

# Big Data for Labour Market Intelligence

Capacity development programme 2022

**Module 1: Technical training**

**Session 5**

Access to the database, data model and key features - Practical guidance

# Data production system

## Tunisia (04/2020 to 07/2022)

680,191 OJVs - > 175,203 deduplicated

## Ukraine general (04/2020 to 07/2022)

2,571,655 OJVs - > 1,304,262 deduplicated

## Georgia (04/2021-07/2022)

129,271 OJVs - 84,817 deduplicated

## Egypt (new)

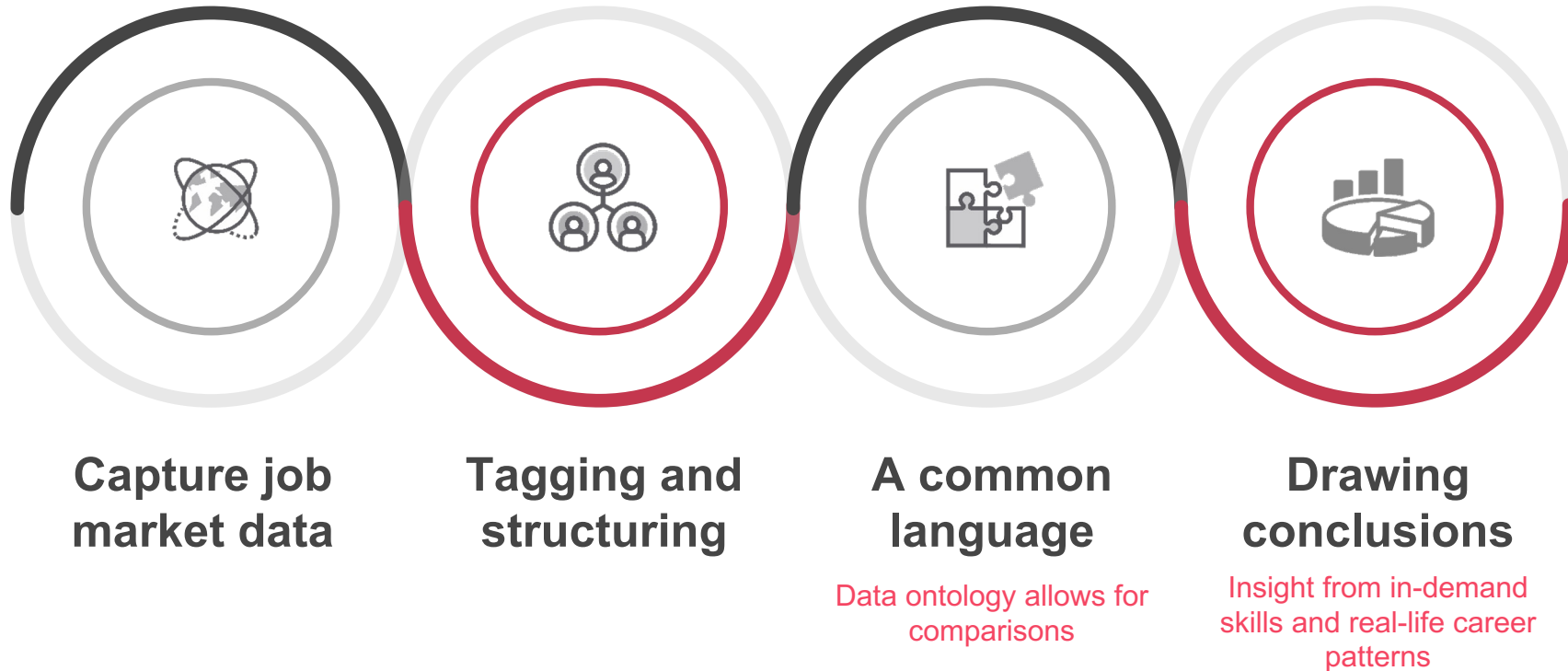
1,307,678 OJVs – 391,701 deduplicated

## Kenya (new)

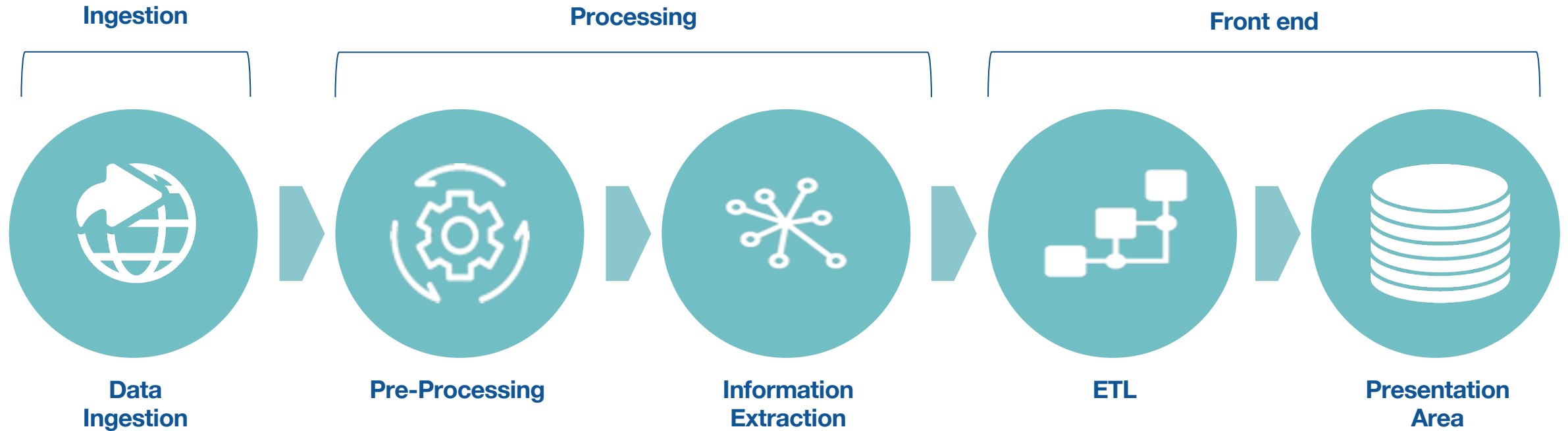
(collection started in september 2022)

# Collecting and decoding labor market data

Real-time job market data offer up-to-date insights not possible through traditional sources



# Overall Data Flow



# Data Ingestion

- **Goal:**
  - Collect massive amount of heterogeneous data from several unpredictable number of sources
- **Challenges:**
  - Guarantee data completeness and consistency
- **Approach:**
  - Develop a multi-technique framework (crawling, scraping, API) to fit different website characteristics: volume of vacancies, different languages, technology structure, non-invasive approach and policy agreements
  - Prevent data sources losses, via redundancy policies
  - Detect and collect metadata to improve information value
- **Features:**
  - To grant process governance in data Ingestion phase, we developed a monitoring system with a scheduling tool and an alerting module

# Data Pre-Processing

- **Goal**
  - Feed information extraction phase with proper data
- **Challenges**
  - Measure, monitor and increase Data Quality, to maximize completeness, consistency, complexity, timeliness and periodicity
- **Approach**
  - Develop a multi-phase pipeline, focused on:
    - Vacancy Detection: analyze website page to select only content referred to vacancies
    - Deduplication: detect duplicated vacancy posts to obtain a single vacancy entity
    - Date detection: identify release and expire dates through vacancy description analysis
    - Vacancy duration: method to define expire date, when not explicitly available
- **Features**
  - Guarantee Data Quality during all processing phases

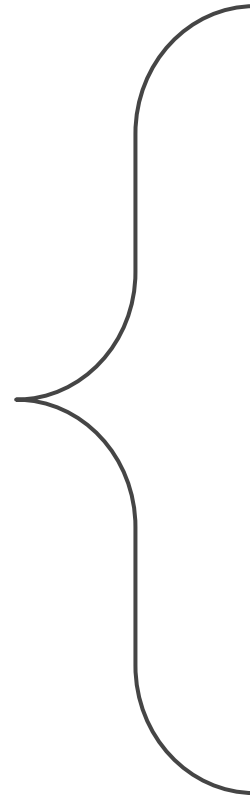
# Data Classification

- **Goal:**
  - Extract and structure information from data, to be provided to the presentation layer
- **Challenges:**
  - Handle massive amount of heterogeneous data written in different languages
- **Approach:**
  - Develop an adaptable framework, language dependent, tailored on different information features. Some relevant challenges:
    - Occupation feature classification: combined methods such as Machine Learning, Topic Modeling and Unsupervised Learning
    - Skill feature classification: another different combined methods, such as Text Analysis with corpus based or Knowledge based similarity
- **Features:**
  - Guarantee Explainable information extraction, logging classification methods and relevant features.

# Presentation Area



Presentation  
Area



Data Discovery  
Capabilities

Publish, Share  
and Collaborate

Self-service  
analytics and BI

Embedded  
Advanced  
Analytics

Visual-Based  
Investigation

Systems of  
Insight

Spatial location  
intelligence

Graph Discovery  
and investigation

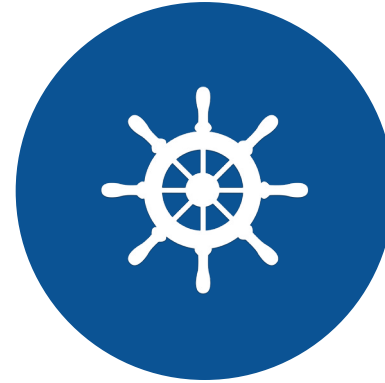
Data  
storytelling



# Identify users and navigation patterns



Data  
Scientists and  
Analysts



Decision  
Makers and  
Business Users

# Data Scientists and Analysts



Data Discovery  
Capabilities



Publish, Share and  
Collaborate



Machine-Learning  
Integration



Embedded  
Advanced Analytics

# Decision Makers and Business Users



Self-service analytics  
and BI



Visual-Based  
Investigation

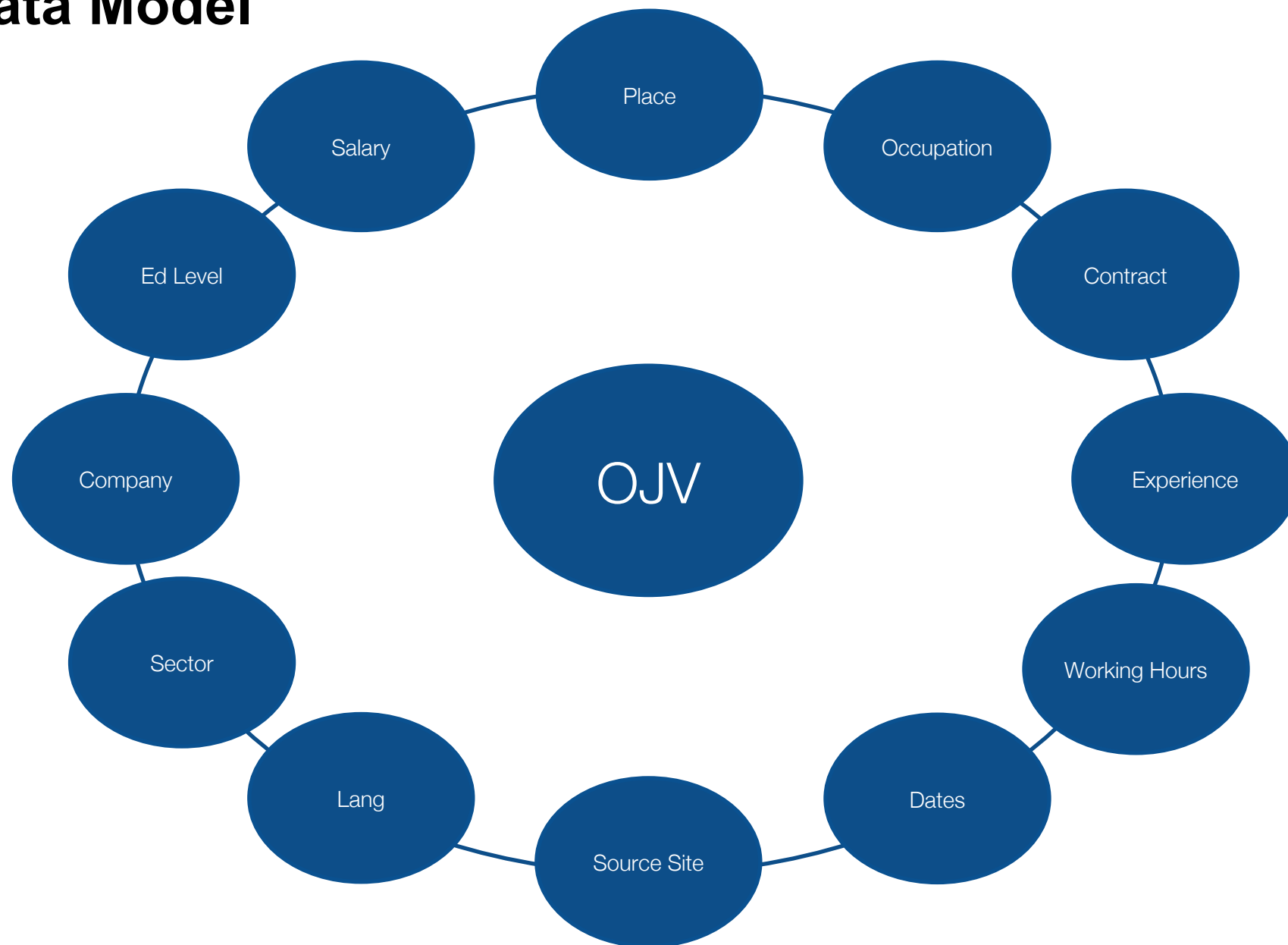


Governed Data  
Discovery



Data  
storytelling

# (Basic) Data Model

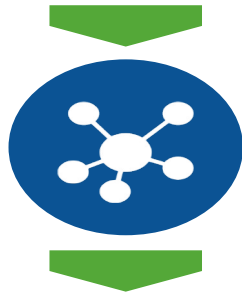


# Sector

Job Reference: 990-NHSE8576N  
Industry: Health  
Salary: 56,665 – 69,168 per annum  
Location: Leeds

NHS England leads the National Health Service (NHS) in England. We set the priorities and direction of the NHS and encourage and inform the national debate to improve health and care. We want everyone to have greater control of their health and their wellbeing, and to be supported to live longer, healthier lives by high quality health and care services that are compassionate, inclusive and constantly-improving...

Nace was present as structured field in OJV (valued as "Health", that matches ontology)



*Classification value:*  
86 - Human health activities

## Reference

NACE  
1° & 2°  
Level

# Contract

Labourer - Aylesbury  
Contract: Temporary (3 Month)  
Salary: £10 per hour

We are currently looking for a hard working an honest labourer in the Aylesbury area. You will be the main site labourer with duties including cleaning the site and helping trades out around the site. You can access this site with public transport, and has parking access if you....

Reference

Permanent  
Self Employment  
Temporary

Custom  
taxonomy

Contract was present as  
structured field in OJV (valued as  
"temporary", that matches  
ontology value)



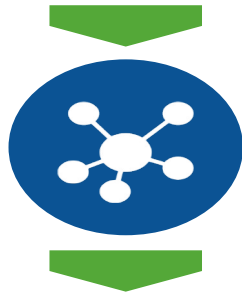
*Classification value:*  
Temporary

# Working Hours

Job Reference: 184-SS.GEN.38  
Department: Dementia  
Location: Bracken House, Chard

The Chard Older Persons Community Mental Health Team are actively seeking **to recruit a Part time** Band 5 Community Mental Health Nurse to assist with the Memory Assessment Service and Day Hospital. As part of an innovative Integrated Team you will be working closely with District Nursing, Integrated Rehab team and the Medical team as well as GP's, Adult Social care, Acute sector and Voluntary sector....

Working Hours was not present as structured field in OJV, but text contains reference to working hours ("part time") that matched ontology



*Classification value:*  
Part Time

Reference

Full time  
Part time

Custom  
taxonomy

# Educational Level

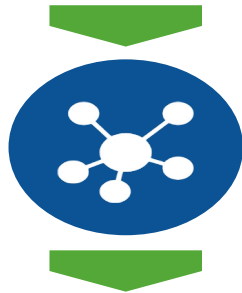
Role: Rolling Stock Team Leader  
Location: South London  
Salary: Approx. £47,500  
Education requirements: Associate Degree  
Experience: Less than 1 Year

The purpose, to lead the day to day activities to achieve timely stock delivery whilst ensuring that both relevant maintenance standards are achieved and passenger environment activities are enabled.  
Roles and responsibilities include but are not limited to: Daily delivery of the fleet into service, reliably and consistently To be part of the leadership team that delivers a cost effective and efficient maintenance ...

Reference

ISCED  
2011

Educational Level was present as structured field in OJV as "Education Requirements" (valued as "Associate degree", that matches ontology's alternate title)



*Classification value:*  
Bachelor or equivalent



# Salary

Role: Rolling Stock Team Leader  
Location: South London  
Salary: Approx. £47,500  
Education requirements: Associate Degree  
Experience: Less than 1 Year

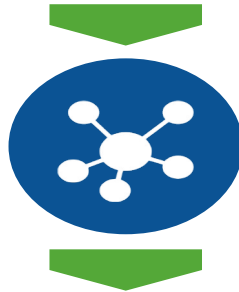
The purpose, to lead the day to day activities to achieve timely stock delivery whilst ensuring that both relevant maintenance standards are achieved and passenger environment activities are enabled.  
Roles and responsibilities include but are not limited to: Daily delivery of the fleet into service, reliably and consistently To be part of the leadership team that delivers a cost effective and efficient maintenance ...

Reference

13 levels

Custom  
taxonomy

Salary was present as structured field in OJV (valued as "£47,500 per Year" and converted to EUR currency)



*Classification value:*  
48.000 - 54.000 EUR Per Year

# Experience

Role: Rolling Stock Team Leader  
Location: South London  
Salary: Approx. £47,500  
Education requirements: Associate Degree  
Experience: Less than 1 Year

The purpose, to lead the day to day activities to achieve timely stock delivery whilst ensuring that both relevant maintenance standards are achieved and passenger environment activities are enabled.  
Roles and responsibilities include but are not limited to: Daily delivery of the fleet into service, reliably and consistently To be part of the leadership team that delivers a cost effective and efficient maintenance ...

Reference

8 levels

Custom  
taxonomy

Experience was present as structured field in OJV (valued as "Less than 1 Year", that matches ontology)



*Classification value:*  
Up to 1 year

# Place

Job Reference: 990-NHSE8576N  
Industry: Health  
Salary: 56,665 – 69,168 per annum  
Location: Leeds

NHS England leads the National Health Service (NHS) in England. We set the priorities and direction of the NHS and encourage and inform the national debate to improve health and care. We want everyone to have greater control of their health and their wellbeing, and to be supported to live longer, healthier lives by high quality health and care services that are compassionate, inclusive and constantly-improving...

Place was present as structured field in OJV (valued as "Leeds", that matches ontology)



*Classification value:*  
Leeds

## Reference

NUTS/ISO  
3166  
&  
LAU

# Occupation

## Unix Technician

In this role you will be responsible for these activities:

- o Install and support the server operating system, system management software and operating system utilities
- o Manage the operating system configuration
- o Manage file systems and print queues
- o Monitor and maintain operating system log files
- o Recommend operating system updates and configuration modification ...

Reference  
(ESCO 4<sup>th</sup> level)

ESCO  
v1.1  
ISCO08  
(4<sup>o</sup> DIGIT)

Machine Learning algorithm  
matched the correct  
Occupation, not present in  
ontology



*Classification value:*  
2522 - Systems administrators

# Skill

Are you an experienced Administrator, seeking your next contract in the Bristol area?  
My client is a large property maintenance specialist with an immediate opportunity for a Branch Administrator to join the team on an initial interim basis.  
The successful candidate will complete a range of **administration** tasks, including **answering incoming calls, liaising with contractors and raising invoices**.

Responsibilities:

- **Use the I.T systems** to provide an administration service in the preparation, processing and selection of estimates, bids and tenders
- **Ordering** of goods, materials and services to enable the requirements of contracts are met
- **Deal with internal and external communications** and record and or report information as necessary
- Ensure all necessary contract data, documentation and reports are accurate and produced on time
- Support Management in **meeting the business needs**.
- **Deal with Client / Customer queries and or communications** professionally and efficiently.

Requirements:

- Confident IT skills, **proficient in the use of MS Office**
- Excellent **communication skills** both written and verbal
- Must be an excellent organiser with proven **time management skills**
- ....

Reference



# Dataset



# Dataset – Model & Datastore

## Keywords

- Denormalized tables
- Columnar Store

# Dataset Model

2 Tables

## FT Document

1 Row for each:

- General\_ID □ Key
- OJV
- Source
- Place

Why? Because, for each OJV, we can detect a multi-place Vacancy

e.g. «Software Developer in London / Liverpool»

## FT Skill Analysis

1 Row for each:

- General\_ID □ Key
- OJV
- Source
- Place
- Skill

Why? Because, for each OJV, we can obviously detect multiple skills

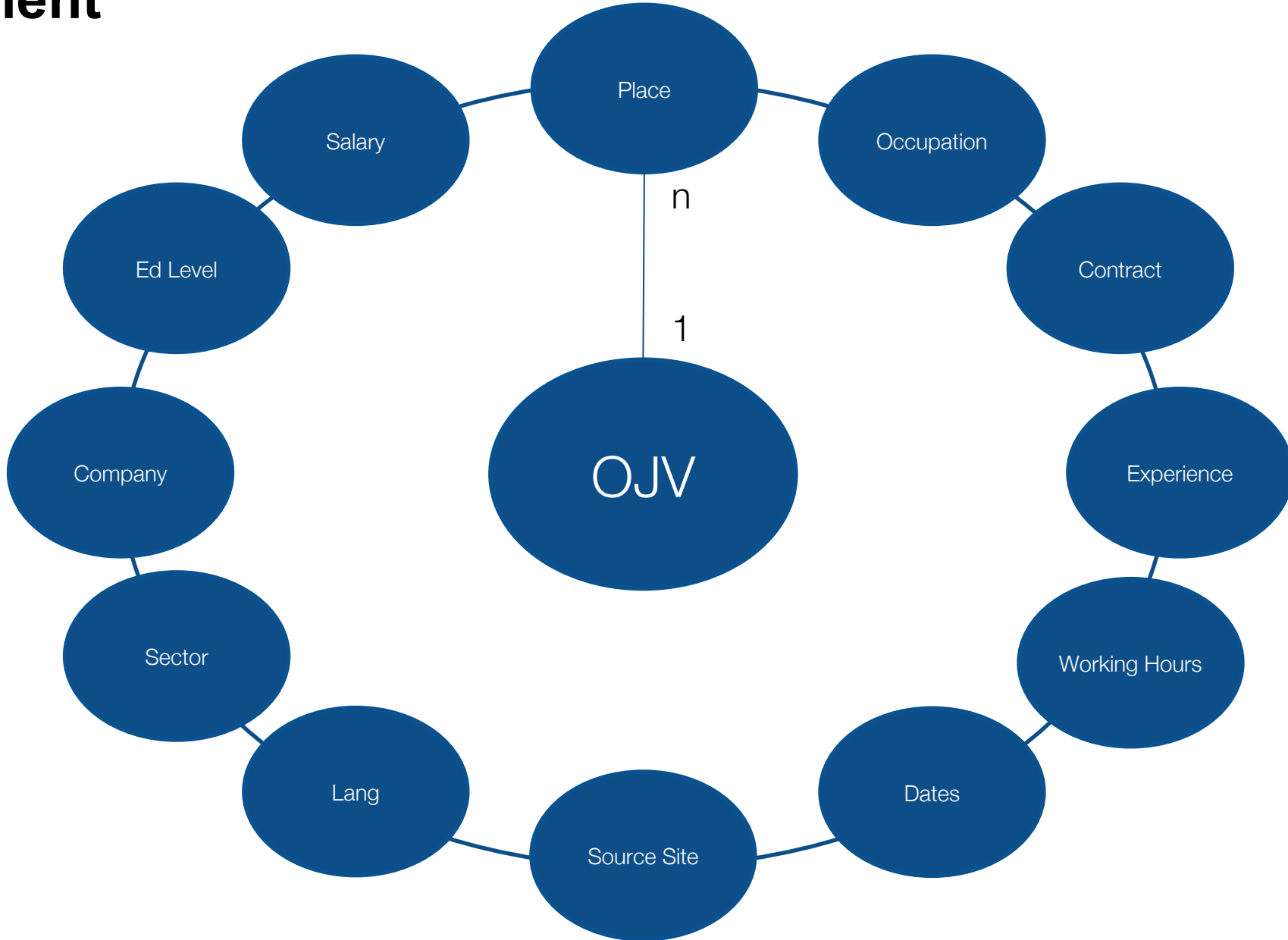
e.g. «Software Developer in London / Liverpool, with customer orientation culture, that speaks english and tolerates stress»



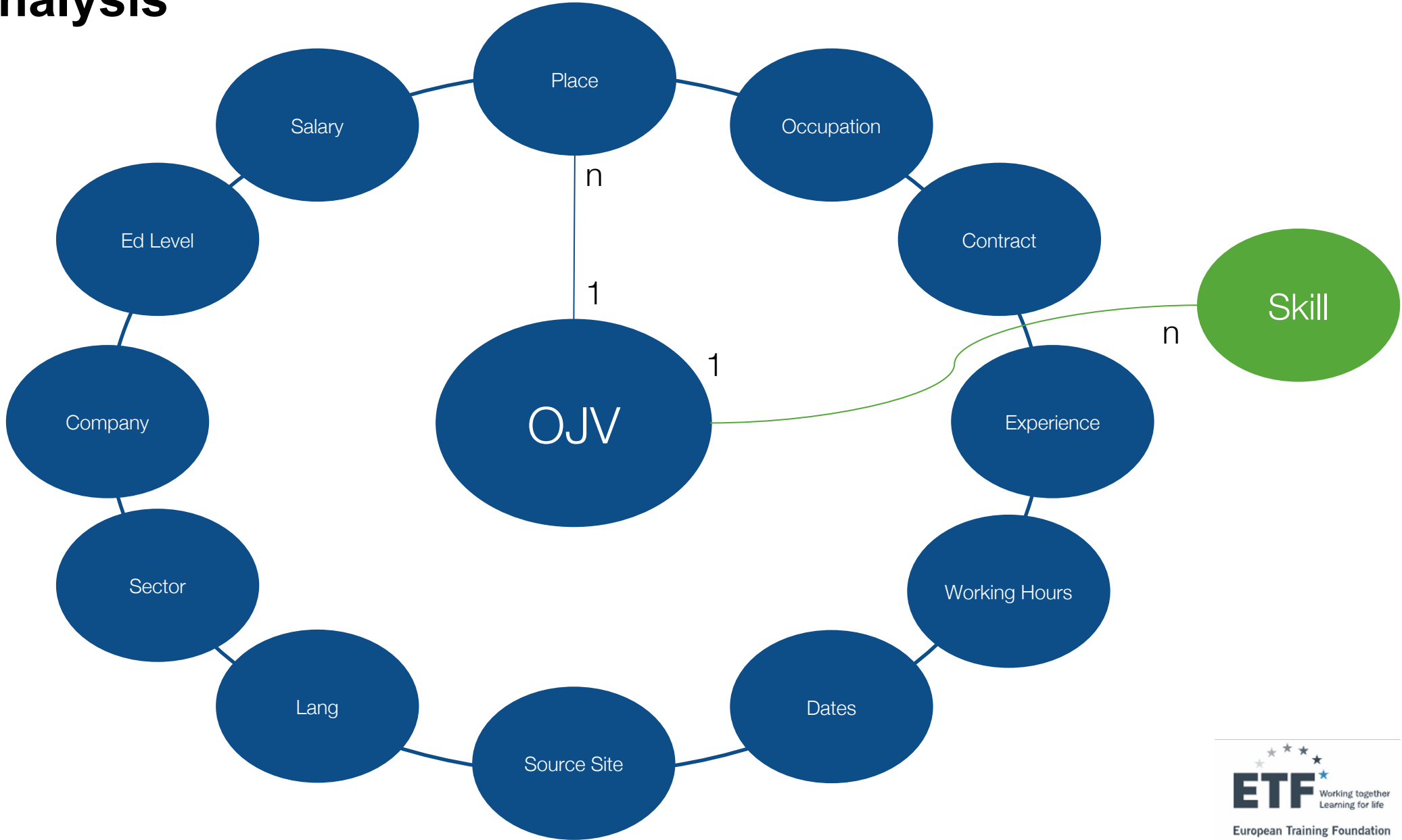
# Dataset Model

So, to have the number of job vacancies you always have to compute an unique count by General\_ID

# FT Document



# FT Skill Analysis



# Tools

