



## Real-time labour market information and skill requirements:

### Setting up the infrastructure for EU system

Expert workshop

20-21 March 2018, Milano, Italy

#### General information

<b>Workshop venue</b>	CRISP, University Milano-Bicocca, U7 Building, 2 <sup>nd</sup> floor, room 2062, Piazza Ateneo Nuovo 1, 20126 Milano, Italy
<b>Organisers</b>	Cedefop – European Centre for the Development of Vocational Training, Thessaloniki, Greece  CRISP - Interuniversity Research Centre on Public Services University of Milano-Bicocca, Italy  ESS-net – network of European Statistical Systems
<b>No. of participants</b>	approx. 30
<b>Participants' profile</b>	Experts in big data statistics, data scientists, labour market analysts
<b>Working language</b>	English

#### Introduction

Cedefop is setting up a pan-European system for gathering and analysing information from online vacancies across all EU countries. To ensure the highest possible quality of the system, Cedefop is bringing together members of the ESSNet (work package 1: Webscraping/Job vacancies) and independent researchers to discuss and fine-tune the methodology as well as various algorithms and rules for extracting vacancies and classifying the information they contain.

#### Workshop rationale and objectives

The workshop will provide a platform for presenting and discussing the proposed methodology as well as for sharing the experience and knowledge of various experts in the meeting with the aim to fine-tune various algorithms and arbitrary set rules, related to:

- Data ingestion (web scraping, crawling, APIs download)
- Extraction, de-duplication and expansion of vacancies
- Classifications of occupations
- Extraction of skills

The ultimate objective of the workshop is to find a joint agreement on the methodology, including the various algorithms and rules for extracting vacancies and classifying the information obtained.

# Agenda

Tuesday, 20<sup>th</sup> March 2017

**8:30– 09:00** *Arrival & Registration*

Chair: Vladimir Kvetan

**09:00 – 09:30** **Welcome and introduction to the workshop**

*Vladimir Kvetan (Cedefop)*

*Nigel Swier (ESS-net)*

**09:30-11:00** **Session 1 - Current state of the art of Cedefop's project**

*Jiri Branka (Cedefop)* – Landscape of EU online vacancies

*Emilio Colombo (CRISP)* – Current state of the art and the next steps

Topics to be discussed:

- Confronting the findings against ESS net experts' knowledge
- Clarifying aims and goals of the project with EC, ESS-net and ESTAT
- Agreeing on overall feasibility of the plan, identification of critical points

**11:00-11:30** *Coffee break*

**11:30 – 13:00** **Session 2 - Data ingestion**

*Mario Mezzanica (CRISP)* - Overview of methodological and technical approach

*Ettore Colombo & Matteo Fontana (Tabulaex)* - Data ingestion process

Topics to be discussed:

- Data ingestion techniques
- Source selection strategy
- Expected volumes and preliminary data

**13:00 – 14:00** *Lunch break*

Chair: Jiri Branka

**14:00 – 15:00** **Session 3 - Data pre-processing**

*Mauro Pelucchi & Alessandro Vaccarino (Tabulaex)*: Data pre-processing

Topics to be discussed:

- Data processing pipeline
- Vacancy detection with machine learning techniques
- Deduplication issues
- Date detection and vacancy's duration

**15:00 – 16:00** **Session 4 Classifying enterprises by economic activities**

Short presentations by ESS net colleagues (Thomas Delclite (BE), Maxime Bergeat (FR), Fero Hajnovic (UK))

	<p>Topics to be discussed:</p> <ul style="list-style-type: none"> <li>• Methods for extracting economic activity from jobs ads</li> <li>• Assessment of the algorithms (accuracy, speed, applicability)</li> </ul>
<b>16:00-16:30</b>	<b>Coffee break</b>
<b>16:30 – 17:30</b>	<p><b>Session 5 Text processing and information extraction</b></p> <p><i>Mauro Pelucchi, Alessandro Vaccarino (Tabulaex) &amp; Fabio Mercorio (CRISP):</i> Text processing and information extraction</p> <p>Topics to be discussed:</p> <ul style="list-style-type: none"> <li>• Focus on text processing and multi-language environment</li> <li>• Classification processes, feature extraction and machine learning</li> <li>• Focus on occupation’s categorization</li> <li>• Focus on skill’s categorization</li> <li>• Preliminary results</li> <li>• How to analyse and assemble skills</li> </ul>

## Wednesday 21<sup>st</sup> March 2017

<b>08:45 – 09:00</b>	<b>Arrival</b>
	<b>Chair: Vladimir Kvetan</b>
<b>09:00 – 10:00</b>	<p><b>Session 6 - Data quality</b></p> <p><i>Nigel Swier (ONS/ESSNet)</i> A quality framework for on-line job vacancy data?</p> <p>Topics to be discussed:</p> <ul style="list-style-type: none"> <li>• How do we evaluate and measure quality from collection to final output?</li> <li>• What kind of quality information/metrics should be made available to end users?</li> </ul> <p><i>Piergiorgio Lovaglio (CRISP):</i> Data significance in OJV – Preliminary outcomes</p>
<b>10:00 – 11:00</b>	<p><b>Session 7 - Data presentations and analysis</b></p> <p><i>Matteo Fontana (Tabulaex):</i> General framework for data presentation</p> <p><i>Alena Zukersteinova (Cedefop):</i> Possible data applications and plans for early data release</p> <p>Topics to be discussed:</p> <ul style="list-style-type: none"> <li>• Assessment of feasibility and value added of the proposal</li> <li>• Ideas for further developments</li> </ul>
<b>11:00 – 11:30</b>	<b>Coffee break</b>
<b>11:30-12:30</b>	<b>Session 8 – Validation exercise</b>

	Interactive session prepared and facilitated by Cedefop
<b>12:30 – 13:00</b>	<b>Conclusions &amp; final remarks</b>
<b>13:00</b>	<b>End of Cedefop workshop</b>
<b>13:00 -14:00</b>	<b><i>Lunchbreak</i></b>

## Contact details

### **Cedefop**

Vladimir Kvetan  
 PO Box 22427, 55102 Thessaloniki, Greece  
 Phone: +(30) 2310 490 194  
 Fax: +(30) 2310 490 117  
 Email: [vladimir.kvetan@cedefop.europa.eu](mailto:vladimir.kvetan@cedefop.europa.eu)  
 Cedefop homepage: [www.cedefop.europa.eu/EN/](http://www.cedefop.europa.eu/EN/)

### **CRISP**

Emilio Colombo, Mario Mezzanica and Silvia Dusi  
 University of Milano-Bicocca (IT)  
 Phone: +39 0264482170  
 Email: [emilio.colombo@unicatt.it](mailto:emilio.colombo@unicatt.it), [mario.mezzanica@unimib.it](mailto:mario.mezzanica@unimib.it), [silvia.dusi@unimib.it](mailto:silvia.dusi@unimib.it)  
 CRISP homepage: [www.crisp-org.it](http://www.crisp-org.it)

### **ESS –net**

Nigel Swier  
 Office for National Statistics (ONS)  
 Government Buildings, Cardiff Road, Newport South Wales (UK), NP10 8XG  
 Phone:  
 Email: [nigel.swier@ons.gov.uk](mailto:nigel.swier@ons.gov.uk)  
 ONS homepage: [www.ons.gov.uk](http://www.ons.gov.uk)