

# El Viajero: A Provenance Enabled Service for News and Blog Aggregation

**Daniel Garijo** • **Boris Villazón-Terrazas** • **Oscar Corcho**  
 dgarijo@delicias.dia.fi.upm.es    bvillazon@fi.upm.es    ocorcho@fi.upm.es

Ontology Engineering Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Spain.

## Motivation, Background and Approach

Provenance of a resource is the record describing all the processes and steps involved in how the resource has come to be at its current state. It is relevant in a wide range of domains (scientific, business and social networking among others), and it is useful to determine whether to trust a resource or not.

### El Viajero Platform

El Viajero is a news and blog platform focused in the domain of travelling, belonging to the Prisacom Group.

It integrates heterogeneous contents (e.g. images, videos, photos, news, etc.) from different types of users: journalists, bloggers and normal users. Provenance is key to determine where the information comes from and how it has been modified.

### Background

Many provenance models exist for different domains:

- Focused in the description of the process that led to the resource (OPM, Provenir, Provenance Vocabulary, PML, etc.)
- Focused in the description of the resource (SIOC, DC, etc.)

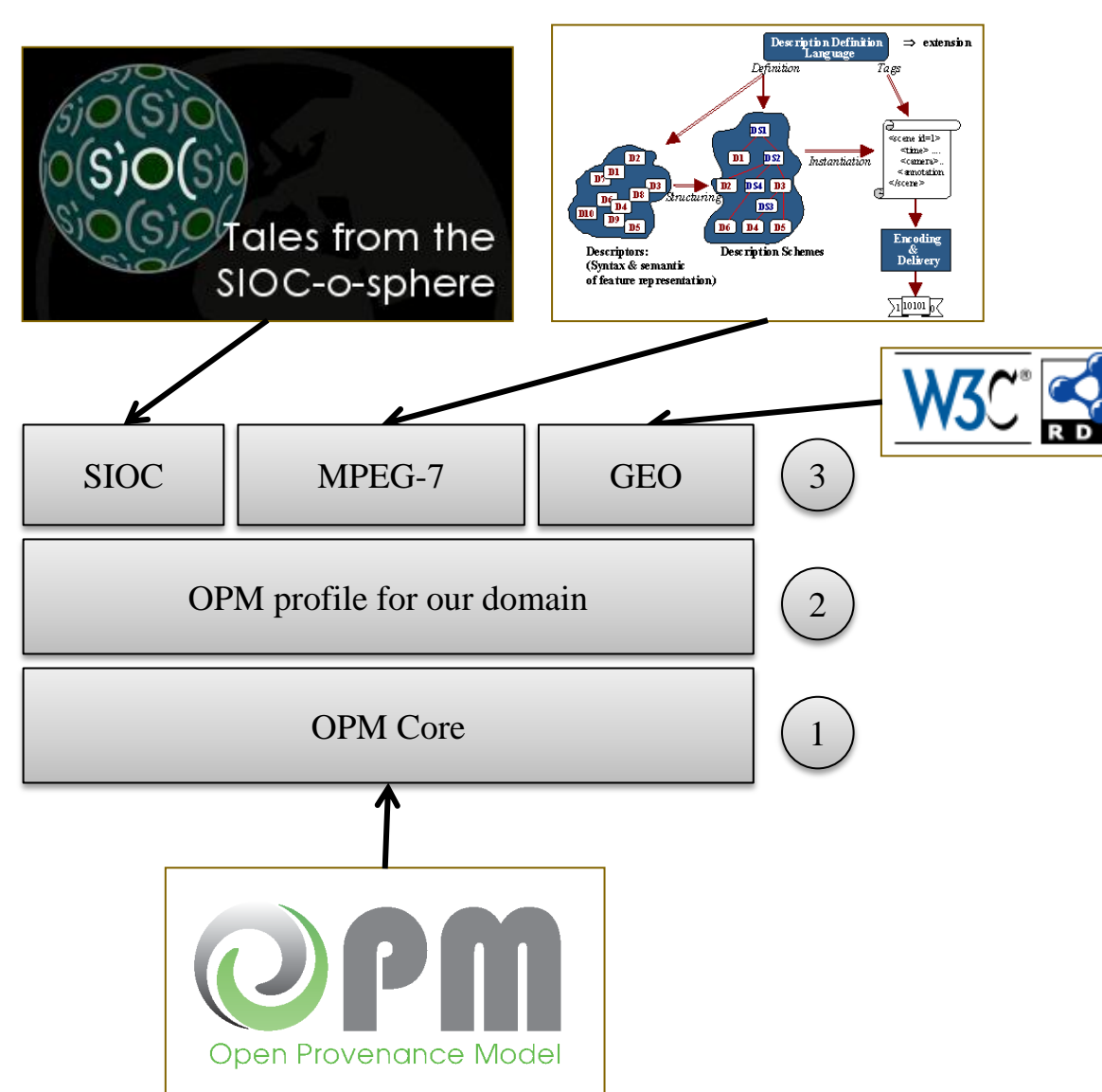
The W3C Provenance Incubator Group developed a roadmap and the first steps towards a standard, which is being currently discussed by the W3C Provenance Interchange Group.

However, no standard guidelines for provenance publishers exist at the moment.

### Open Provenance Model

Domain generic model centered in the description of the process involved in the evolution of the resources. It is the outcome of the Provenance Challenge series and has two different implementations: OPMO (OPM Ontology) and OPMV (OPM Vocabulary).

### Our approach



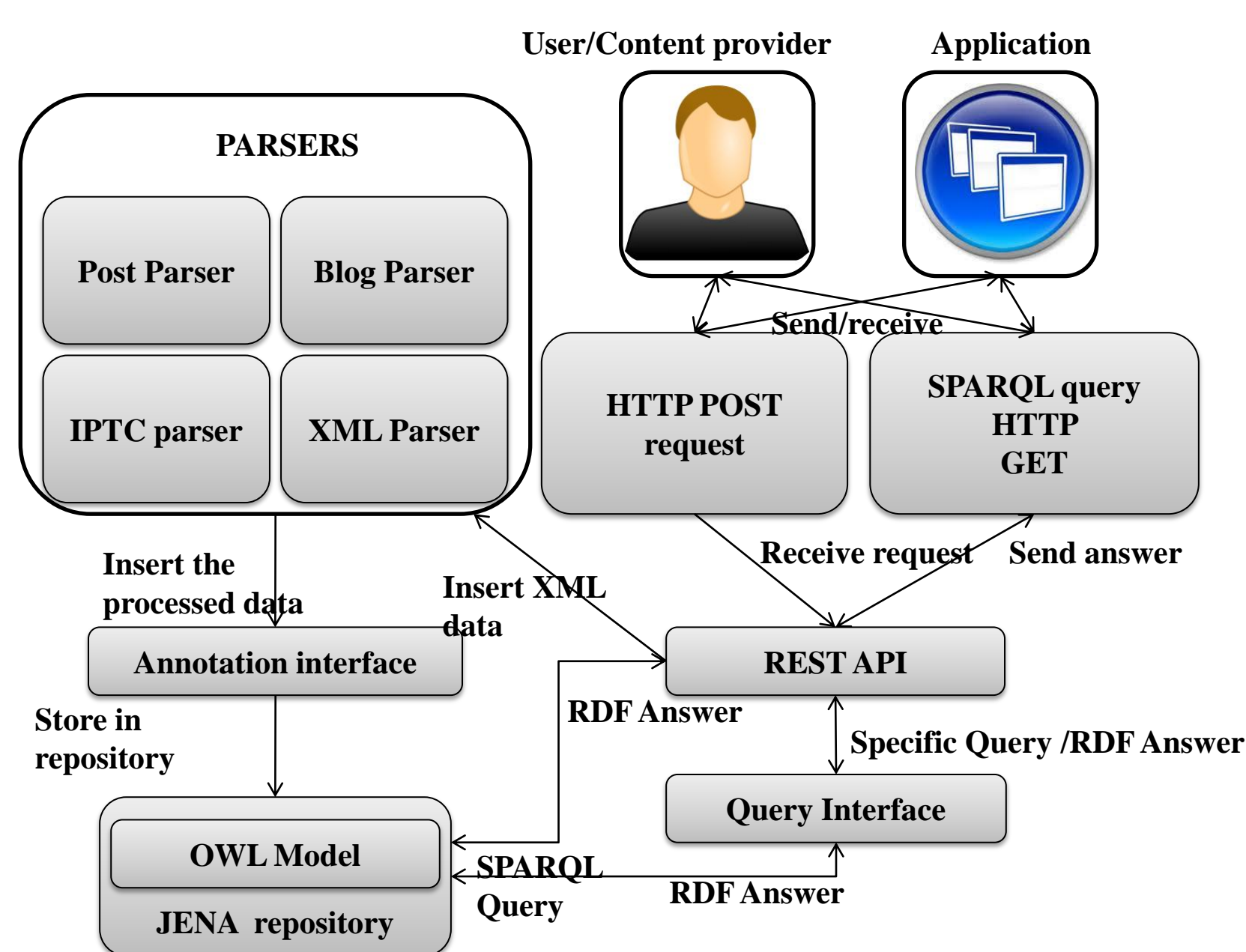
Our approach uses OPM as model to capture the provenance of the platform. It has been divided in three different layers. The first layer (1) is the core of the model (OPMO) and it is used to represent the evolution of the resources in their lifetime.

The second layer (2) extends the first one to adapt it to the specific processes and resources that can be found in our scenario.

Finally, in the last layer (3) we reuse three resource-centric vocabularies to describe the provenance of different resources available in the platform.

SIOC is used to represent the metadata from posts, blogs and social networking features. MPEG-7 is used for describing images and videos and W3C Geo is used for representing the geospatial information of the resources.

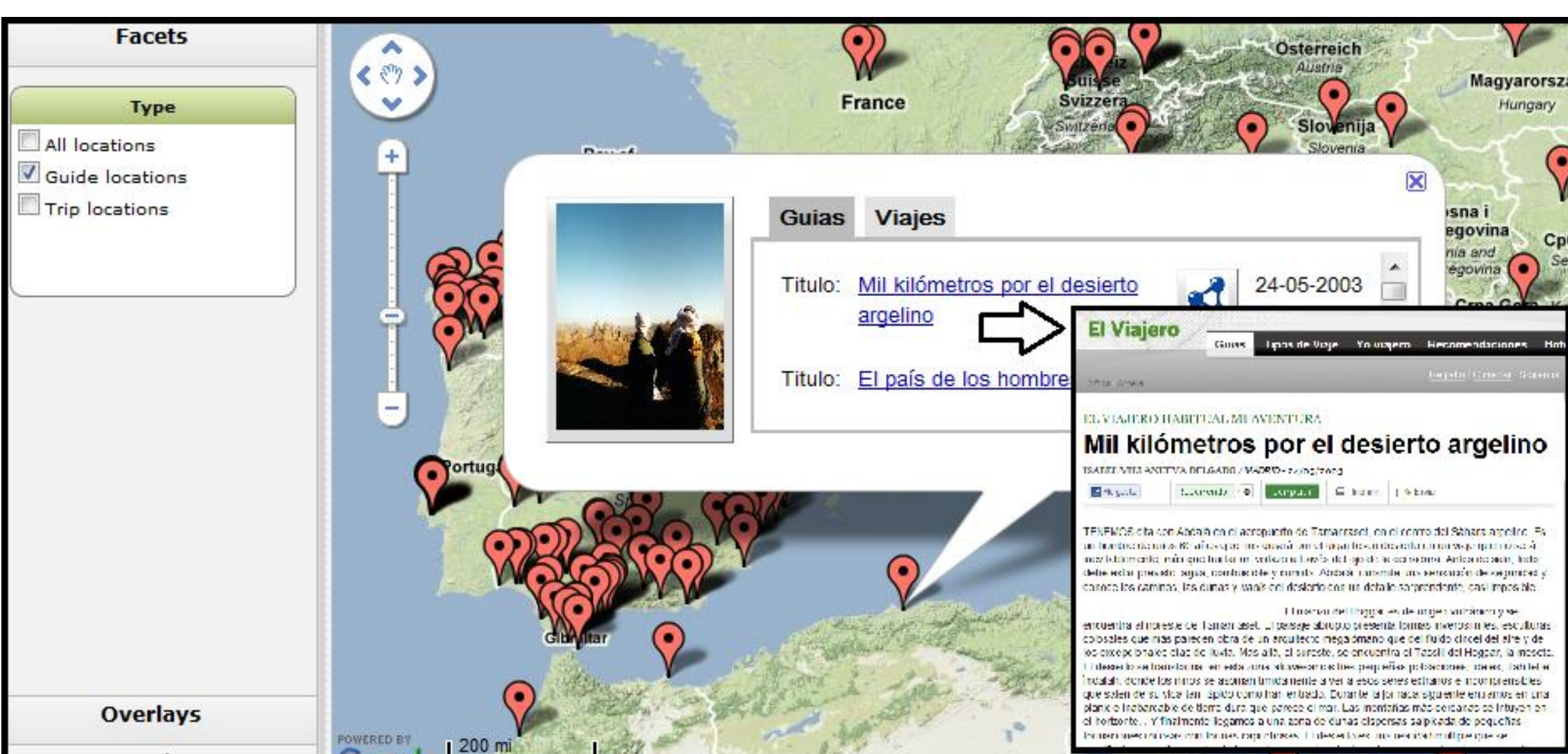
## Overview



The service for El Viajero is divided in different modules:

- The parsers process and send the information from the platform (posts, blogs, comments, IPTCs (i.e. travel guides), etc.) to the annotation interface.
- The annotation interface inserts into the repository the input received from the parsers, using our OPM adaptation as domain model.
- The query interface provides some default queries to retrieve the provenance from the resources of the platform.
- The REST API handles all the requests and inserts new contents to the repository or retrieves the requested information depending on the request.
- The repository stores the provenance for all the available resources in the platforms, under the proposed domain model.

## WebN+1 Use Case



• Guide browser and viewer (Using Map4RDF)



• Guide, post and Image Searcher

All the examples are available at: <http://webenemasuno.linkeddata.es>