

# Tracking deputies in the news using OpenData and Free software

Vincent Michel - Logilab

Law Factory 2012 - 6 juillet 2012

# Merge news articles and open data

- Open databases : *General data* (e.g. Dbpedia), *Political data* (e.g. NosDeputes.fr, data.gouv.fr)
- Information feeds : *Twitter, Blogs, Newspaper RSS*
- Semantic web framework : *CubicWeb* (LGPL licence)

## Parse the news

Extract nouns of people and places :

... **Olivier Falorni**, qui a battu **Ségolène Royal** à **La Rochelle**, a déclaré ce mardi en arrivant à l'**Assemblée nationale** qu'il "siégerait à gauche" pour soutenir l'action de **François Hollande** rappelant les liens anciens qui l'unissent au chef de l'Etat...

# URIs as Identifiers

... relate nouns to entries in open databases ...

- **Ségolène Royal :**

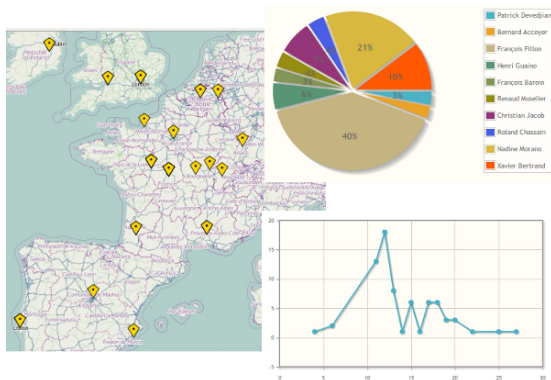
- ▶ *[http://elections.gouv.fr/Ségolène\\_Royal](http://elections.gouv.fr/Ségolène_Royal)*
- ▶ *[http://dbpedia.org/resource/Ségolène\\_Royal](http://dbpedia.org/resource/Ségolène_Royal)*

- **La Rochelle :**

- ▶ *[http://elections.gouv.fr/La\\_Rochelle](http://elections.gouv.fr/La_Rochelle)*
- ▶ *[http://dbpedia.org/resource/La\\_Rochelle](http://dbpedia.org/resource/La_Rochelle)*

# Analyse the data collected

Input **queries** / Pick **visualization**



# Examples of queries

- *Give me the news articles about "Nadine Morano"*
- *Give me the number of articles about UMP deputies born after 1960*
- *Give me the number of articles about (Socialist) deputies that have been candidate to the 2012 legislatives elections*

## Examples of visualisation

- *Give me, for each news where these information appear, the latitude and longitude of a place, and the name of the deputy that are linked in this news.*
- *Give me, for each candidat of the 2012 legislatives election that appears in the news, his/her political group and name, and the corresponding number of articles.*
- *Give me, for each day, the number of news about Ségolène Royal and Olivier Falorni.*

# Conclusion

## Automatic tracking of deputies in the news, and powerful analyze possibilities

- **Still work in progress !**

- ▶ Improving the named entities recognition part.
- ▶ Improving the multiple databases query system.
- ▶ Add more reference databases, more views...

- **However, interesting features !**

- ▶ *Use of URI as unique reference.*
- ▶ *Allow complex queries with large amount of available information.*