

Ithaca:

From Open Data to Linked Open Data

Michalis Mountantonakis
mountanton@csd.uoc.gr

Pavlos Fafalios
fafalios@csd.uoc.gr



Need

The provided datasets are not “linked”

Their data is not accessible through an API

The user/developer cannot:

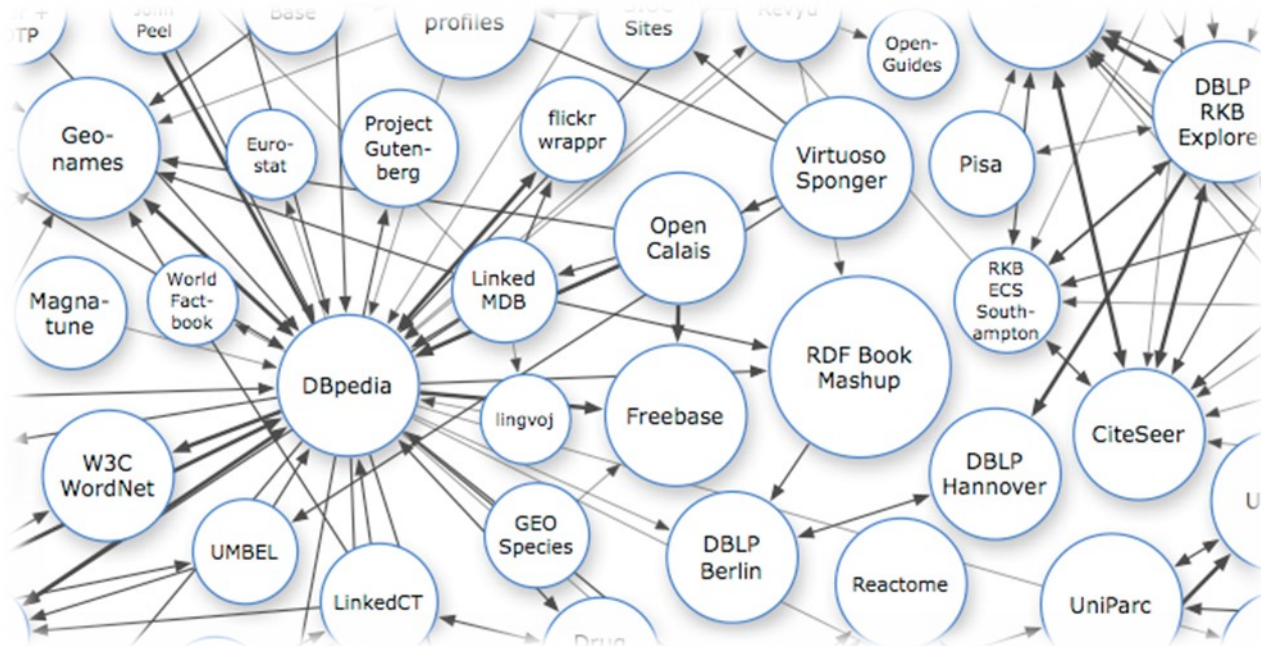
- Browse into the data
- Inspect if and how the underlying data are connected
- Combine information coming from different datasets

The user/developer must:

- Manually download one or more datasets
- Examine the data of each dataset and its structure
- Build software that reads, cleans and exploits the data

Our hack...

solves this problem by exploiting the **Linked Open Data (LOD)** and their dynamic and distributed nature



Data

We utilized 21 federated datasets:

- Wineries
- Ports
- Ancient building, places, monasteries
- Churches
- Beaches, Routes (bicycle, botanic, E4, ...), Museums, Caves
- ...

However:

- Our approach can be exploited with ALL datasets in ANY language

Under the hood

[A] We designed an OWL ontology for describing:

- The datasets (their metadata)
- The data in the each dataset

The ontology contains:

- 19 Classes, 4 Object properties, 57 Datatype properties

[B] We created a Java program that:

- Reads the data (by creating a parser for each different type of sources)
- Creates Linked Data (i.e. RDF triples), using the ontology and other well-known vocabularies (RDFS, VoID, DublinCore, vCard, Geo, FOAF, XSD)
- Ingests the RDF triples in a Virtuoso Warehouse, offering an API for accessing the data (**SPARQL Endpoint**)



Impact

Ithaca...for developers!

The user/developer can now run complex (SPARQL) queries over all the data and combine information coming from multiple datasets. E.g.:

- *Give me all wineries (together with a description in English) that are in Heraklion and have bus accessibility, and that there is at least one point of interest (e.g. a museum) in distance < 5 km.*
- *Give me all ports with fuel availability and internet, with a gorge or an ancient place in near distance and a description in Greek of the gorge or the ancient place .*
- ...

The user/developer can now build smart applications that use this data and combine information exploiting also other external LOD datasets, e.g. DBpedia, Freebase, etc.

Next steps

- Automate as much as possible the procedure
- Transform and ingest all HOMER federated datasets in the Warehouse
- Improve accordingly the Ontology
- Create same-as triples in order to link the data of the warehouse to data of other datasets (of the LOD cloud)



Harmonising
Open data in the
Mediterranean through better access and
ReUSE of public sector information



Demo...



Harmonising
Open data in the
Mediterranean through better access and
ReUSE of public sector information



Ithaca

Thank you!