Entities – An Overview

Individuals

3. CAO Entities – An Overview

Data Types

Classes Object properties Data properties **Datatypes** Individuals

Datatypes:

- owl:rational
- owl:real
- rdf:langString
- rdf:PlainLiteral
- rdf:XMLLiteral
- rdfs:Literal
- xsd:anyURI
- xsd:base64Binary
- xsd:boolean
- xsd:byte
- xsd:dateTime
- xsd:dateTimeStamp
- xsd:decimal
- xsd:double
- xsd:float
- xsd:hexBinary
- xsd:int
- xsd:integer
- xsd:language
- xsd:long
- xsd:Name
- xsd:NCName
- xsd:negativeInteger
- xsd:NMTOKEN
- xsd:nonNegativeInteger
- xsd:nonPositiveInteger
- xsd:normalizedString
- xsd:positiveInteger
- xsd:short
- xsd:string
- xsd:token
- xsd:token
- xsd:unsignedByte
- xsd:unsignedInt
- xsd:unsignedLong
- xsd:unsignedShort

3. CAO Entities – An Overview

Data Properties
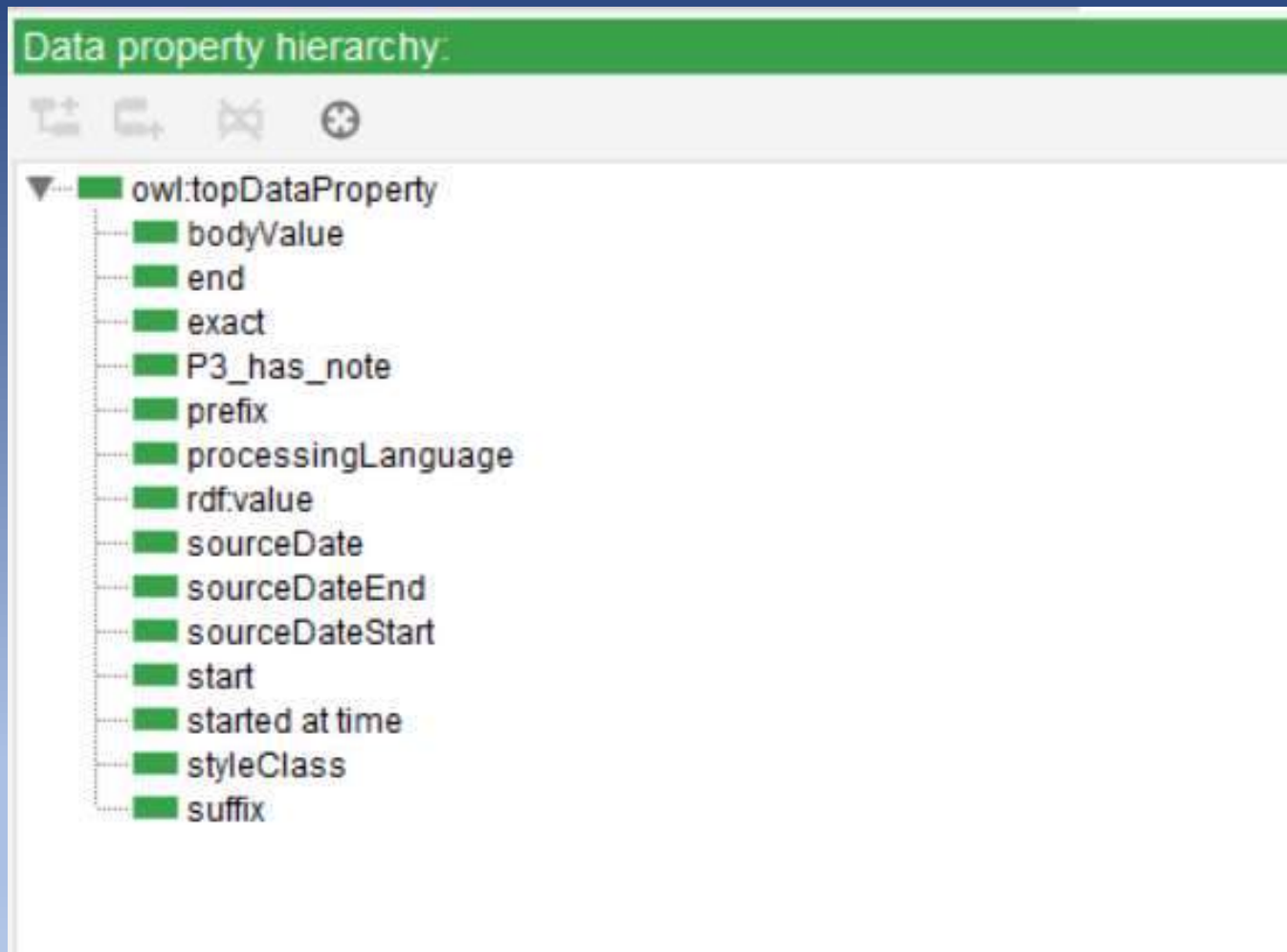
3. CAO Entities – An Overview

Annotation Properties

Annotation property hierarchy:

- cachedSource
- dc:contributor
- dc:creator
- dc:date
- dc:description
- dc:rights
- dc:title
- dcterms:creator
- dcterms:description
- dcterms:license
- dcterms:modified
- foaf:name
- hasBody
- hasScope
- hasSource
- hasTarget
- owl:backwardCompatibleWith
- owl:deprecated
- owl:incompatibleWith
- owl:priorVersion
- owl:versionInfo
- preferredNamespacePrefix
- preferredNamespaceUri
- rdfs:comment
- rdfs:isDefinedBy
- rdfs:label
- rdfs:seeAlso
- renderedVia
- **was revision of (prov:wasRevisionOf)**

OWLWiz Class-Tree
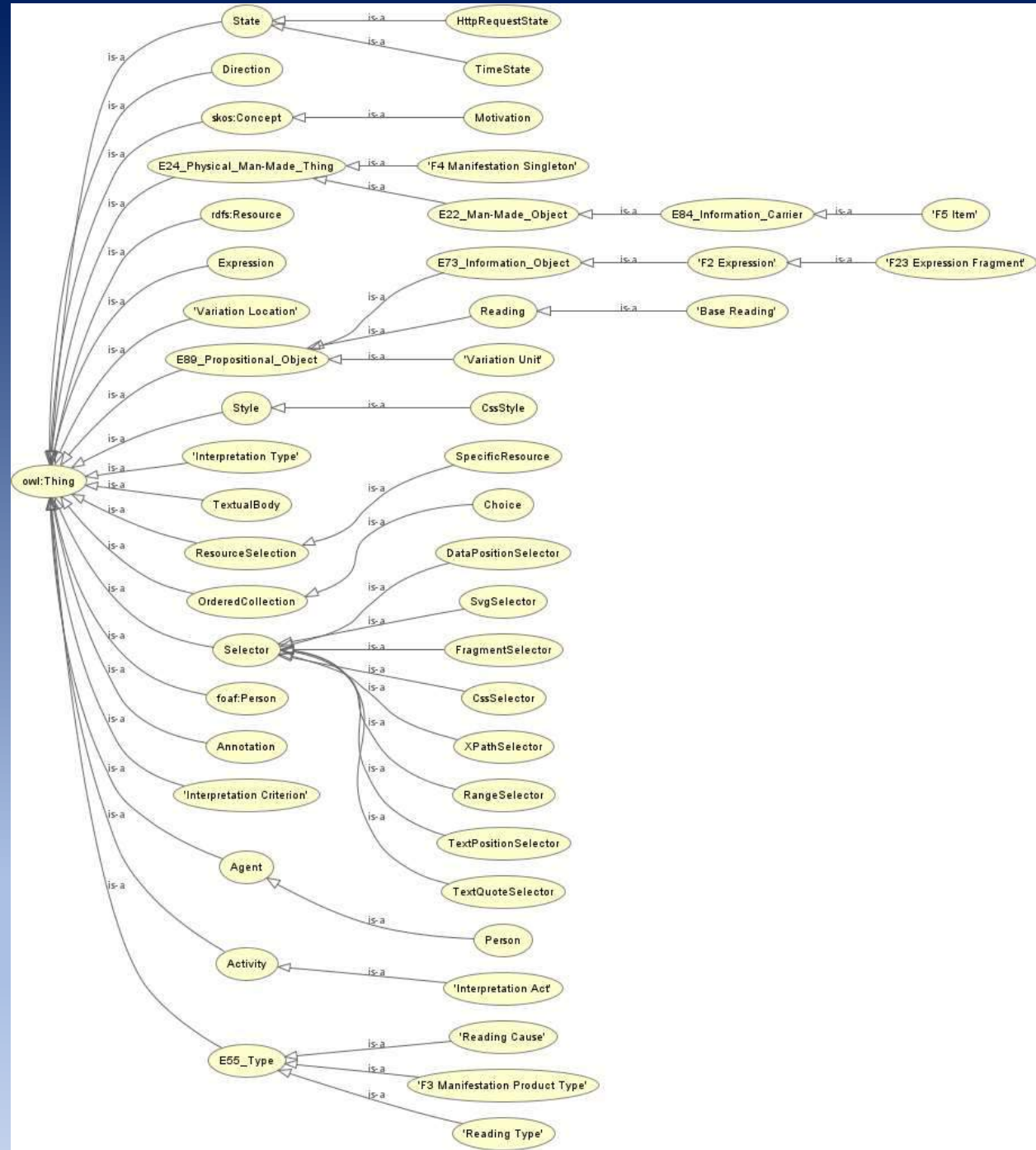
4. Conclusions

- The first ontology specifically created for representing scholarly digital editions' critical apparatus.

- A project which is part of a bigger one: **_LIFT_** *(Linked Data From TEI)* https://projects.dharc.unibo.it/lift/ .

- Space for further researches.

4. Conclusions

- What to improve?
  - Version 0.9.
  - Correct some transcription errors.
  - Improve the representation of *transposition.*
  - Create a specific SPARQL endpoint.
  - Try to implement scholarly digital editions in order to test the CAO.

5. Bibliography

- **Giovannetti F. 2015**

Combining TEI and Semantic Web Technologies to Annotate and Explore the Notebook of Paolo Bufalini (1915-2001).

- **Giovannetti F. 2019**

The Critical Apparatus as a Semantic Graph. A Combined Approach to Describing Textual Variation by TEI, RDF, and Ontologies

6. Web Pages

- CAO: https://w3id.org/cao
- PROV-O: https://www.w3.org/TR/2013/REC-prov-o-20130430/
- HiCO: https://marilenadaquino.github.io/hico/
- FRBR: https://fgiovannetti.github.io/cao/#http://iflastandards.info/ns/fr/frbr/frbroo/R4
- OA: https://www.w3.org/TR/annotation-vocab/#classes

7. Credits and contacts

- Francesca Giovannetti: francesc.giovannett6@unibo.it
- Diego Canali: diego.canali@studio.unibo.it

Thank you for the attention

Diego Canali for TALOS – AI4SSH
ERA Chair in Artificial Intelligence for Humanities and Social Sciences
Rethymno, 6 November 2023