

Measuring the size and characteristics of the demand for workers is not an easy task. Increasing non-response and under-reporting in business sample surveys may introduce bias in the job vacancy estimates. The limited scope of job vacancy surveys by official statistics does not allow to provide estimates of the demand for skills and competencies or to cover all types of work contracts (e.g. contracts for specific work, contracts of mandate, business-to-business contracts). Administrative data and the Internet (the so-called "found data") cover thousands of job ads at any moment that are full of valuable information. However this information take an unstructured form (e.g. text) and requires the application of machine learning and natural language processing (ML/NLP) algorithms. Direct use of these data as sources of statistics is limited due to the fact that Internet data, although detailed, may be selective and do not conform to official classifications. Plenty of applications of job ads for describing the labour demand and employer – job seeker search and matching use simple error correction methods (e.g. post-stratification) while additional errors resulting from imperfect ML/NLP methods are ignored. Such a naive approach may lead to biased estimates, and thus incorrect policy recommendations.

**The main objective of the project is to develop new methods to measure the size and characteristics of the job vacancies (including flexible work contracts) based on survey, online and administrative data, and apply it to analyse fast changes observed in the vacancy market in the latest years, including the COVID-19 recession, refugees' inflow and technological advancements (e.g. artificial intelligence), as well as the job seeker – employer search and matching.**

**The novelty of the project results from:** 1) proposing a new approach to estimate the total number of vacancies based on solely on non-statistical data sources; 2) correcting non-representativeness of admin and online data using data integration techniques under mis-classification error; 3) inferring about the demand for explicitly defined skills and other characteristics of workers and companies, such as types of contract or work mode – remote, hybrid, mobile, traditional; 4) accounting for uncertainty due to non-random character of data to answer questions regarding the structural changes observed in the vacancy market (e.g. the ones resulting from the COVID-19 pandemic, refugees' inflow, artificial intelligence); 5) the analysis of search and matching between job seekers and employers using unit-level data instead of common aggregate data approach.

**The study will bring new statistical and economic solutions to labour market research.** Innovative nature of the proposed statistical approach is connected to the use of *big data* (from administrative and online sources) supported with methods designed for inference based on non-probability samples with full representativeness treatment to support official statistics. The developed new methods with *big data* will enable recognition of labour market transformation, such as new job search and matching possibilities (online platforms), flexible work contracts (e.g. B2B), human capital measurement – skills and qualifications. Innovative nature of this approach results from the fact that this transformation cannot be fully tracked by traditional methods.

The project will be based on a unique collection of micro data, in particular: Administrative job ads (Central Job Offers Database; CBOP), Online job advertisements, the Demand for Labour survey; the Labour Force Survey and the Register of Unemployed.

We aim to enhance national scientific cooperation. The project will be led by a consortium of two Polish Universities: University of Information Technology and Management in Rzeszów (Leader; UITM) and Poznań University of Economics (PUEB). Tasks associated with these goals will be assigned to the research team of the Principal Investigator (**dr hab. Robert Pater**), co-investigator from PUEB (**dr Maciej Beręsewicz**), 2 co-investigators and 2 PhD students.