Module 2: Dissemination and analysis

Session 6
Role of ESCO and O*Net classifications in OJVs analysis.

Speaker: Mauro Pelucchi
16/11/2022
Data Scientist
<table>
<thead>
<tr>
<th>Role</th>
<th>Company</th>
<th>Location</th>
<th>Full-time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Science Analyst (Remote)</td>
<td>Yelp 3.4 ★</td>
<td>Remote</td>
<td>Full-time</td>
<td>A knack for communicating quantitative results to a broad audience through writing and compelling data visualization.</td>
</tr>
<tr>
<td>Machine Learning Research Engineer</td>
<td>Evolution Artificial Intelligence Remote</td>
<td>Remote</td>
<td>Full-time</td>
<td>You'll be putting state of the art deep learning technology into production use. Designing, developing and rigorously testing machine learning models.</td>
</tr>
<tr>
<td>Biostatistician - Global Biostatistics</td>
<td>IQVIA 3.8 ★</td>
<td>Remote</td>
<td>Full-time</td>
<td>Perform protocol development, sample size calculation, protocol and CRF review, and data management on database design and data.</td>
</tr>
<tr>
<td>Applied Scientist (Remote)</td>
<td>Yelp 3.4 ★</td>
<td>Remote</td>
<td>Full-time</td>
<td>Experience with data analysis/statistical software and packages (pandas/statsmodels/sklearn within Python, R, etc.).</td>
</tr>
<tr>
<td>Data Science Manager</td>
<td>Consortia</td>
<td>Remote</td>
<td></td>
<td>As Data Science Manager, you will lead and grow a world-class data science team.</td>
</tr>
<tr>
<td>Senior Analyst, BI &amp; Analytics</td>
<td>WorldRemit</td>
<td>Remote</td>
<td></td>
<td>Experience in an analytics or data science role. Strong SQL skills, able to handle large, complex data structures from multiple sources.</td>
</tr>
</tbody>
</table>
Labour Market Analysts

Interactive Data Analytics

Employment Agencies and Public Employment Services

Job Portals

Classified Ads Sites

Newspaper, Companies, University Job Placement

Web Scraper

Pre-Processing

Web Crawler

Information Extraction and Classification

Direct Access

Data Management and Presentation

Document store

Job Vacancies Classified on ISCO
Recognised NUTs
Other dimension (contract, sector, education, …)

Labour Market Analysts

ETF: Working together Learning for life
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.
Data Classification - An example

Junior Software Developer

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 API's, and third party integration.

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.

2512 – Software Developer

Skills: develop software, implement web based applications, problem solving, develop user experiences

Harwell, UK

Do we really need a classification standard?
The quality of big data analysis depends on the quality of the classification system. It's like the road to follow in a big maze.
Can you imagine some big data analysis collected so far?
### Most requested Skills in EU in ICT professionals

<table>
<thead>
<tr>
<th>Skill</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adapt to change</td>
<td>1,682,070</td>
</tr>
<tr>
<td>English</td>
<td>1,491,122</td>
</tr>
<tr>
<td>Project management</td>
<td>1,437,838</td>
</tr>
<tr>
<td>Use a computer</td>
<td>1,339,540</td>
</tr>
<tr>
<td>Teamwork principles</td>
<td>1,274,141</td>
</tr>
<tr>
<td>Work as a team</td>
<td>1,258,546</td>
</tr>
<tr>
<td>Foreign languages for international careers</td>
<td>1,250,600</td>
</tr>
<tr>
<td>Computer programming</td>
<td>1,243,285</td>
</tr>
<tr>
<td>Team building</td>
<td>1,099,985</td>
</tr>
<tr>
<td>Business ict systems</td>
<td>1,070,484</td>
</tr>
<tr>
<td>Think creatively</td>
<td>1,007,100</td>
</tr>
<tr>
<td>Customer service</td>
<td>995,272</td>
</tr>
<tr>
<td>Administer ict system</td>
<td>933,873</td>
</tr>
<tr>
<td>Create solutions to problems</td>
<td>882,993</td>
</tr>
<tr>
<td>Develop strategy to solve problems</td>
<td>857,275</td>
</tr>
<tr>
<td>Use microsoft office</td>
<td>811,946</td>
</tr>
<tr>
<td>Use object-oriented programming</td>
<td>769,825</td>
</tr>
<tr>
<td>Use software design patterns</td>
<td>735,207</td>
</tr>
<tr>
<td>Analyse software specifications</td>
<td>733,080</td>
</tr>
<tr>
<td>Database</td>
<td>711,456</td>
</tr>
</tbody>
</table>
Most relevant skill in digital occupations

- **Soft Skills**
  - teamwork principles
  - think creatively
  - analyse problems for opportunities
  - adapt to change
  - assist customers
  - use analytics
  - lead a team
  - project management
  - communication principles
  - show responsibility

- **Digital Skills**
  - analyse software specifications
  - computer programming
  - administer ICT system
  - database
  - ERP system
  - ABAP
  - SQL
  - web programming
  - Java
  - manage IT security

Source: WollyBI Data - Italy 2018
Estimated on 146,567 digital OJVs over 1,330,623 OJVs
Number of Web Job Vacancy by Programming Language

Languages:
- SQL
- Java
- PHP
- C#
- C++
- Python
- JavaScript
- Scala
- MATLAB

European Training Foundation
Working together Learning for life
Structured data vs Big data

**Structured data**
Purposefully collected and collated data which comes in neat, tidy structure. This is typically data from government statistical surveys, designed to ask explicit questions of targeted samples of specific audiences.

**Big data**
Extremely large scale data captured from some transactions rather than as a specific data collection exercise. This means harvesting job postings and worker profiles from different web-based sources.
Why use Big data?

We make data available through three primary channels

- **Granular**: big data provides coverage down to 4-digit industry and occupation, down to Local Area Units, with more detail possible.
- **Comprehensive**: big data is made available across all industry/occupation/area cells, covering all cases in the country.
- **Fresh**: job postings data is updated monthly
- **Integrated**: big data is presented as a single set without exceptions, and is made available through tools which allow access for decision-makers rather than only expert analysts
Online Profile

Skill

Job Posting

Employee Surveys

Person

Job

Task

Workplace

Enterprise

Company Accounts

Education Provider

Qualification

Employer Surveys

ETF: Working together Learning for Life
Taxonomies
The three main taxonomies

<table>
<thead>
<tr>
<th>ISCO08</th>
<th>NACE</th>
<th>ESCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>The International Standard Classification of Occupations is an International Labour Organization classification structure for organizing information on labour and jobs. It is part of the international family of economic and social classifications of the United Nations.</td>
<td>The Statistical Classification of Economic Activities in the European Community, commonly referred to as NACE, is the industry standard classification system used in the European Union. The current version is revision 2 and was established by Regulation No 1893/2006.</td>
<td>ESCO (European Skills, Competences, Qualifications and Occupations) is the European multilingual classification of Skills, Competences and Occupations.</td>
</tr>
</tbody>
</table>
The International Standard Classification of Occupations is an International Labour Organization classification structure for organizing information on labour and jobs. It is part of the international family of economic and social classifications of the United Nations.
Taxonomies are fully hierarchical

ESCO

ESCO (European Skills, Competences, Qualifications and Occupations) is the European multilingual classification of Skills, Competences and Occupations.

![ESCogram](chart.png)
ESCO Skills main groups

- **Abilities** - Enduring attributes of the individual that influence performance
- **Knowledge** - Organized sets of principles and facts applying in general domains
- **Skills** - Developed capacities that facilitate learning or the more rapid acquisition of knowledge
- **Work Activities** - General types of job behaviors occurring on multiple jobs
- **Work Styles** - Personal characteristics that can affect how well someone performs a job.
- **Technology Skills & Tools** - Information technology and software skills or machines, equipment, and tools essential to the performance of an occupational role
Occupation taxonomy

- Group similar roles together to enable easier analysis.
- A role is defined by a distinct mix of knowledge, skills, abilities, activities, and tasks.
- They may include many variations of similar job titles.
Occupation taxonomy

Clustering: Occupations represent clusters of Job Titles and Skills

ADB Shell Engineer
Android Developer
JavaScript, XML, Struts
Senior Engineer – Android Mobile Dev
Linux Mobile

= 2512 Software Developer
Occupation taxonomy are building blocks

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5821</td>
<td>Publishing computer games</td>
</tr>
<tr>
<td>5829</td>
<td>Other software publishing</td>
</tr>
<tr>
<td>6110</td>
<td>Wired telecommunications activities</td>
</tr>
<tr>
<td>6120</td>
<td>Wireless telecommunications activities</td>
</tr>
<tr>
<td>6130</td>
<td>Satellite telecommunications activities</td>
</tr>
<tr>
<td>6190</td>
<td>Other telecommunications activities</td>
</tr>
<tr>
<td>6201</td>
<td>Computer programming activities</td>
</tr>
<tr>
<td>6202</td>
<td>Computer consultancy activities</td>
</tr>
<tr>
<td>6203</td>
<td>Computer facilities management activities</td>
</tr>
<tr>
<td>6209</td>
<td>Other information technology and computer services</td>
</tr>
<tr>
<td>6311</td>
<td>Data processing, hosting and related activities</td>
</tr>
<tr>
<td>6312</td>
<td>Web portals</td>
</tr>
</tbody>
</table>

Digital cluster
Pillars

Hierarchy

Pillars

Digital
AI
Pillar n
Results on Pillars Digital skills

Digital

DDH1Digital

1.000

Abilities

Knowledge

Skills

Technology Skills & Tools

Work Activities

Work Styles

Esco1Digital

1. Managers

2. Professionals

3. Technicians and associate professionals

4. Clinical support workers

5. Service and sales workers

6. Skilled agricultural, forestry and fishery workers

7. Craft and related trades workers

8. Plant and machine operators, and assemblers

9. Elementary occupations

ContryDigital

CEDEFOP
European Centre for the Development of Vocational Training

ETF
European Training Foundation
Results on Pillars: Soft skills
Using ONET data

- ONET data is produced by the US Government Department of Labor (v25.2). A range of occupations are measured against the ONET Content Model using an ongoing survey process. The Content Model is described here: [https://www.onetcenter.org/content.html](https://www.onetcenter.org/content.html)
  - Knowledge, Skills, Tasks and Work Conditions are measured by an Importance scale (IM) from 1 (Not important) to 5 (Extremely important)
  - Knowledge and Skills are also measured by a Level scale (LV) from 0 to 7.
  - Tasks are measured by a Relevance scale (RT) which is a percentage scale.
  - Work Context is measured by Context scale (CX) from 1 to 5.
O*Net-SOC Example

O*NET SOC takes things one step farther

<table>
<thead>
<tr>
<th>Major Group</th>
<th>Minor Group</th>
<th>Broad Occupation</th>
<th>Detailed Occupation</th>
<th>Detailed O*NET-SOC</th>
<th>SOC or O*NET-SOC 2019 Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Management Occupations</td>
</tr>
<tr>
<td>11-1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Top Executives</td>
</tr>
<tr>
<td>11-1010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chief Executives</td>
</tr>
<tr>
<td>11-1011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chief Executives</td>
</tr>
<tr>
<td></td>
<td>11-1011.03</td>
<td></td>
<td></td>
<td></td>
<td>Chief Sustainability Officers</td>
</tr>
</tbody>
</table>
Hierarchical Comparison

O*NET SOC takes things one step farther

<table>
<thead>
<tr>
<th>Major Group</th>
<th>Minor Group</th>
<th>Broad Occupation</th>
<th>Detailed Occupation</th>
<th>Detailed O*NET-SOC</th>
<th>SOC or O*NET-SOC 2019 Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Management Occupations</td>
</tr>
<tr>
<td>11-1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Top Executives</td>
</tr>
<tr>
<td>11-1010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chief Executives</td>
</tr>
<tr>
<td>11-1011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chief Executives</td>
</tr>
<tr>
<td>11-1011.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chief Sustainability Officers</td>
</tr>
</tbody>
</table>
The ‘Occupational Information Network’ (O*NET) is a free online database that contains hundreds of occupational definitions developed during the 1990s under the sponsorship of the US Department of Labor/Employment and Training Administration (USDOL/ETA).

USDOL/ETA describes the O*NET as: a database of occupational requirements and worker attributes. It describes occupations in terms of the **skills** and **knowledge** required, how the work is performed, and typical work settings.
About O*NET-2

- The **Content Model** is the conceptual foundation of O*NET.
- The Content Model is organized into six major domains and provides a framework that identifies the most important types of information about work and integrates them into a theoretically and empirically sound system.
- **O*NET OnLine** is a tool for career exploration and job analysis, in this tool you can browse O*NET data according to the O*NET Content Model.
At its first level O*NET includes the following Data Descriptors:

- **Abilities** - Enduring attributes of the individual that influence performance.
- **Interests** - Preferences for work environments and outcomes.
- **Knowledge** - Organized sets of principles and facts applying in general domains.
- **Skills** - Developed capacities that facilitate learning or the more rapid acquisition of knowledge.
- **Work Activities** - General types of job behaviors occurring on multiple jobs.
O*NET OnLine categories - 2

- **Work Context** - Physical and social factors that influence the nature of work.
- **Work Styles** - Personal characteristics that can affect how well someone performs a job.
- **Work Values** - Global aspects of work that are important to a person's satisfaction.
- **Technology Skills & Tools** - Information technology and software skills or machines, equipment, and tools essential to the performance of an occupational role.
Real-Time Labour Market Information System on Skill Requirements

Continuously evolving Labour Market

Number of job vacancies collected
2,624,510

Number of job vacancies deduplicated
1,348,241

Number of unique Vacancies by Web Source

Distribution by Release Date (date of publication of the OJV)