

Understanding heterogeneous data sources with ConnectionLens

Nelly Barret – Inria (CEDAR team) & LIX
Supervised by Ioana Manolescu



In the shoes of ...

Article from Liberation (text)

Balkany mineur de fonds

L'élu de **Levallois-Perret** est soupçonné d'avoir touché 5 millions de dollars de commission en 2009 grâce à son rôle d'intermédiaire entre **Areva** et la **Centrafrique** dans le dossier **Uramin**.
[...]

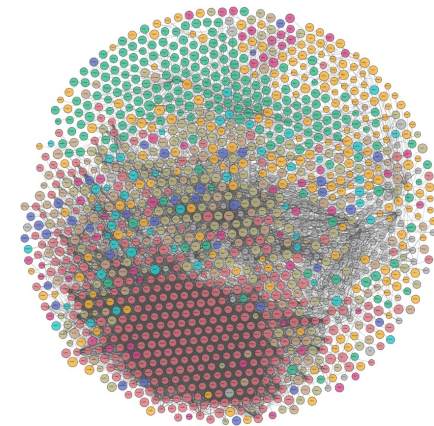
National directory of elected officials (JSON)

```
{  
  name: "Levallois-Perret",  
  mayor: "P. Balkany",  
  city-council: [  
    {name: "I. Balkany"}, ... ]  
}, ...]
```

DBpedia.fr (RDF)

```
{  
  dbr:Marrakech  
    dbr:name      "Marrakech"  
    rdf:type      dbo:City ;  
    dbo:country   dbr:Morocco .  
  dbr:Morocco  
    dbr:name      "Morocco"  
    rdf:type      dbo:Country  
    dbo:locatedIn dbr:Africa .  
  dbr:CentralAfricanRepublic  
    dbr:name      "Central African Republic"  
    dbo:locatedIn dbr:Africa .  
}
```

Linked Open Data Cloud (RDF)



What is the general purpose of this huge knowledge graph?

I work in the Big Data and can't analyse each data source

What is the general purpose of these documents?

Are they linked?



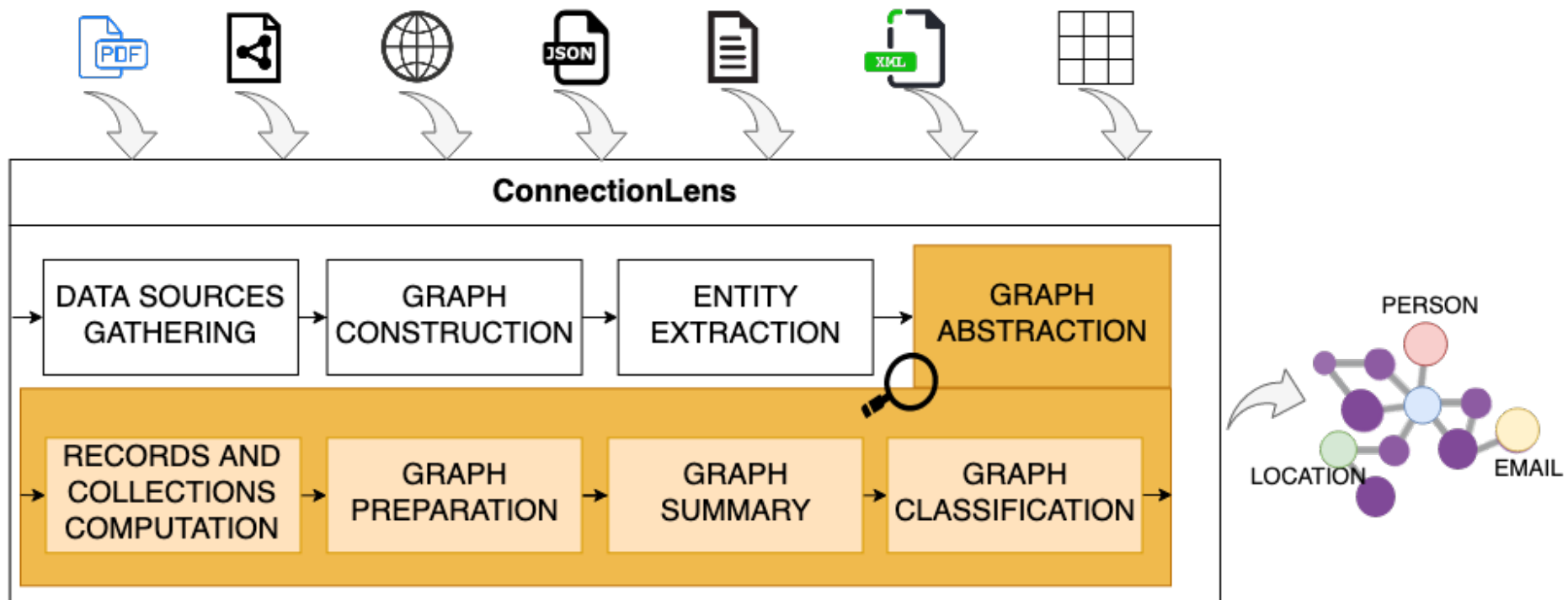
Big Data (any kind)



These questions are answered by our approach: ConnectionLens

The ConnectionLens approach

Objective: help a human user grasp the content of a data source



A concrete example

- You are the CEO a lab and you have to write a report about your employees and the worplaces.



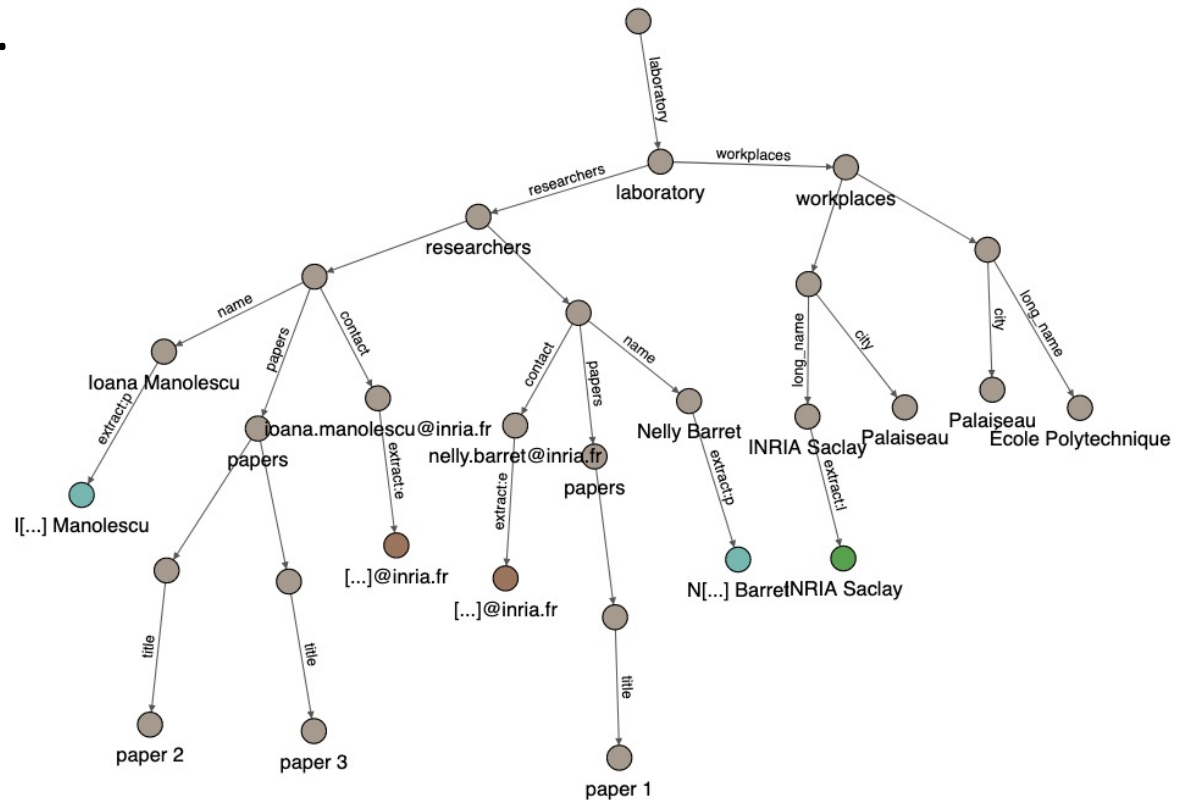
researchers.json



papers.sql

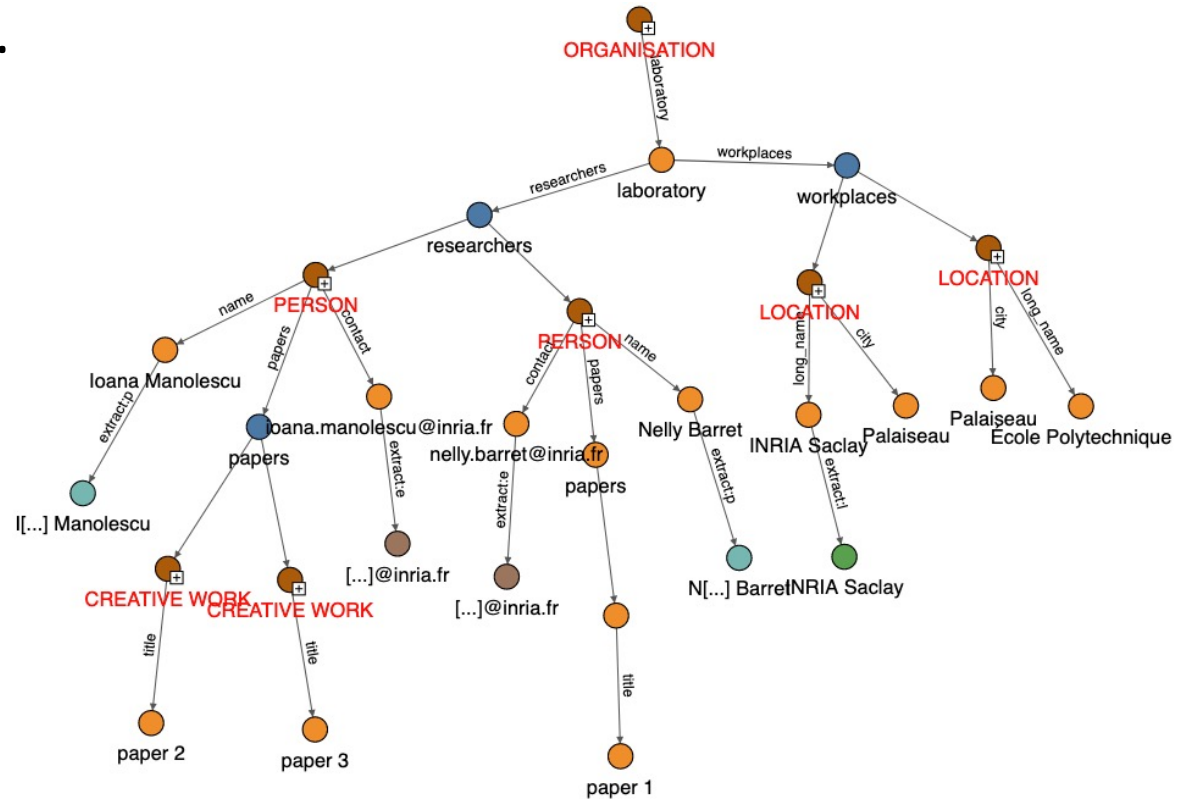


workplaces.xls



A concrete example

- You are the CEO a lab and you have to write a report about your employees and the worplaces.



If you enjoyed this presentation or if you have questions,
feel free to come during the poster session!

Nelly Barret – 1st year PhD student (Inria & LIX)

 nelly.barret@inria.fr

 nellybarret.fr