Ontogenia

Miguel Ceriani , Anna Sofia Lippolis, Andrea Giovanni Nuzzolese, Sara Zuppiroli

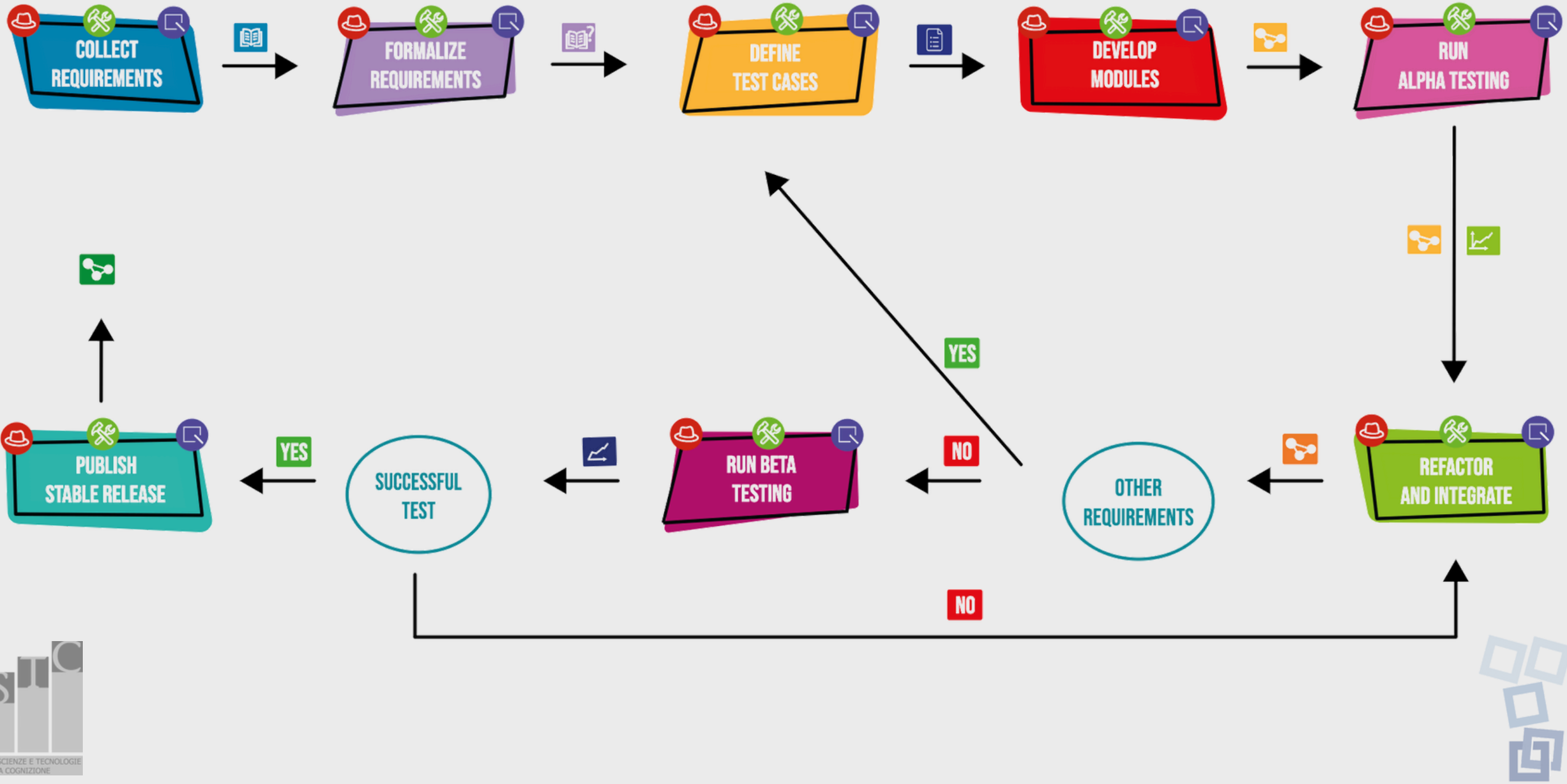
Semantic Technology Laboratory

CNR -Institute of Cognitive Sciences and Technologies

Bologna , Italy

eXtreme Design

XD
EXTREME DESIGN



Context

Knowledge engineering

"[...] the collection of activities for eliciting, capturing, conceptualising and formalising knowledge for the purpose of being used in information systems"1

1. M. Sabou, et al., Knowledge Engineering in the Age of Neurosymbolic Systems. Submitted to Neurosymbolic Artificial Intelligence. Pre-print

Fortunately

Illustration from “Fortunately” by Remy Charlip

We have established KE methodologies

Unfortunately

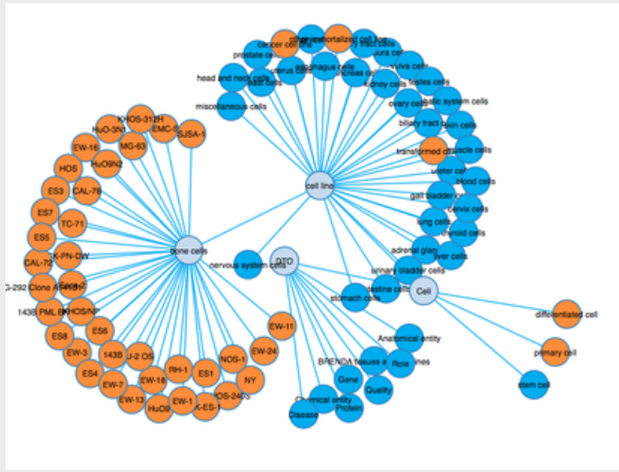
XD requires deep understanding of the domain

Time-consuming activity that requires a lot of cognitive effort by *knowledge engineers*

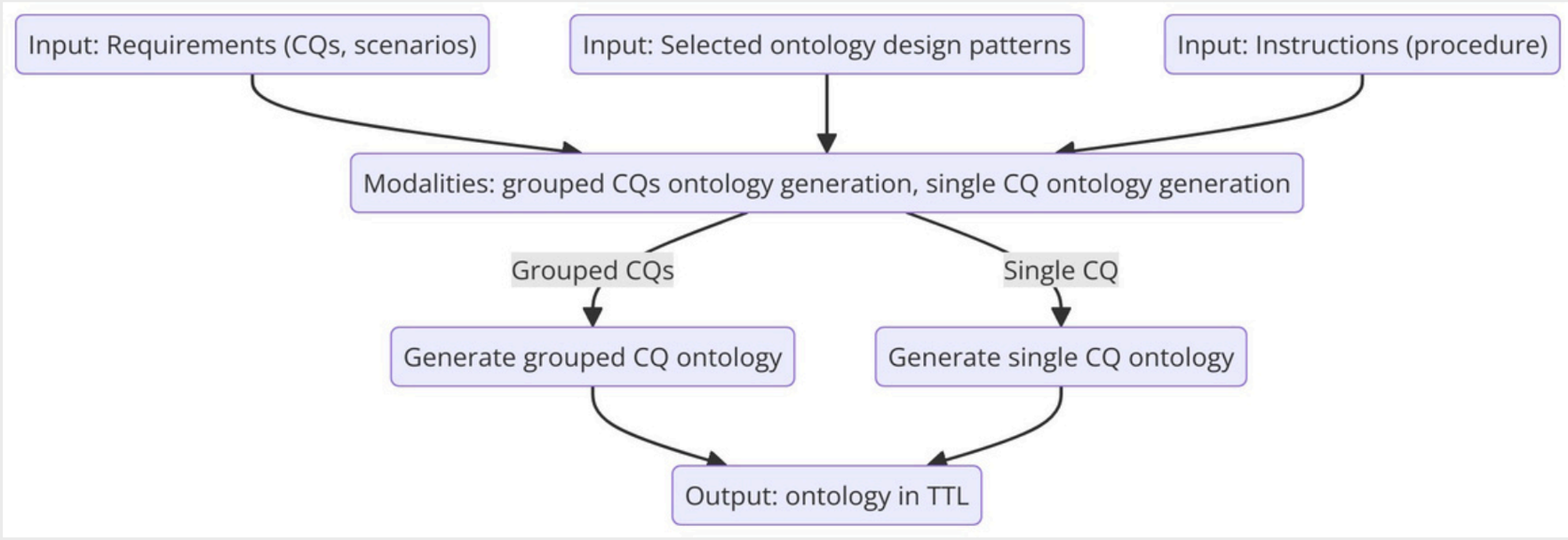
Illustration by Christoph Neimann

Proposed Solution

Ontogenia: Humans + LLMs to generate semantic knowledge graphs



Ontogenia steps



Input: Requirements(CQs, scenarios), selected ontology design patterns, instructions (procedure)

Modalities: CQs to be merged together or to be inputted one at a time

Output: ontology in Turtle format

Example of requirement

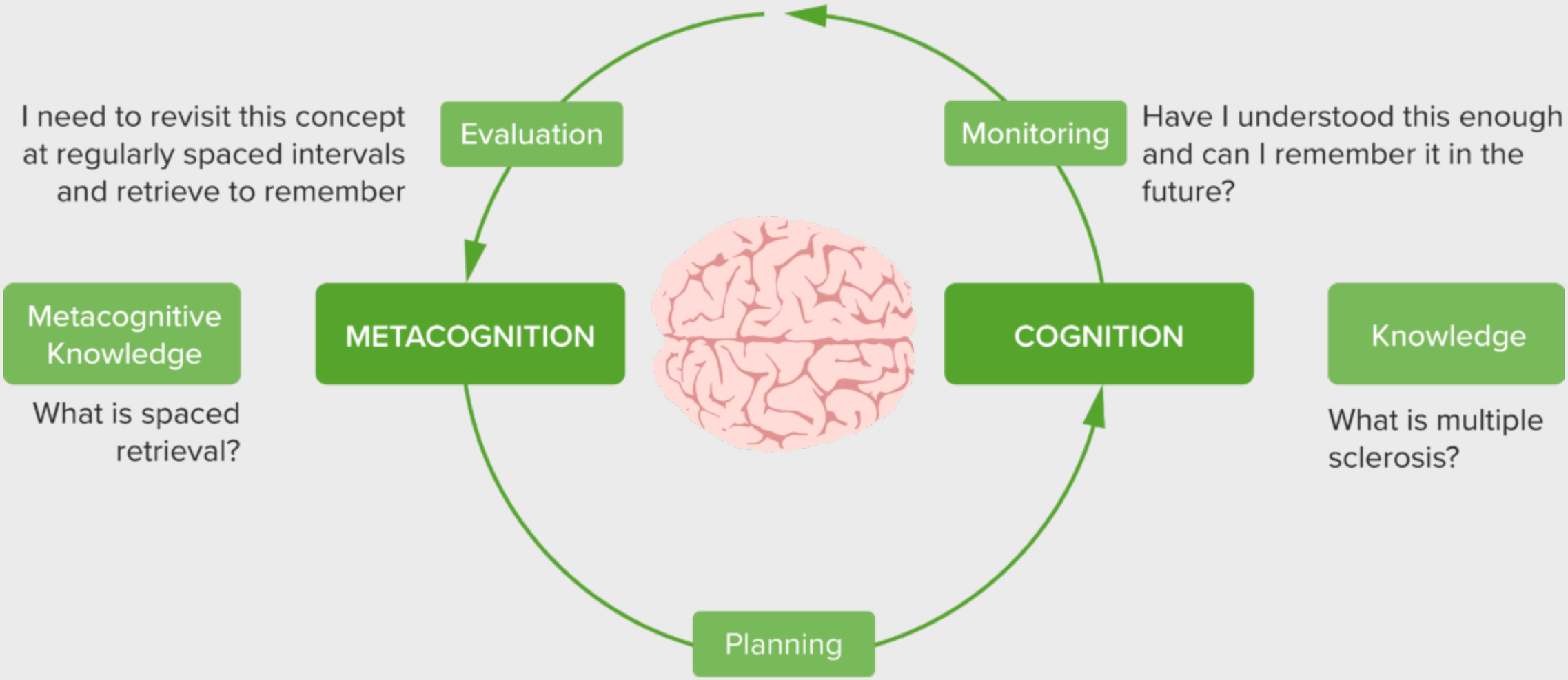
Competency question: When was a certain album released?

Scenario:

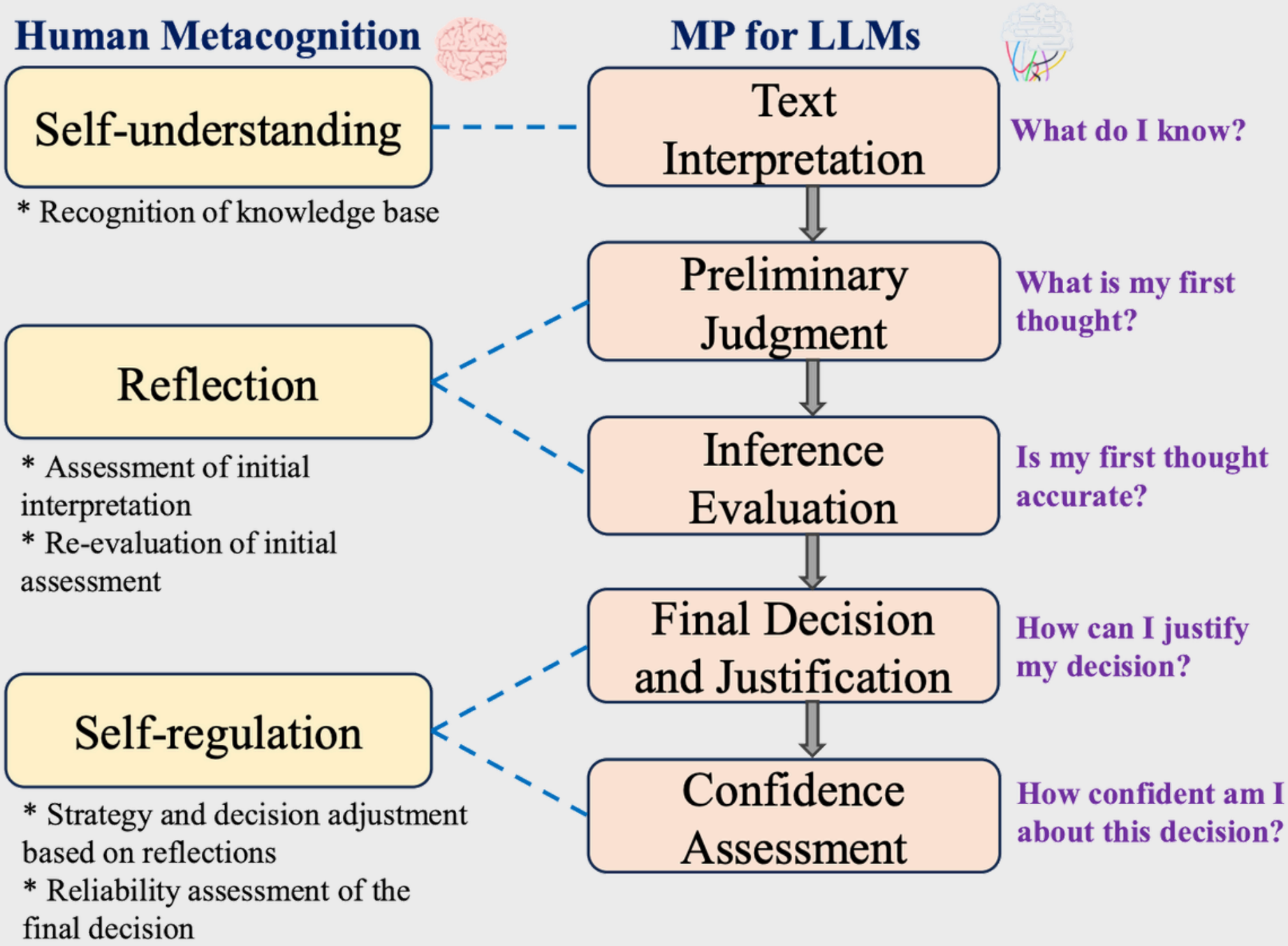
The current configuration of the 'Red Hot Chili Peppers' are: Anthony Kiedis(vocals), Flea(bass, trumpet, keyboards, and vocals), John Frusciante (guitar), and Chad Smith (drums).The line-up has changed a few times during the years, Frusciante replaced Hillel Slovakin 1988, and when Jack Irons left the band he was briefly replaced by D.H. Peligo until the band found Chad Smith. In addition to playing guitars for Red hot Chili Peppers Frusciante also contributed to the band “The Mars Volta” as a vocalist for some time.

From September 2004, the Red Hot Chili Peppers started recording the album 'Stadium Arcadium'. The album contains 28 tracks and was released on May 5 2006. It includes a track of the song 'Hump de Bump', which was composed in January 26, 2004.

Human metacognition



Metacognitive prompting



1.Y. Wang , Y. Zhao., (2023) Metacognitive Prompting Improves Understanding in Large Language Models. arXiv preprint arXiv:2308.05342

Ontogenia steps

Comprehension Clarirification

Competency question understanding

Understanding the question: "Is [this animal] a herbivore?" We need to determine if a specific animal falls under the category of herbivores, which are animals that primarily eat plants.

Preliminary judgement

Preliminary identification of the context. Divide the CQ in subject, preducate, object and predicate nominative

Context: The domain involves zoology, specifically animal diets. We need classes to represent animals and their dietary habits.

Critical evaluation

Starting from your knowledge, extend the ontology with these restrictions

```
<owl:Class rdf:about="&zoo;Animal"/>

<owl:Class rdf:about="&zoo;Herbivore">
  <rdfs:subClassOf rdf:resource="&zoo;Animal"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="&zoo;hasDietType"/>
      <owl:hasValue rdf:datatype="&xsd;string">plant</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="&zoo;Carnivore">
  <rdfs:subClassOf rdf:resource="&zoo;Animal"/>
  <owl:disjointWith rdf:resource="&zoo;Herbivore"/>
  <owl:disjointWith rdf:resource="&zoo;Omnivore"/>
</owl:Class>

<owl:Class rdf:about="&zoo;Omnivore">
  <rdfs:subClassOf rdf:resource="&zoo;Animal"/>
  <owl:disjointWith rdf:resource="&zoo;Herbivore"/>
  <owl:disjointWith rdf:resource="&zoo;Carnivore"/>
</owl:Class>

<owl:ObjectProperty rdf:about="&zoo;hasDiet"/>
<owl:DatatypeProperty rdf:about="&zoo;hasDietType"/>
```

- Subject: [this animal] (individual instance of the class Animal)
- Predicate: is (relates the subject to its dietary habit)
- Object: herbivore (instance of the class Herbivore)

n terms of ontology:

- Class: `Animal`
- Class: `Herbivore` (subclass of Animal)
- Object property: `hasDiet` (domain: `Animal`, range: `Diet`)
- Datatype property: `isHerbivore` (referring to the dietary habit

Decision confirmation

Confirm the final answer and explain the reasoning

The ontology correctly models the relationships and classifications needed to determine if an animal is a herbivore. It uses the `hasDietType` property to differentiate between herbivores, carnivores, and omnivores, ensuring the correct classification of any given animal based on its diet.

To evaluate, we can instantiate an individual animal and test its classification:

Confidence assessment

Make a confidence evaluation and explanation using the ontology reasoning and test it on some instances

xml — Copia codice

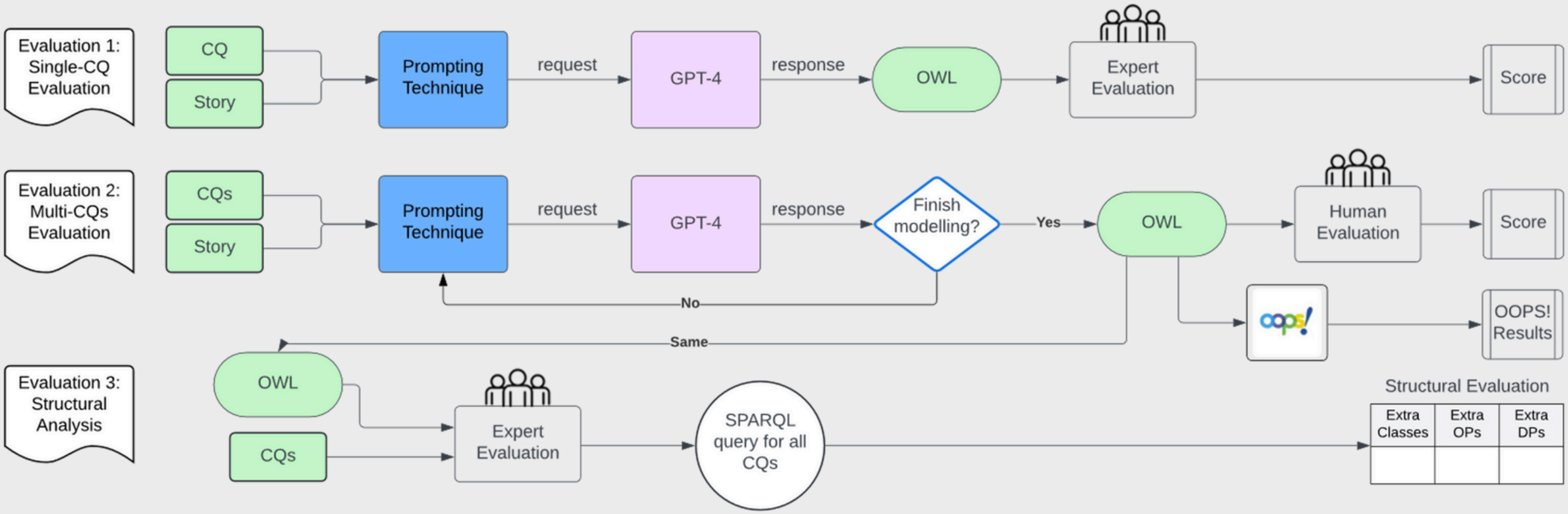
```
<owl:NamedIndividual rdf:about="&zoo;Elephant">
  <rdf:type rdf:resource="&zoo;Herbivore"/>
  <hasDietType rdf:datatype="&xsd;string">plant</hasDietType>
</owl:NamedIndividual>
```

Using a reasoner, we can check that `Elephant` is correctly inferred to be an instance of `Herbivore`.

Metacognitive stages

Ontogenia stages

Benchmark testing and evaluation

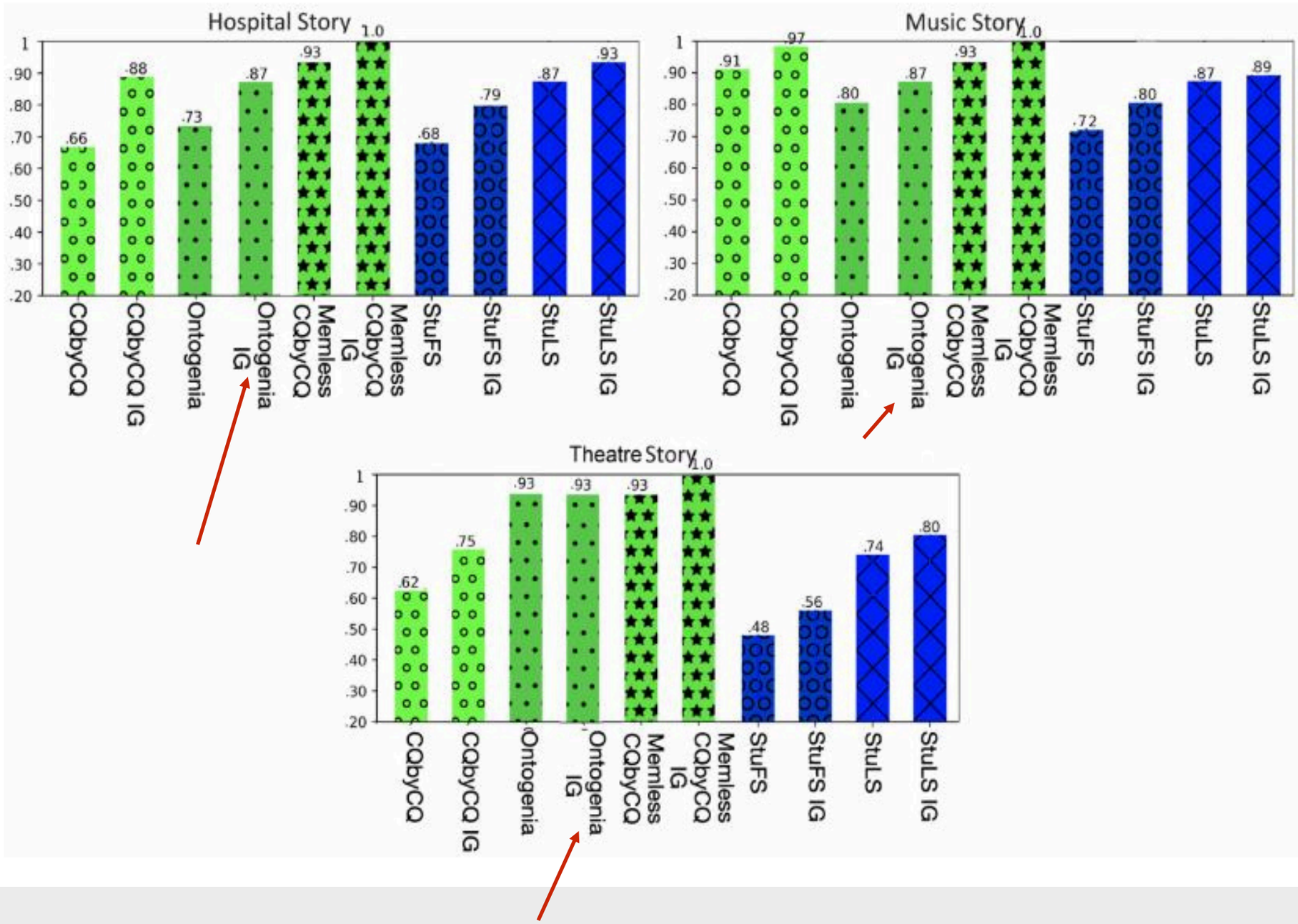


Results: Structural evaluation

Ontometrics	Case1 No pattern No MP	Case2 Pattern No MP	Case3 No pattern MP	Case4 Pattern MP	Reference Ontology
Axioms count	49	119	64	118	108
Logical axioms count	26	74	36	76	56
Class count	14	17	14	21	31
Object property count	8	11	8	14	5
Data property count	0	2	3	2	0
Properties count	8	13	11	16	5
Individual count	1	19	0	11	0
DL expressivity	ALCROI	AL(D)	ALC(D)	ALCI(D)	SRI

- **Classes and properties successfully identified**
- **Richer formalisation with MP**
- **Larger set of terms and axioms when ODP used**
- **Low cost and time**

Results: Comparison with state of the art systems and humans

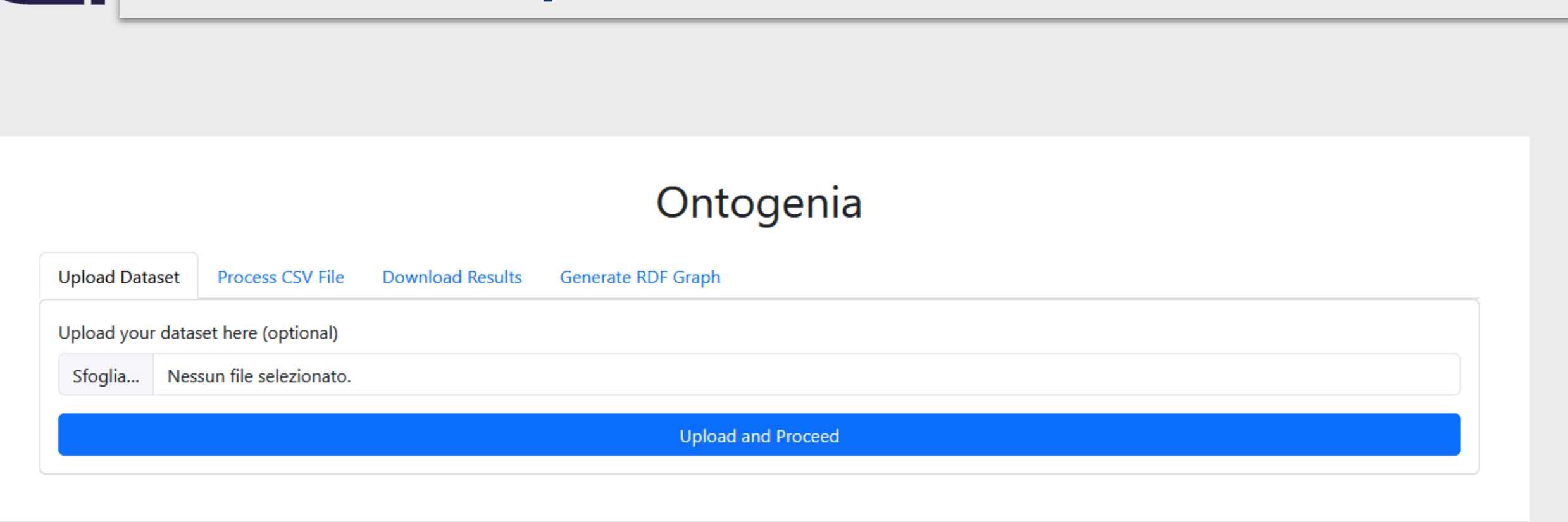


- IG indicates the overseeing of minor errors.
- Humans were students of a knowledge engineering course

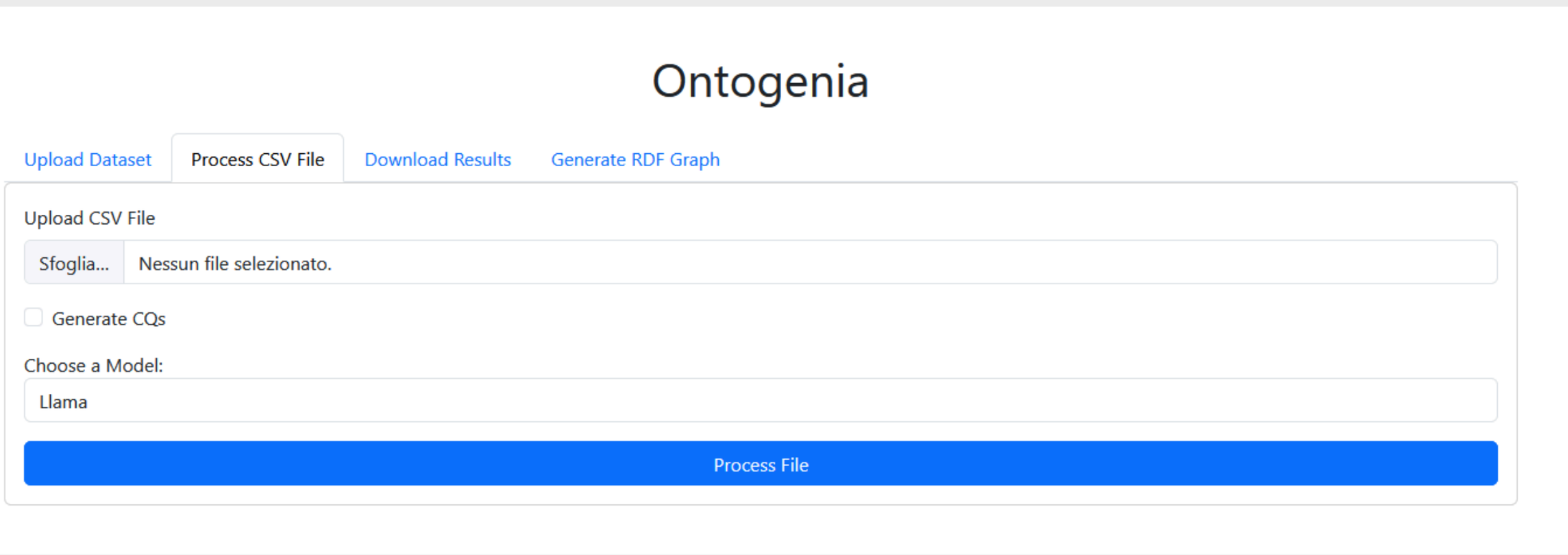
Work in progress

- Possibility to choose open or closed-source LLM
- Management of more input formats (dataset)
- CQs generation
- Generation of declarative mapping (RML, SPARQL Anything) to populate a KG
- Testing the same task with other LLMs to verify reproducibility
- Representing prompts as KG for FAIRness

1. Upload dataset

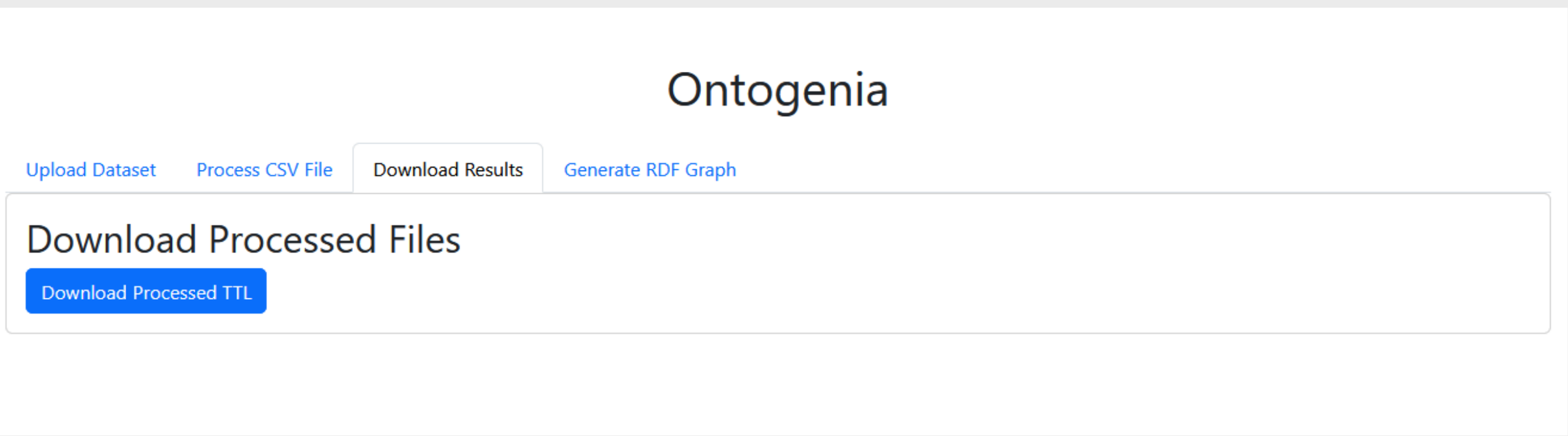


2. Upload or generate requirements



Requirements in the form of CQs and user stories can be uploaded in a CSV file or generated automatically

3. Download ontology



The ontology is processed in the background and can then be downloaded in Turtle format

In the background

```
[1 rows x 3 columns]
Designing ontology for CQID FestSCQ1: When did a certain theatre festival take place?
127.0.0.1 - - [25/Sep/2024 17:39:44] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [25/Sep/2024 17:39:44] "GET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [25/Sep/2024 17:39:46] "GET / HTTP/1.1" 200 -
```turtle
@prefix : <http://example.org/ontology/theatre#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix time: <http://www.w3.org/2006/time#> .

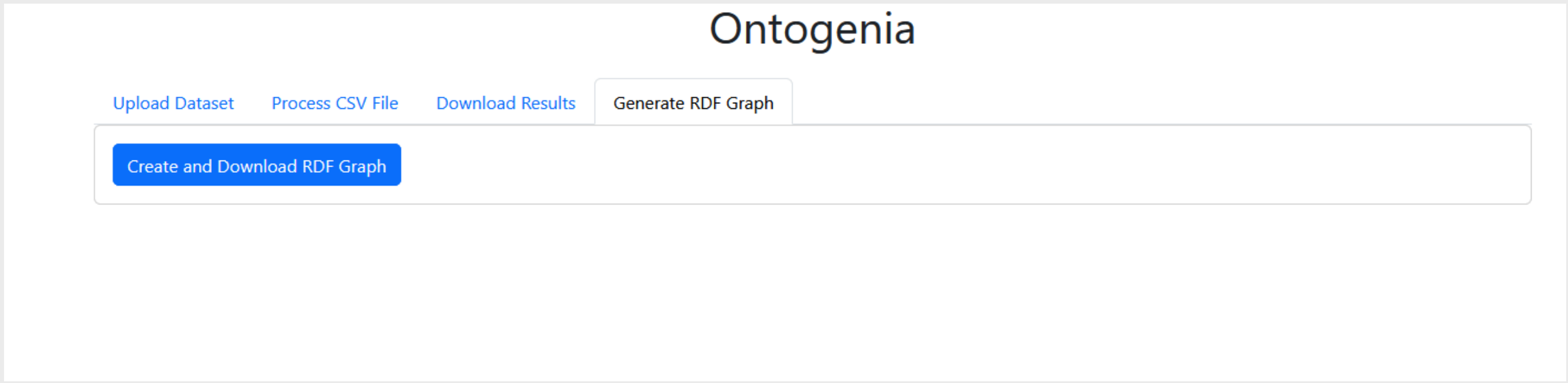
# Classes

:Performance a owl:Class ;
    rdfs:label "Performance" ;
    rdfs:comment "A specific instance of a play being performed." .

:TheatreFestival a owl:Class ;
    rdfs:label "Theatre Festival" ;
    rdfs:comment "A festival event that includes multiple theatre performances." .

:City a owl:Class ;
    rdfs:label "City" ;
    rdfs:comment "A city where theatre festivals and performances can take place." .
```
```

4. Create and download RDF graph



Basing on the ontology and the dataset, RML rules are defined to generate a RDF Graph sample that can be downloaded