Cross Forest

INEA/CEF/ICT/A2017/1566738

Tools for data exploration & live demonstration

Guillermo Vega-Gorgojo

Co-financed by the Connecting Europe Facility of the European Union

cross forest

Recap of some of the source datasets

- IFN3 (Spain)
  - 100 Microsoft Access files (2 per province)
- IFN6 (Portugal)
  - 1 shapefile (excerpt of the dataset)
- MFE50 (Spain)
  - 50 shapefiles (1 per province)
- COS6 (Portugal)
  - 1 shapefile

cross forest

Why Cross-Forest was commissioned

- Source data is of high-quality, but **difficult to access**
  - Different formats
  - Require specialized tools
  - Many different files
  - Data complexity
  - Lack of data transformations
    - no WGS84 coordinates
    - no layers with different resolutions
    - missing Genus/Family/Class abstractions
  - Sources not integrated

- As a result: datasets only exploited by a minority of the potential audience

cross forest

The Cross-Forest endpoint

- Our SPARQL endpoint
  - A single entry point
  - All datasets integrated
  - Overarching Iberian ontologies
  - Country-specific ontologies
  - Linked to other external datasets
  - 5 ⋆ dataset
  - Accessible through a regular web browser
    https://forestexplorer.gsic.uva.es/sparql

cross forest

Querying our endpoint: trees

- Counting triples

```
select count(*) where {
  ?s ?p ?o .
}
```

=> 190M

- Counting trees

```
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
select count(distinct ?tree) where {
  ?tree a ifn:Tree .
}
```

=> 1,399K

- All trees with species data?

```
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
select count(distinct ?tree) where {
  ?tree a ifn:Tree .
  FILTER EXISTS { ?tree a/rdfs:subClassOf+ ifn:Plantae }
}
```

=> **1,302K (!)**

cross forest

Querying our endpoint: trees and species

- Counting trees per species

```
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-
ambiente/ifn/>
select ?sp count(distinct ?tree) AS ?count where {
  ?tree a ifn:Tree, ?sp .
  FILTER EXISTS { ?sp rdfs:subClassOf+ ifn:Plantae }
}
GROUP BY ?sp
ORDER BY DESC(?count)
```

cross forest

Querying our endpoint: wrong species

- Discovering a wrong species in the data

```
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
select distinct ?sp where {
  ?tree a ifn:Tree, ?sp .
  FILTER NOT EXISTS { ?sp a ifn:Species }
  FILTER NOT EXISTS { ?sp a ifn:Genus }
  FILTER NOT EXISTS { ?sp a ifn:Family }
  FILTER NOT EXISTS { ?sp a ifn:Class }
  FILTER ( strstarts(str(?sp), "https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/Species") )
}
```

=> **ifn:Species269 (!)**

cross forest

Querying our endpoint: plots

- Obtaining the plots in a bounding box

```
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
PREFIX pos: <http://crossforest.eu/position/ontology/>
SELECT DISTINCT ?plot ?lat ?lng WHERE {
  ?plot a/rdfs:subClassOf* ifn:PrimaryPlot ;
    pos:hasPosition ?pos .
  ?pos pos:hasCoordinateReferenceSystem <http://epsg.w3id.org/data/crs/4326> ;
    <http://epsg.w3id.org/ontology/axis/1> ?lat ;
    <http://epsg.w3id.org/ontology/axis/2> ?lng .
  FILTER (?lat < 42.17)
  FILTER (?lat > 41.77)
  FILTER (?lng < -6.51)
  FILTER (?lng > -7.57)
}
```

cross forest

Querying our endpoint: plots and trees

- Obtaining the trees of a specific plot

```
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
PREFIX pos: <http://crossforest.eu/position/ontology/>
PREFIX dplot: <https://datos.iepnb.es/recurso/sector-publico/medio-
ambiente/ifn/plot/>
SELECT DISTINCT ?tree ?lat ?lng WHERE {
  ?tree a ifn:Tree ;
    pos:hasPosition ?pos ;
    ifn:isInPlot dplot:32-1904-A-1 .
  ?pos pos:hasCoordinateReferenceSystem <http://epsg.w3id.org/data/crs/4326> ;
    <http://epsg.w3id.org/ontology/axis/1> ?lat ;
    <http://epsg.w3id.org/ontology/axis/2> ?lng .
}
```

cross forest

Querying our endpoint: patches

- Counting patches in layers

```
PREFIX ilu: <http://crossforest.eu/ilu/ontology/>
PREFIX pos: <http://crossforest.eu/position/ontology/>
PREFIX epsg: <http://epsg.w3id.org/ontology/>
SELECT ?layer ?countryPatch count(DISTINCT ?patch)
WHERE {
  ?patch a ?countryPatch ;
     pos:hasPolygon ?poly .
  ?countryPatch rdfs:subClassOf* ilu:Patch .
  ?poly pos:isInLayer ?layer .
}
GROUP BY ?layer ?countryPatch
ORDER BY ASC(?countryPatch) ASC(?layer)
```

cross forest

Querying our endpoint: Spanish patches

- Obtaining the **Spanish patches** with +1M km^2 in a bounding box

```
PREFIX mfe: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/mfe/>
PREFIX pos: <http://crossforest.eu/position/ontology/>
PREFIX epsg: <http://epsg.w3id.org/ontology/>
SELECT DISTINCT ?patch ?poly ?west ?east ?north ?south ?area
WHERE {
  ?patch a mfe:Patch ;
    pos:hasPolygon ?poly .
  ?poly epsg:hasLeftBound107 ?west ;
    epsg:hasRightBound107 ?east ;
    epsg:hasUpperBound106 ?north ;
    epsg:hasLowerBound106 ?south ;
    pos:hasAreaInSquareMeters ?area ;
    pos:isInLayer <http://crossforest.eu/ilu/data/layer/s5s5> .
  FILTER (?south < 42.17)
  FILTER (?north > 41.77)
  FILTER (?west < -6.51)
  FILTER (?east > -7.57)
  FILTER (?area > 1000000)
}
```

cross forest

Querying our endpoint: Iberian patches

- Obtaining the **Iberian** patches with +1M km^2 in a bounding box

```
PREFIX ilu: <http://crossforest.eu/ilu/ontology/>
PREFIX pos: <http://crossforest.eu/position/ontology/>
PREFIX epsg: <http://epsg.w3id.org/ontology/>
SELECT DISTINCT ?patch ?poly ?west ?east ?north ?south ?area
WHERE {
  ?patch a/rdfs:subClassOf* ilu:Patch ;
    pos:hasPolygon ?poly .
  ?poly epsg:hasLeftBound107 ?west ;
    epsg:hasRightBound107 ?east ;
    epsg:hasUpperBound106 ?north ;
    epsg:hasLowerBound106 ?south ;
    pos:hasAreaInSquareMeters ?area ;
    pos:isInLayer <http://crossforest.eu/ilu/data/layer/s5s5> .
  FILTER (?south < 42.17)
  FILTER (?north > 41.77)
  FILTER (?west < -6.51)
  FILTER (?east > -7.57)
  FILTER (?area > 1000000)
}
```

cross forest

Querying our endpoint: plots and patches

- Obtaining the plots contained in a patch (approximation)

```
PREFIX pos: <http://crossforest.eu/position/ontology/>
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
PREFIX epsg: <http://epsg.w3id.org/ontology/>
PREFIX dpatch: <https://datos.iepnb.es/recurso/sector-publico/medio-
ambiente/mfe/patch/>
SELECT DISTINCT ?plot WHERE {
  dpatch:s5s5-49-994 pos:hasPolygon ?poly .
  ?poly epsg:hasLeftBound107 ?west ;
    epsg:hasRightBound107 ?east ;
    epsg:hasUpperBound106 ?north ;
    epsg:hasLowerBound106 ?south .
  ?plot a ifn:Plot ;
    pos:hasPosition ?pos .
  ?pos pos:hasCoordinateReferenceSystem <http://epsg.w3id.org/data/crs/4326> ;
    <http://epsg.w3id.org/ontology/axis/1> ?lat ;
    <http://epsg.w3id.org/ontology/axis/2> ?lng .
  FILTER (?lat <= ?north)
  FILTER (?lat >= ?south)
  FILTER (?lng <= ?east)
  FILTER (?lng >= ?west)
}
```

cross forest

Querying our endpoint: GeoSPARQL

- Obtaining the plots contained in a patch using GeoSPARQL

```
PREFIX pos: <http://crossforest.eu/position/ontology/>
PREFIX ifn: <https://datos.iepnb.es/def/sector-publico/medio-ambiente/ifn/>
PREFIX dpatch: <https://datos.iepnb.es/recurso/sector-publico/medio-
ambiente/mfe/patch/>
SELECT DISTINCT ?plot WHERE {
  dpatch:s5s5-49-994 pos:hasPolygon ?poly .
  ?poly <http://www.opengis.net/ont/geosparql#asWKT> ?wkt_patch .
  ?plot a ifn:Plot ;
      pos:hasPosition ?pos .
  ?pos pos:hasCoordinateReferenceSystem <http://epsg.w3id.org/data/crs/4326> ;
      <http://epsg.w3id.org/ontology/axis/1> ?lat ;
      <http://epsg.w3id.org/ontology/axis/2> ?lng .
FILTER ( bif:st_within( bif:st_point(xsd:double(?lng), xsd:double(?lat)),
?wkt_patch, 0) )
}
```

cross forest

Recap of our live queries

- Quite powerful, i.e. very expressive queries
  - Queries to get an overview of the data
  - Queries to zoom and filter
  - Queries to get details if needed
- Seamless querying of the integrated dataset
  - Exploiting the Iberian ontology
  - Exploiting geolocations to find connections between entities from different datasets, e.g. patches and plots
- But…
  - Requires expert knowledge on SPARQL
  - Need to check the ontology documentation
  - Error-prone

cross forest

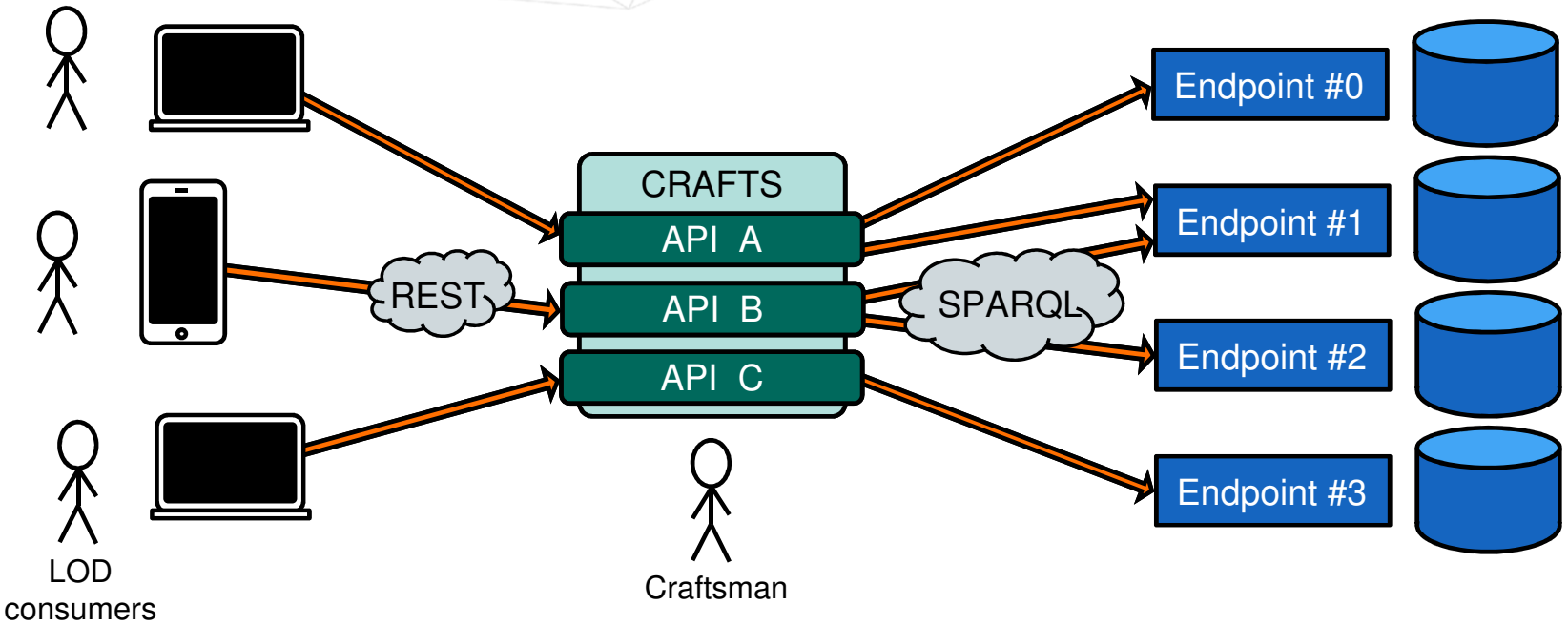
Forest Explorer

- Aim of Forest Explorer
  - Simplify access to the Cross-Forest endpoint
  - Interactive map
  - Requires no knowledge of SPARQL, RDF/S, OWL
- It works!
  - https://forestexplorer.gsic.uva.es/
  - G. Vega-Gorgojo, J.M. Giménez-García, C. Ordóñez, F. Bravo. Pioneering Easy-to-Use Forestry Data with Forest Explorer. Semantic Web. 2021
- But…
  - Development required expert knowledge of SPARQL
  - Adapting to ontology changes and dataset extensions (e.g. Portuguese data) is painful

cross forest

CRAFTS

- CRAFTS: Configurable RESTful APIs For Triple Stores
- Key idea
  - Simplify access to Linked Open Data by providing configurable RESTful APIs
- Generic solution, not tied to the Cross-Forest case



cross forest

Forest Explorer is now CRAFTS-based

- Current version of Forest Explorer uses a CRAFTS API
  - Portuguese data included
  - Code simplification

- Quantifying code simplification in Forest Explorer
  - Source lines of code (SLOC) analysis
  - Same developer => main problems of SLOC analysis are vanished
  - 10% code reduction (overall)
  - 31% code reduction of the *Data manager* component

cross forest

DEMO TIME!

https://forestexplorer.gsic.uva.es/

cross forest

cross forest

INEA/CEF/ICT/A2017/1566738

Co-financed by the Connecting Europe Facility of the European Union