



# WPB: Online Job Vacancies

## ESSNet Big Data II

Implementation kick-off

Tomaž Špeh

Vienna, 3.12.2018

# Agenda

- Introduction
- WP1 summary
- WPB description

# Online Job Vacancies WPB

- To identify and develop statistical estimates, methodology and prototypes in the statistical theme of OJV
  - at the level of individual NSI and at the level of the ESS
- BG, CH, DARES, DE, IT, LT, PL, RO, SE, SI, UK
- CEDEFOP
- Cooperation with WPC and WPF
- M1 – M24
- Physical meetings M2, M8, M20, BDES M25
  - Non-physical meetings
  - Communication and supporting arrangements
- Participation to relevant Eurostat Working groups and Task Forces

# WP1 Summary

- ESSnet Big Data, February 2016 - May 2018
- Demonstrate by concrete estimates approaches (techniques, methodology etc.) how to produce statistical estimates in the domain of job vacancies
  - Under which conditions these approaches can be used in the ESS.
- Explored a mix of sources
  - job portals, job adverts on enterprise websites, and job vacancy data from third party sources.
- Carried out by UK, DE,GR, IT, SE, SI, BE, DM, FR(DARES), PT
- The aim was not to develop a system suitable for production.

# WP1 OJV Data Use Cases

- Improving current job vacancy statistics
  - to produce high frequency statistics in near real-time
  - Contains additional information on occupations, skills, location information
  - reduce the number of surveys to reduce collections costs
- Classifying data from text descriptions (ESCO, NACE)
- New statistics on aspects of the labour market
  - International advertised jobs
- New job titles
- Measuring OJV coverage
- Time series analysis
- Data driven analysis

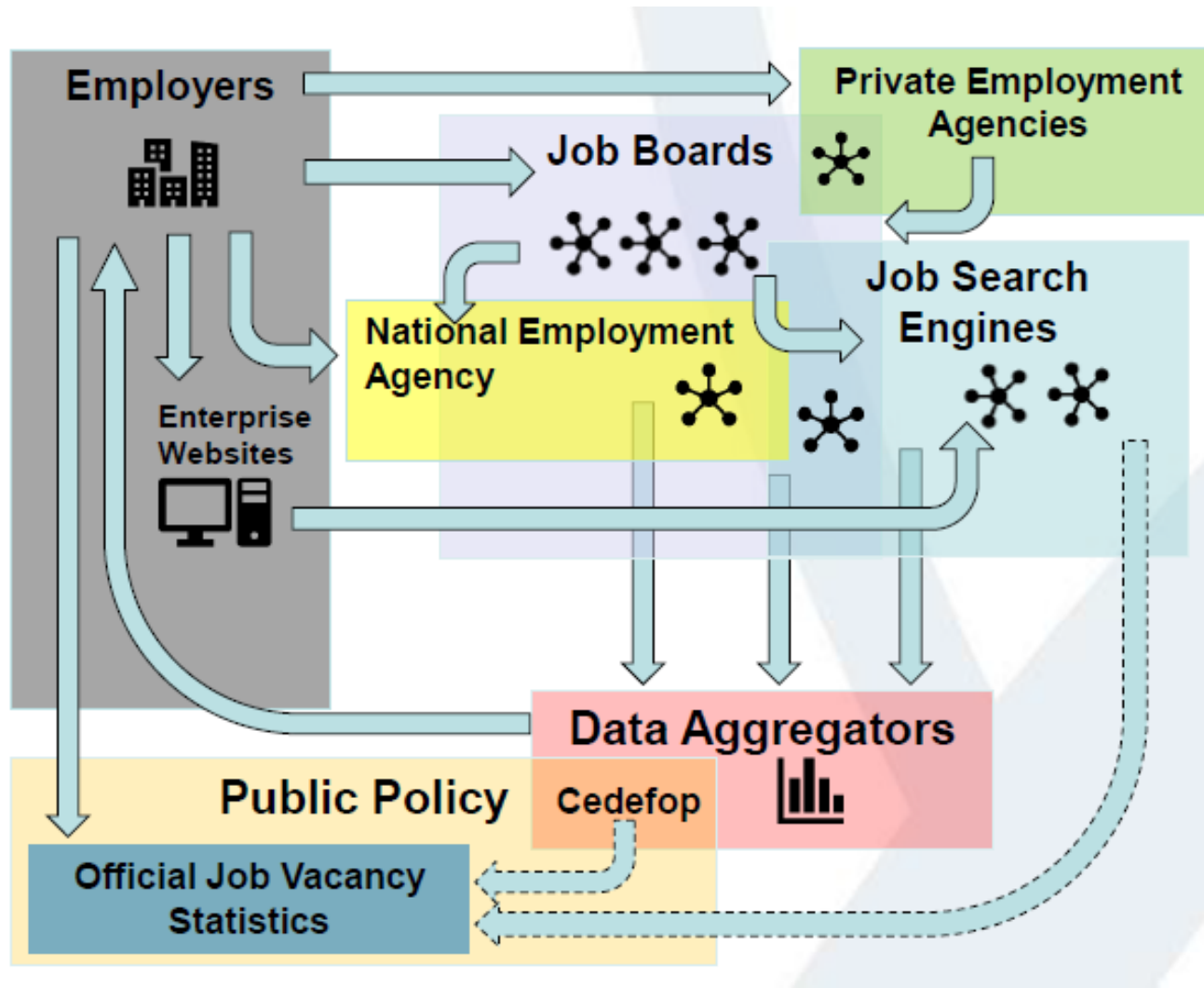
Table 1: Differences between the Official survey based estimates of job vacancy statistics

Dimension	Official estimates (JVS based)	OJV Data
Frequency	Quarterly	Real-time
Industry Sector	Yes	Yes
Enterprise size	Yes	Yes
Job title / Occupation / Skills <sup>6</sup>	No	Yes
Sub-national	No	Yes
National totals (estimates)	Yes	No

# WP1 Data Access

- Study the technical aspects of job portals and evaluate legal aspects.
- Direct web scraping
  - Point and click web scraping, Programmatic web scraping, Web scraping enterprise websites
- Arranged access
  - CEDEFOP, National Employment Agencies (NEAs), Private Job Portal Owners, Commercial suppliers of OJV data
- Evaluate the role of business and administrative registers to support the collection and quality assurance of web scraped data.
  - Evaluate the availability of information on the enterprise
  - possible use for linkage with business registers.

# WP1 OJV data landscape



# WP1 Data Access by Country

Country	Direct Web scraping		Agreed Access			
	Enterprise websites	Job Portals	National Employment Agency	Private Employment Agencies	Other data aggregators	
Germany		Yes		Yes		Yes (CEDEFOP)
Greece		Yes				
Slovenia	Yes	Yes	Yes	Yes	Yes	
Sweden			Yes	Yes		
United Kingdom	Yes	Yes	Yes			Yes (CEDEFOP, Burning Glass)
France		Yes		Yes		
Belgium	Yes	Yes		Yes		
Portugal		Yes				



# WP1 Data Handling

- Develop data processing steps to transform semi-structured web data into a structure suitable for analysis.
- Data cleaning, correction of formatting errors, evaluation and treatment of missing data.
- De-duplication of job adverts both within and across job portals.
- Text Pre-processing
  - Text standardization, Stop word removal, White/black lists, Lemmatization
- Text Analysis and Classification of data
  - e.g. occupation and geography coding.
- Flow to Stock transformation

# WP1 Methodology

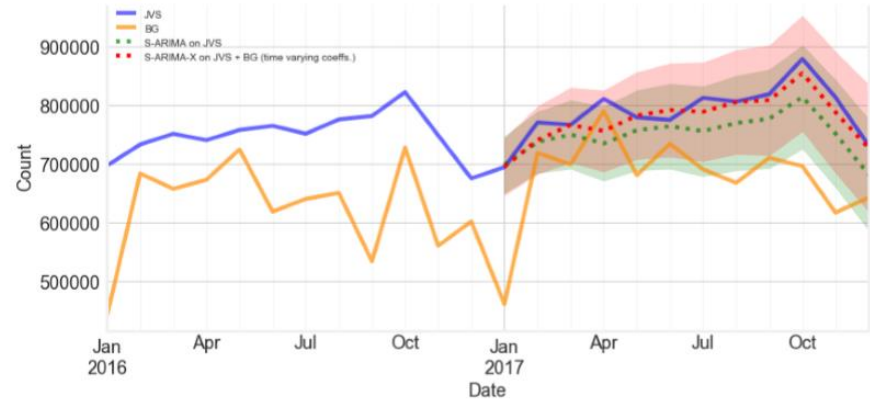
- Methodological challenges in using OJV data to produce official statistics
  - issues around the quality of OJV data, what these represent, and how they compare with official estimates produced from the JVS.
- Initial quality assessment of job portal data
  - assessment of job adverts advertised by agencies versus direct employers
  - evaluation of coverage by job portal/employment sector/ industry sector
- Qualitative assessment of the information available.
  - job title, occupation, economic activity, location, etc.
- Measuring Coverage
  - Micro-level comparisons, Aggregate comparisons, Measuring use of advertising channels via the JVS.
- Matching and linking
- Time series analysis
- Visualise results to show potential of the methods.

# WP1 Statistical Outputs

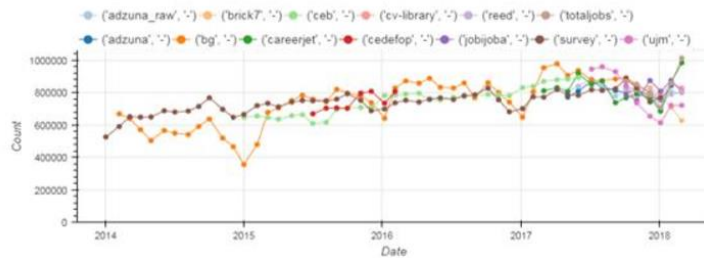
**Table 3: Experimental on-line job vacancy statistics for Slovenia**

Estimates	Type	28 August 2017 (Q3)	30 November 2017 (Q4)
Detected job ads for quarter	Stock	6849	6327
Official JVS estimate	Stock	17221	15243
Available in reference month	Stock	3542	4493
Available on reference day	Stock	1368	1285
Newly available on reference month	Flow	1984	2115
Newly available on reference day	Flow	123	76

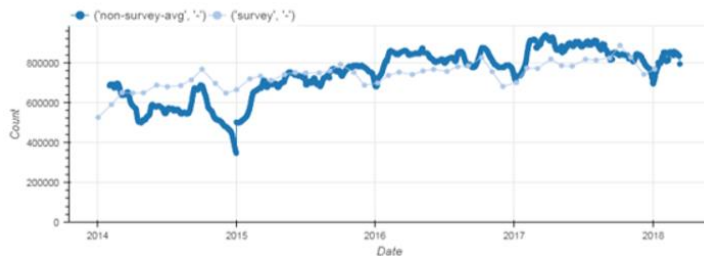
**Figure 5: Nowcasts based on the S-ARIMA-X time series model.**



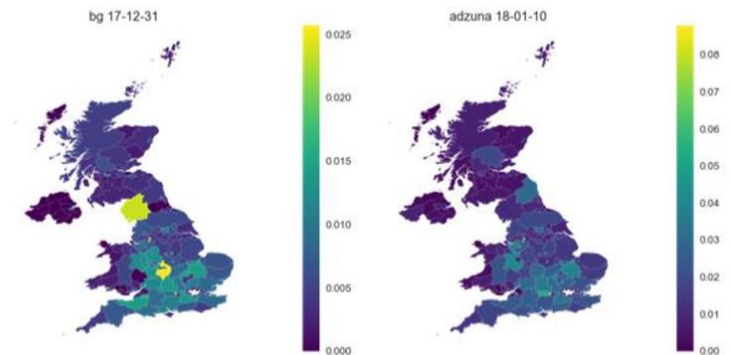
**Figure 4a: Time series of the total JV counts, averaged per month (scaled to the JVS scale).**



**Figure 4b: Time series of the JVS and daily average of the online sources (scaled to the JVS).**



**Figure 6: Number of job vacancies as a proportion of working age population<sup>28</sup>**



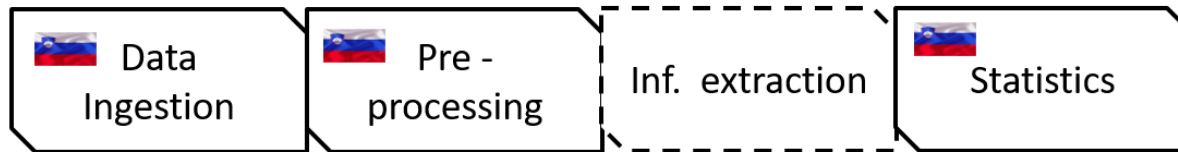
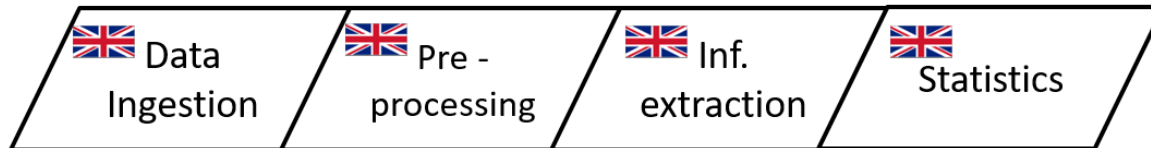
# WP1 Challenges

- Coverage
  - Not all jobs are advertised on-line. Coverage is incomplete and not representative.
  - Duplication
- There is no definitive source of OJV data
  - National Employment Agencies, Job portals, Enterprise websites, Data aggregators
- Complex information extraction
  - Much OJV data is unstructured. Text processing and analysis is required to extract useful information.
- Official statistics definitions of a job vacancy
  - Some job ads are not within the scope
  - Does not correspond directly to the concept of a live job ad
- The specific job vacancy data landscape varies between countries

# WP1 Key Conclusions

- Agreed access arrangements are generally better than direct web scraping
- OJV data cannot replace the Job Vacancy Survey
- OJV data does not correspond to target concepts and only measures part of the labour market.
- A successful collaboration with CEDEFOP is essential.
  - Common supporting infrastructure to support the adoption of OJV data in the ESS during the ESSnet II and beyond.
  - How do we get the best possible quality data for official statistics purposes?

# WP1 Pipelines



# WPB Objectives

- To identify and produce statistical estimates in the statistical theme of online job vacancies
- To develop and test the methodology, recommendations, specifications and statistical software to facilitate their integration into production
  - at the level of individual NSI and at the level of the ESS
- Based on approaches, methodologies and recommendations developed during the pilot phase of the project

# Task 1 - Methodological Framework

- Definition of the conceptual production processes at national level and at the level of the ESS
- Development and testing of fully functional prototypes
  - Including methodology for collecting, processing and analysis OJV data
  - Cooperation with CEDEFOP
- Developing and evaluating scenarios for linking to the business register and established statistical classifications
- Addressing issues related to sustainability of data sources, as well as data sharing between NSIs
- Cooperation with WPC in the development of the ESS web scraping policy



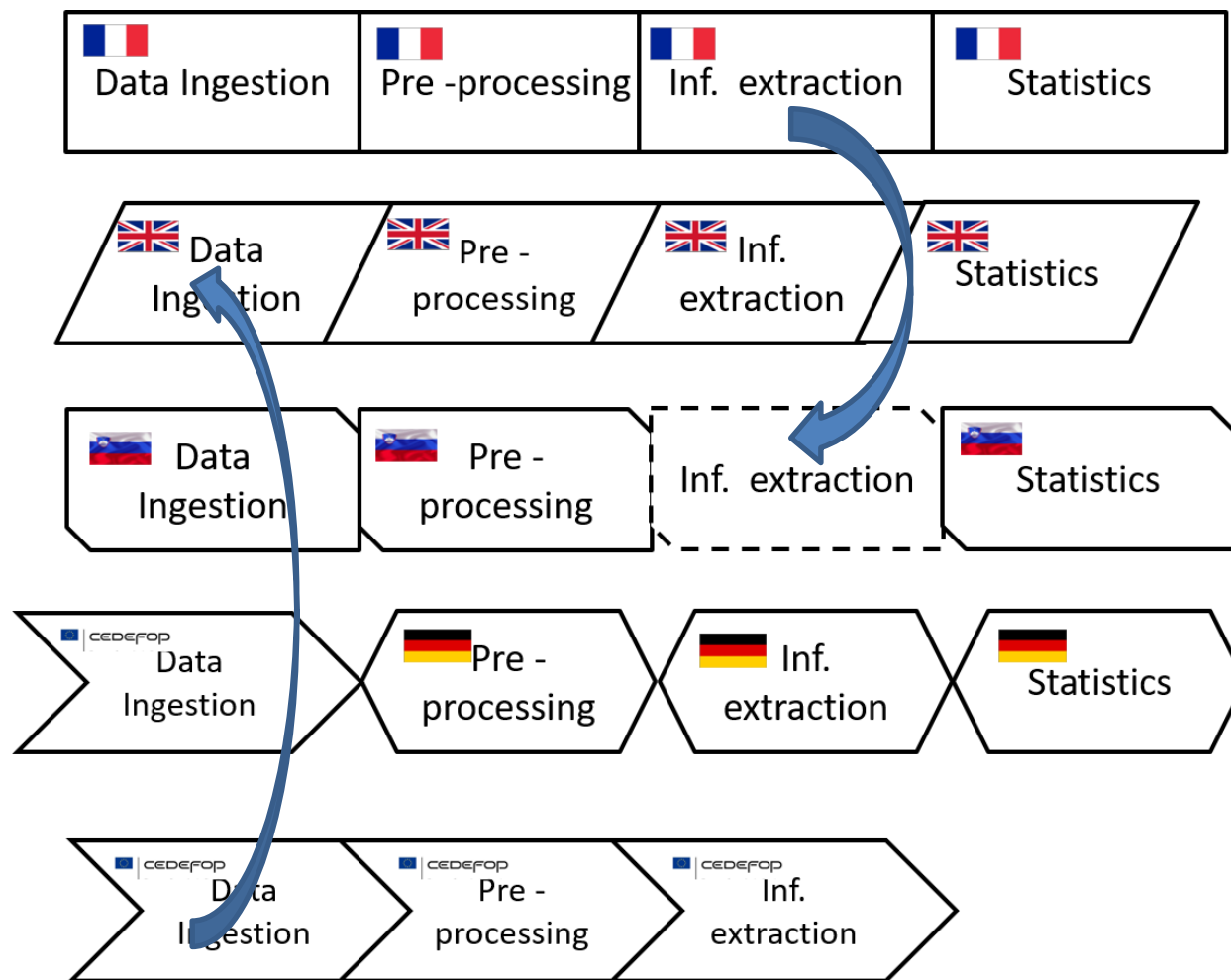
# CEDEFOP

- European Centre for the Development of Vocational Training
  - Online Labour Market Information
  - EC priority on jobs, growth and employment A new skills agenda for Europe
- Successful feasibility study
  - Early data release in spring 2019, fully fledged system in 2020
- Occupations, Geo-location of jobs, Skills and other job requirements, Time dimensions...
- Opportunity for partnerships
  - Economies of scale, Coordination and administration, Serving national and European needs
  - OJV statistics, Skills statistics, Flash estimates

# Task 2 - Statistical output

- Identify and produce statistical indicator(s) in the field of job vacancies statistics using OJV data
- Quality assessment of statistical indicator(s)
- Publication of indicators on OJVs which can supplement existing JV statistics as experimental statistics
- Investigation of other possible statistical indicator(s) beyond job vacancies statistics which could be produced with OJV data

# WPB Common conceptual architecture



- Integrate, replicate, share

# Task 3 - Implementation requirements

- Definition of the implementation requirements of prototypes in the relevant statistical production processes at European and national level
- Outline of the architecture, processes and infrastructure
  - The issue of common infrastructure
- Interaction with WPF on designing and adopting application and information architectures (WPF Task 1) and providing input on WPB solution architectures.
- Definition of the necessary material for the integration process
  - transfer of knowledge, methodologies, IT infrastructure needed, toolbox of methods, software development, testing, and maintenance.

# WPB Deliverables & Milestones

- B1 (M12) Interim technical report
- B2 (M12) Methodological Framework V.1
- B3 (M24) Methodological Framework V.2
- B4 (M24) Report on the statistical output, required quality and definition of the necessary metadata at European and national level
- B5 (M24) Technical report on the implementation requirements of prototypes in the relevant statistical production processes at European and national level
  
- BM1 (M9) Report on the WP meeting mid 2019
- BM2 (M20) Report on the WP meeting mid 2020