

# Big Data for Labour Market Information (LMI) Building the system of demand analysis with online job vacancy data

**Webinar 2 June 2020**

**Preliminary results of skills demand analysis  
Other consolidated uses and cases**

Session 2

Moving ahead: building the system of Big Data for LMI

Alessandro Vaccarino, Tabulaex

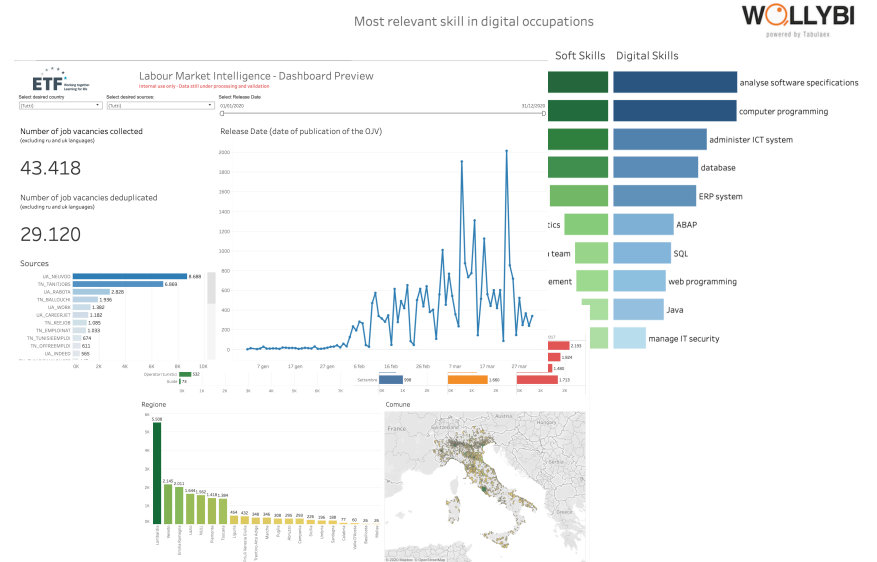
# WHAT'S THE GOAL?

Transform those...

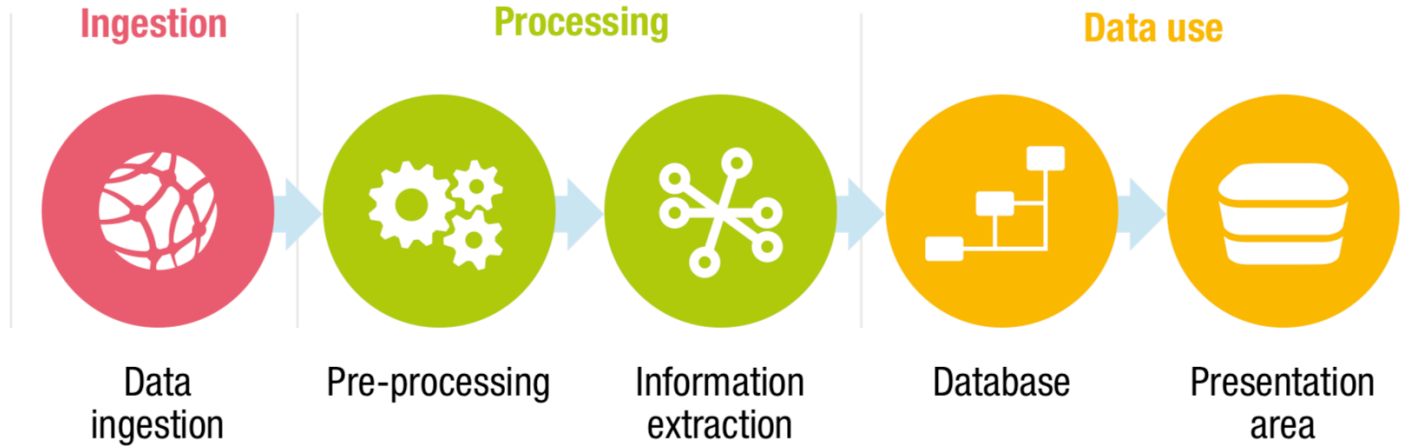
The collage features several overlapping job advertisements. Key elements include:
 

- Junior Software Developer:** Multiple listings with details like 'Candidati facilmente', 'Our client is looking for...', and 'Location: Office or home'.
- Lead Clinical Data Manager - Office or home based:** A prominent listing for 'Warman O'Brien - Milano, Lombardia' with a salary range of '€40,000 - €50,000 all'anno'.
- Supply Chain - Magazziniere:** A listing for 'La Risorsa Umana.it' in 'MODENA, EMR'.
- Database design specifications, edit check:** A listing for 'study teams'.
- Attractive salary and benefits package:** A listing for 'working options' and 'career development opportunities'.
- Lead large or complex EDC studies and study teams:** A listing for 'Understand and have experience in database design specifications, edit check'.

...to this



# HOW TO DO IT?

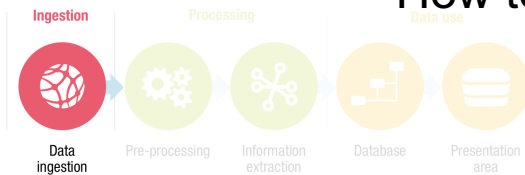


# DATA INGESTION – SELECT THE RIGHT SOURCES

Web Labour market can offer a wide range of sources:

- National/International portals
- Sector specific/generic
- Company websites/Newspapers/Aggregators/Public Employment Services/...
- Social networks
- ...

How to select the most appropriate and rank them?





# DATA INGESTION – COLLECTING DATA – GOAL...

**JUNIOR SOFTWARE DEVELOPER**

Location: United Kingdom  
Application deadline: Saturday, 30 September 2017  
Reference number: 100

APPLY NOW

Home > Now Hiring: Software Developers > Junior Software Developer

Share:

**Description**

As Junior Software Developer, you will develop excellent software for use in field mapping, data collection, sensor networks, street navigation, and more. You will collaborate with other programmers and developers to autonomously design and implement high-quality web-based applications, restful APIs, and third party integrations.

We're looking for a passionate, committed developer that is able to solve and articulate complex problems with application design, development and user experiences. The position is based in our offices in Harwell, United Kingdom.

Title:

Junior Software Developer

Area:

United Kingdom

Time:

Saturday, 30 September 2017

Description:

As Junior Software Developer, you will develop excellent software for use ...



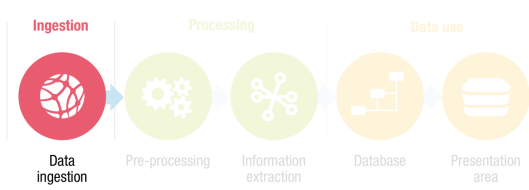
# DATA INGESTION – COLLECTING DATA – ...STRATEGY...



Focus on  
volumes



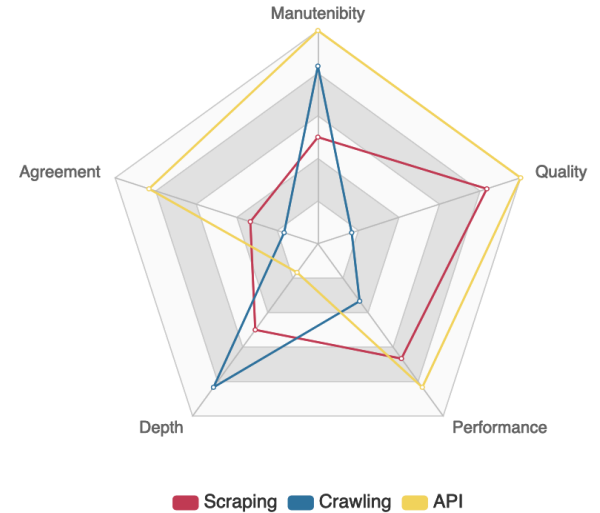
Coverage  
augmentation &  
maximization



# DATA INGESTION – COLLECTING DATA – ...TECHNIQUES

Tailored approach, aimed to maximize efficiency:

- Polite ingestion (no overloading of portals)
- Extract specific attributes when available (sector, place,...)
- Scale when needed



# DATA PREPROCESSING – GOAL...

After Data Ingestion phase, a good amount of Web Pages is collected. But:

- How many of them are really OJVs?
  - Noise (social profile pages, policy pages,...)
  - Fake vacancies (trainings, consultancy services,... published on OJVs' portals)
- How many of them refer to a unique OJV?



# DATA PREPROCESSING – ...STRATEGY...



Language  
detection



Noise  
reduction

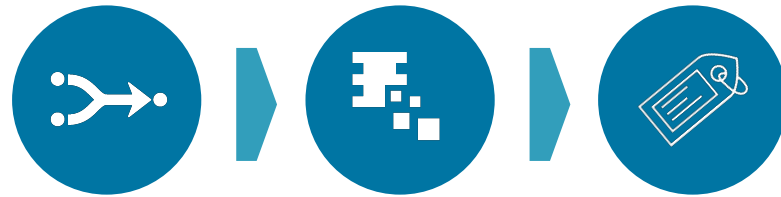


OJVs  
Deduplication



# DATA PREPROCESSING – ...TECHNIQUES

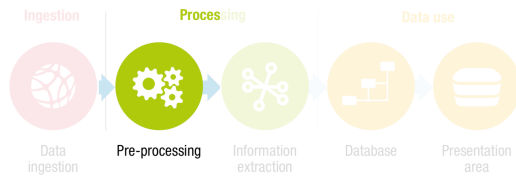
Several Artificial Intelligence techniques are applied to detect noise (such a mail spam filter) and to detect duplicated OJVs.  
Language Detection is the first step, aimed to focus on language specific pipelines.



Merging

Cleaning

Text processing  
and summarizing



# DATA CLASSIFICATION – GOAL...

After Data Preprocessing phase, a relevant amount of Online Job Vacancies is extrated. Now, it's time to **extract information**:

- Which profession is required?
- There is a place where the job will be done?
- Is the job focused on a specific sector?
- Which kind of skills are required?
- ...



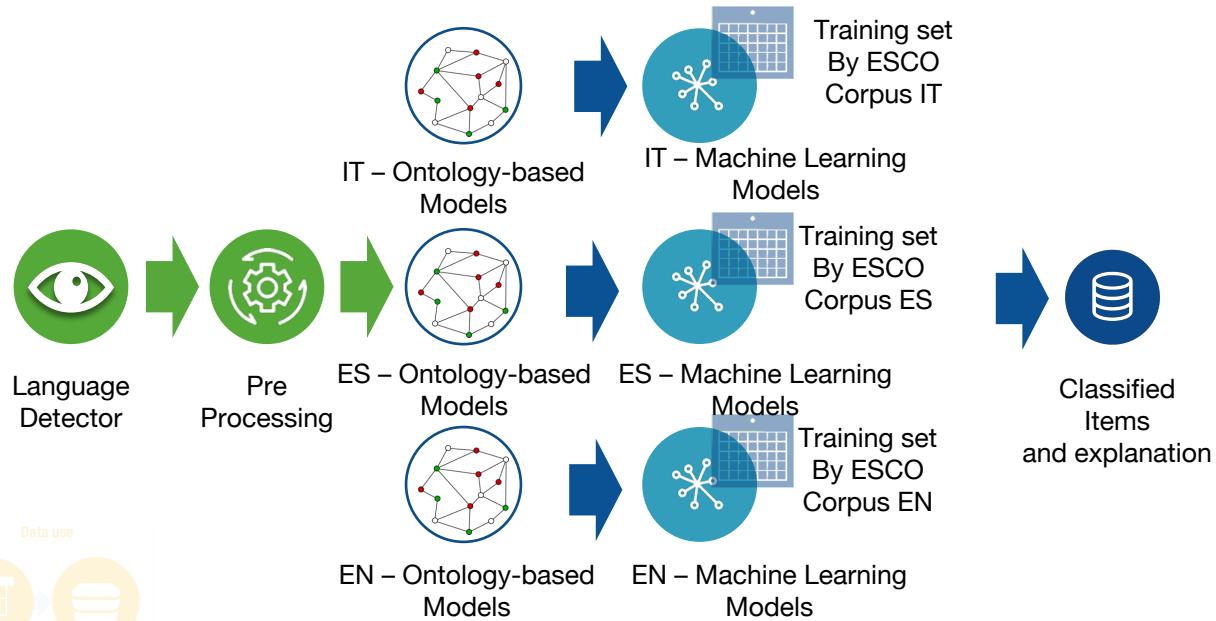
# DATA CLASSIFICATION – ...STRATEGY...

Adoption of Taxonomies for each analytical variable, to set a domain of values:

- Occupation (ESCO)
- Skills (ESCO)
- Place (NUTS / ISO3166)
- Sector (NACE)
- Educational level (ISCED)
- ...



# DATA CLASSIFICATION – ...TECHNIQUES





# THANK YOU

## ALESSANDRO VACCARINO, TABULAEX

