Identifying Knowledge Patterns and Developing an Ontology

João Diogo Martins Romão, Marius Hauenstein, Niccolò Revel Garrone, Ricardo Carvalho

Part I

Introduction

The process of identifying knowledge patterns and developing an ontology involves analyzing a domain of knowledge and identifying the key concepts and relationship within that domain. This involves creating a structure for representing that knowledge in a formal way, using a system such as OWL to define the concepts and relationships within the domain.
The resulting ontology can be used to represent and organize knowledge in a way that makes it more accessible and easier to understand.

For the first part of this project, the objective is to identify four situations that can be modeled using an N-ary relation pattern within the domain of the Camino de Santiago, also known as the way of St James. This involves analyzing Wikipedia pages about the Camino de Santiago and identifying situations that involve multiple entities and can be represented using an N-ary relation pattern. For each situation identified, we will provide a general and a specific description, the Wikipedia page where it was found, a graphical representation of the situation, and the OWL code for the pattern. In addition we will classify each of these new patterns within the N-ary relation pattern taxonomy.

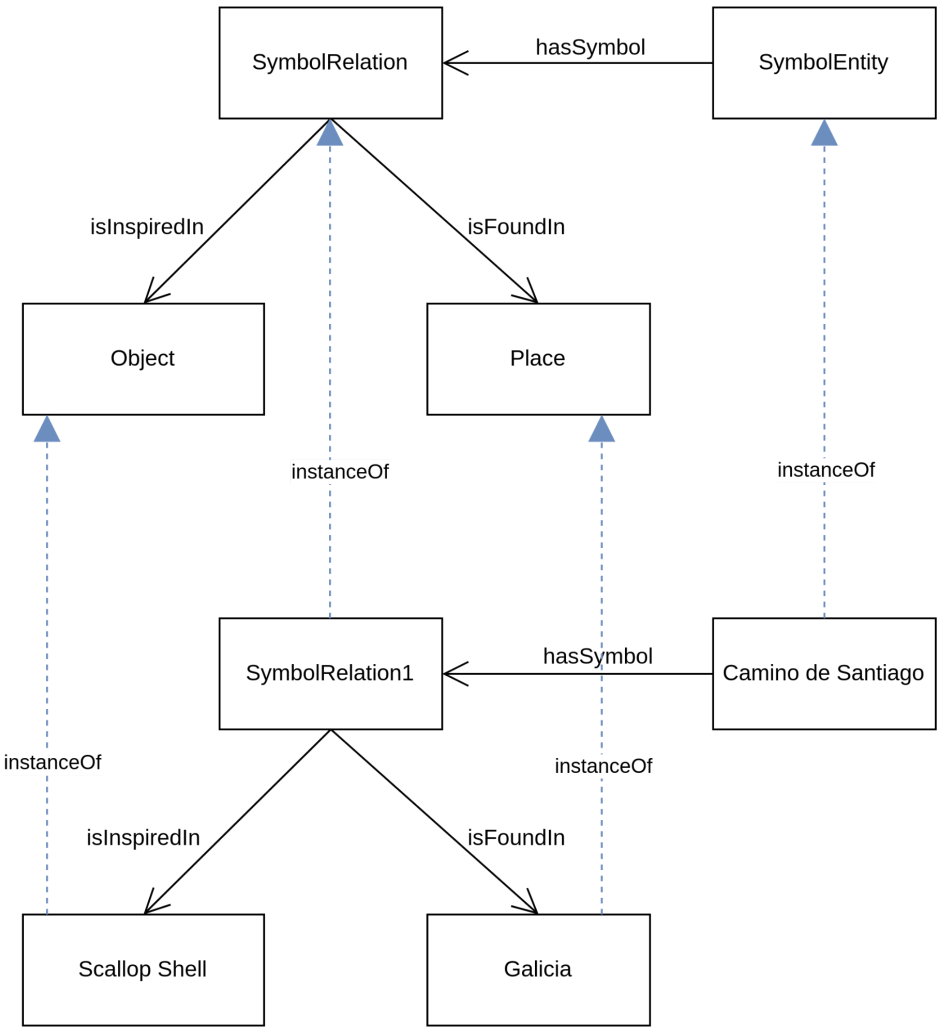
Situation 1

Possible general situation: Famous places often have symbols from a place that represent them.

Particular situation: The scallop shell, often found on the shores in Galicia, has long been the symbol of the Camino de Santiago.

Wikipedia page URL:
https://en.wikipedia.org/wiki/Camino_de_Santiago#Scallop_symbol

Graphical representation:



OWL code: link to code

Classification in the N-ray relation pattern taxonomy: Pattern 1: Describing a Relation

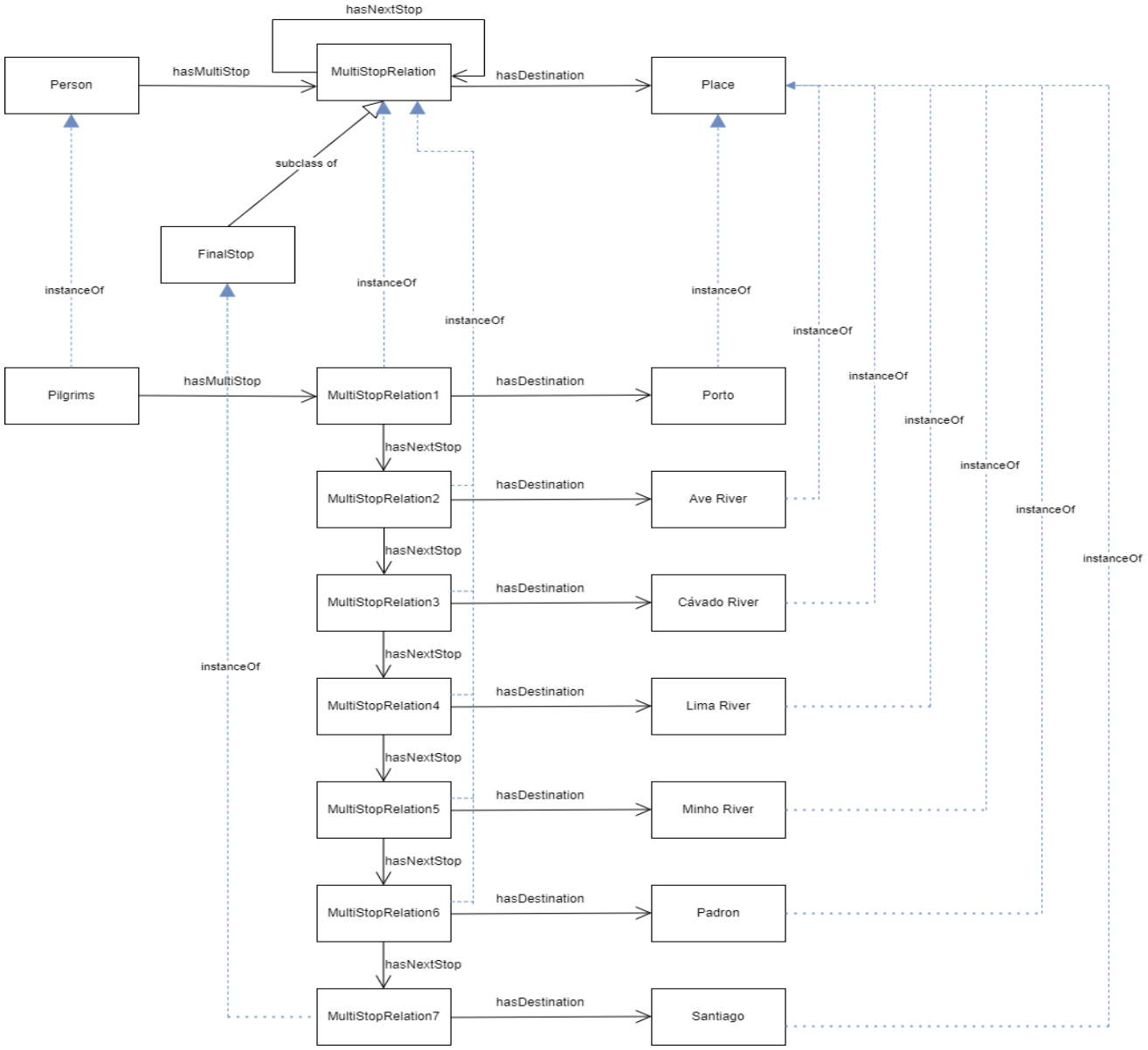
Situation 2

General Situation: People travel through different places before arriving at their destination.

Specific situation: From Porto, along the Douro River, pilgrims travel north crossing the Ave, Cávado, Lima and Minho rivers before entering Spain and then passing through Padron before arriving at Santiago.

Wikipedia page URL:
https://en.wikipedia.org/wiki/Camino_de_Santiago_(route_descriptions)

Graphical representation:



OWL code: link to code

Classification in the N-ray relation pattern taxonomy: Pattern 4: Sequence of Arguments

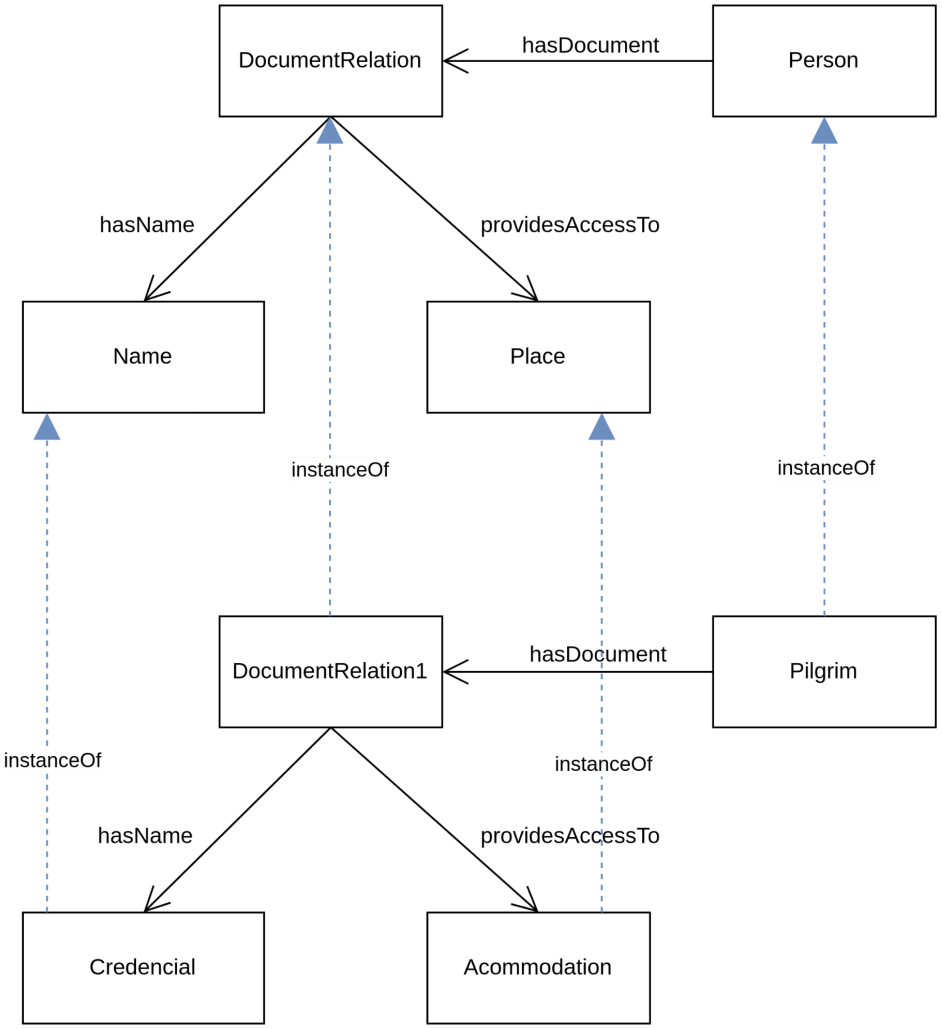
Situation 3

General Situation: Someone gets a document that provides access to something and has a name.

Specific Situation: Most pilgrims purchase and carry a document called the credencial, which gives access to overnight accommodation along the route.

Wikipedia page URL:
https://en.wikipedia.org/wiki/Camino_de_Santiago#Credencial_or_pilgrim's_passport

Graphical representation:



OWL code: link to code

Classification in the N-ray relation pattern taxonomy: Pattern 1: Describing a relation.

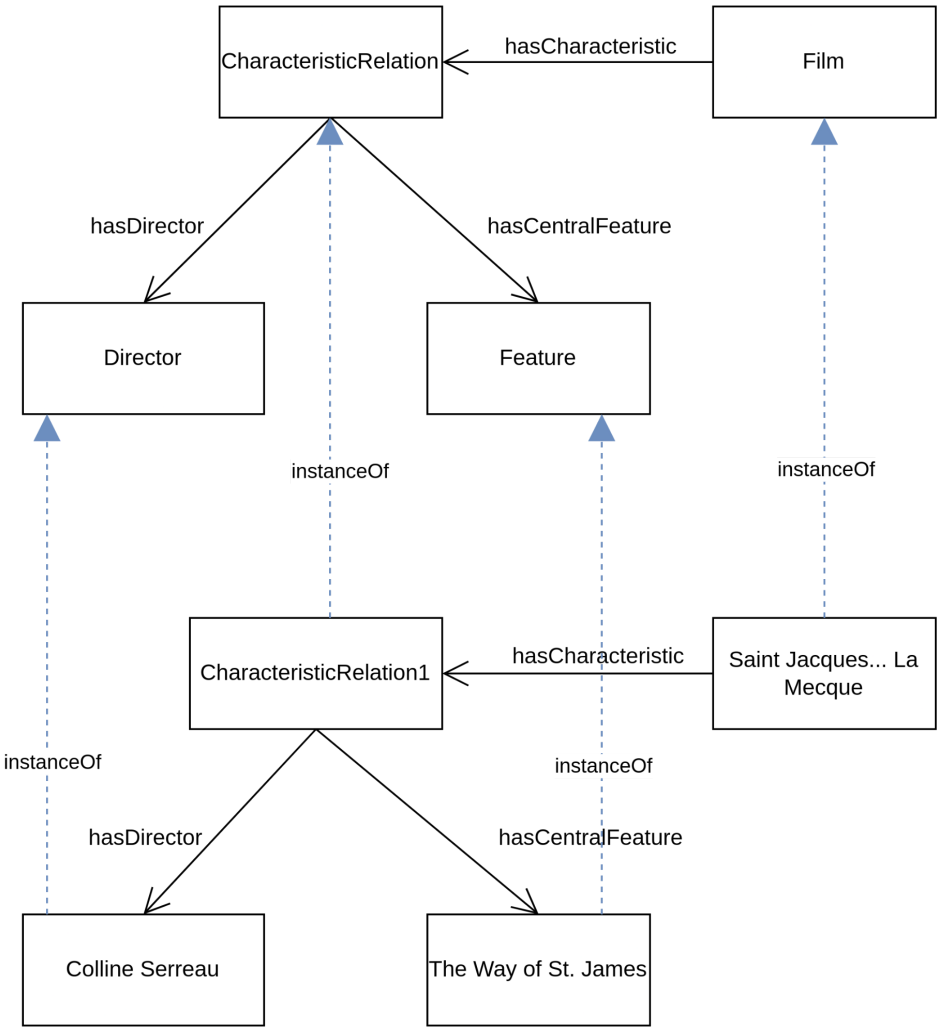
Situation 4

General Situation: A film has a story about something and is directed by someone.

Specific Situation: The Way of St. James was the central feature of the film *Saint Jacques... La Mecque* (2005) directed by Coline Serreau.

Wikipedia page URL:
https://en.wikipedia.org/wiki/Camino_de_Santiago#In_film_and_television

Graphical representation:



OWL code: link to code

Classification in the N-ray relation pattern taxonomy: Pattern 1: Describing a relation.

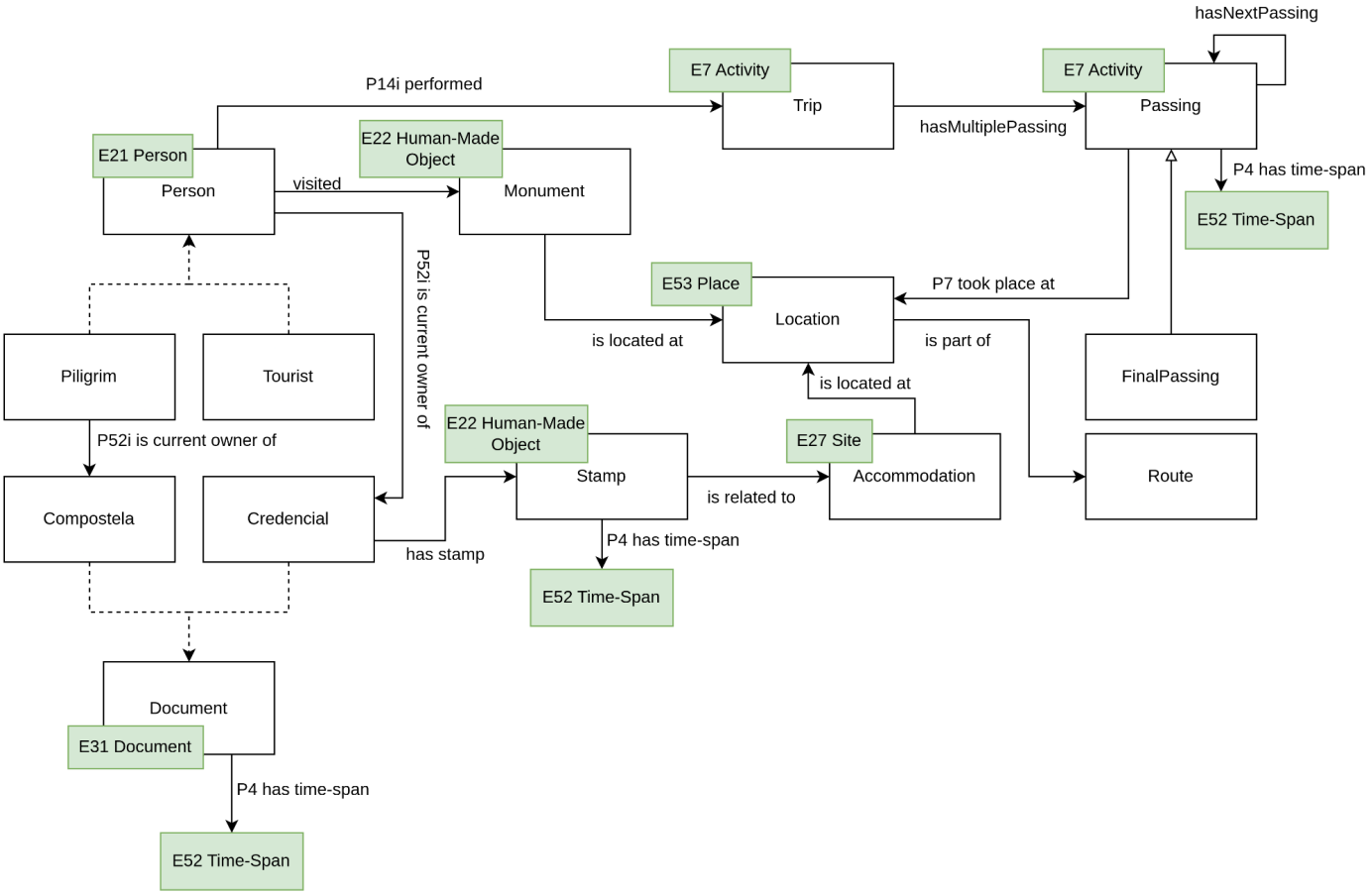
Part II

Introduction

The objective of the second part of the project is to develop an ontology from the wikipedia information on the camino de Santiago.

In developing an ontology, we decided to focus our work on routes, and everything related to it. Specifically, we considered the different locations through which the routes pass, the monuments, the accommodations, and the different people who make the journey.

Ontology



OWL Code: link to code

Wikipedia page URL: https://en.wikipedia.org/wiki/Camino_de_Santiago (source of all the mentioned concepts).

Note: The classes in green are classes from the CIDOC-CRM ontology from which we created subclasses of. For example, *Trip* is a direct subclass of *E7 Activity*.

Design Decisions:

- **Person, Pilgrim, Tourist relationship:** The Person entity is composed of pilgrims and tourists, because pilgrims need to be traced and separated as they may have different relationships to tourists. For example, pilgrims, unlike tourists, are eligible to receive a compostela document, which states that they have made the journey for religious purposes. Since tourists and pilgrims differ in some respects, we decided to create two subclasses of the person class, so that they are tracked correctly.

- **Location:** The *Location* entity is the connector to the *Route*. The idea is "A Person has been in a certain location, and this location is part of a Route" which allows to connect a Person to a Route. This way, instead of just using the *E53 Place* from the CIDOC-CRM ontology, we decided to create our class Location, a subclass of *E53 Place*, which would be uniquely identified in the ontology and would have its own data properties.

- **Person, Monument relationship:** One may ask why Person and Monument share an object property since the Monument and Person are both connected to a Location which would allow them to connect Person and Monument. The idea is that by passing in a certain location a person would not necessarily visit a specific monument. This way, we found it would be helpful to include an object property connecting Person and Monument so that it is clear when a Person visits a Museum.

- **Trip, Passing, FinalPassing, Location relationship:** This relationship suffered multiple changes throughout the development of the ontology. In the end we decided that it would be better to make use of the N-ary Relation Pattern: Sequence of Arguments. This fits well as a *Trip* is formed by a sequence of passages (*Passing*) through different places, having a final passage (*FinalPassing*). Each *Passing* is then connected to the *Location* associated with it and the next *Passing*.

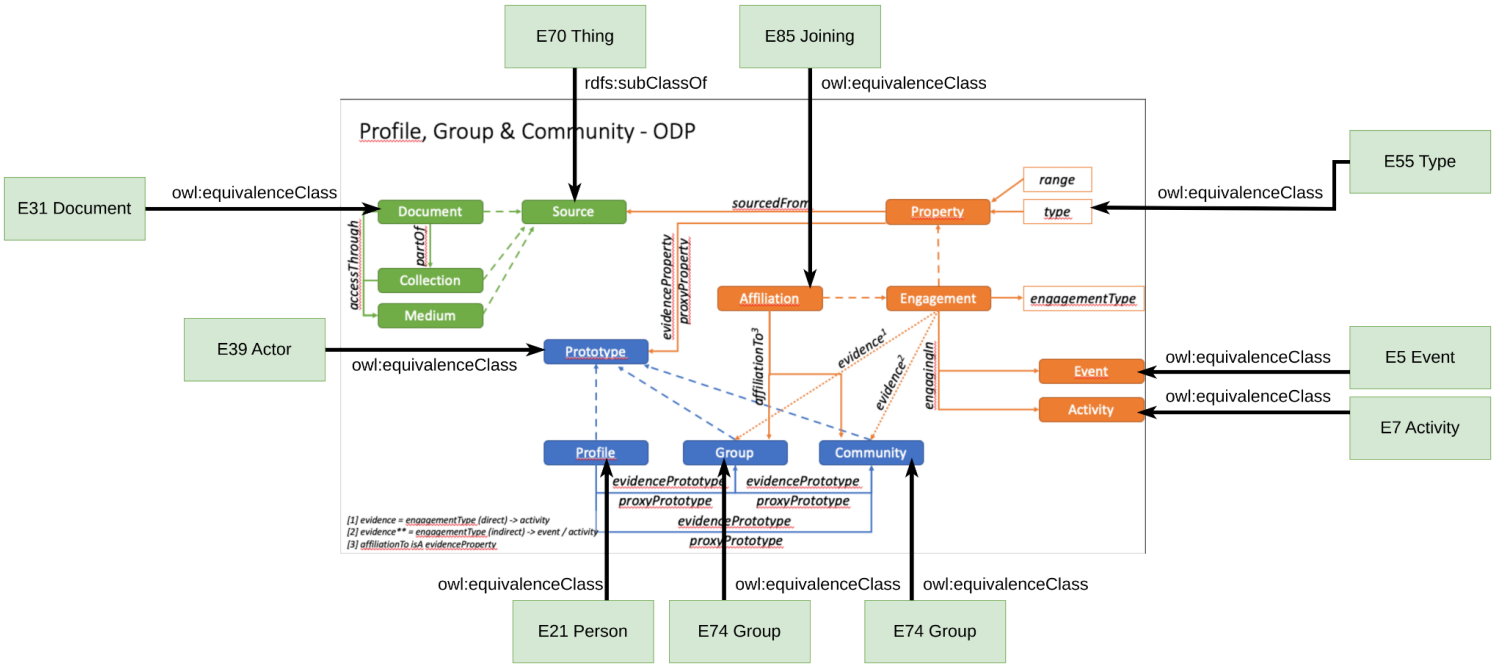
Reused Resources:

The main reused resource was the CIDOC-CRM ontology which can be found in https://www.cidoc-crm.org/. Some of the classes are similar to some defined in CIDOC-CRM and so they can be extended. More straightforward examples include: *Person* is subclass of *E21 Person*, *Document* of *E31 Document*. *E52 Time-Span* and *P4 has time-span* are used to associate a time with an entity. Both *Passing* and *Trip* are subclasses of *Activities* that are performed (*P14i performed*) by a *Person*. *P52i is current owner of* is used to state that a *Person* owns a certain document. Finally both *Stamp* and *Monument* can be related to *E22 Human-Made Object*.

With this, the only missing class to *Route*, and after searching for possible compatible ontologies in the Linked Open Vocabulary website we did not manage to find an ontology that defined a Route in the way we intend to use it.

Pilgrim, Tourist, Credencial and Compostela are very specific to our problem related to Camino de Santiago so it is expected not to find ontologies that could be reused to define these.

CIDOC-CRM module using the pattern *pgc*



Design Decisions:

- Defining equivalenceClass relationships for similar concepts: Document, Event, Activity and Type are present in both the CIDOC-CRM ontology and the pattern.
- *Profile* seems, in our understanding, to refer to an individual that can be part of a group, and this way we defined an equivalenceClass relationship with *E21 Person*. From this a straightforward equivalenceClass relationship can be also inferred from *Group* to *E74 Group*.
- Our understanding of the pattern is that a *Community* can have several inner groups and can also have isolated members (*Profiles* that are not part of a *Group* but are part of a *Community*). Being the description of *E74 Group*: “This class comprises any gatherings or organizations of human individuals or groups that act collectively or in a similar way due to any form of unifying relationship.” we considered that *Community* would also fall within the definition of *Group* and so an equivalenceClass relationship was defined.
- The description of *E39 Actor* states: “This class comprises people, either individually or in groups, who have the potential to perform intentional actions of kinds for which someone may be held responsible.” and being *Prototype* the super class of *Profile, Group* and *Community* we considered that *E39 Actor* is equivalent to *Prototype*.
- We understood *Affiliation* as the act of joining a *Group* or *Community* and so, established an equivalenceClass relationship between these classes.

- Finally, we believed that *E70 Thing* could be established as a subclass of *Source* as it is defined in the following way: “This general class comprises discrete, identifiable, instances of E77 Persistent Item that are documented as single units, that either consist of matter or depend on being carried by matter and are characterized by relative stability. They may be intellectual products or physical things. They may for instance have a solid physical form, an electronic encoding, or they may be a logical concept or structure.” which, in our understanding, encapsulates everything a source of a *Document* can be.

Final remarks: Throughout the document are provided links to OWL code that redirect to a Github repository where the code is hosted. This was done in order not to include a considerable amount of lines of code that could affect the readability of the document. These files are also present in the submitted .zip file.