

## WIN 2025 CONFERENCE PROGRAMME

Tuesday, 4 February 2025	
8:00-9:00	<i>Registration and welcome coffee</i>
9:00-9:30	<b>OPENING SESSION</b> <ul style="list-style-type: none"> <li>- <b>Dominika Rogalińska</b>, Statistics Poland</li> <li>- <b>Albrecht Wirthmann</b>, Eurostat</li> </ul>
9:30-10:30	<b>KEYNOTE SPEECH</b> <ul style="list-style-type: none"> <li>- GenAI for official statistics - opportunities and dangers <b>Marko Grobelnik</b>, Jožef Stefan Institute</li> </ul>
10:30-11:00	<i>Coffee break</i>
11:00-12:30	<b>SESSION I. WEB SCRAPING AND INFRASTRUCTURE</b> <b>Chair: Alexander Kowarik</b> <ol style="list-style-type: none"> <li>1. URL finding: looking back, progress and plans for the future - <b>Heidi Kühnemann</b></li> <li>2. Identifying official firm websites: a comparison of machine learning-based URL retrieval methods and AI-powered search engines - <b>Donato Summa</b></li> <li>3. State of play of the Data Acquisition Service (DAS) of the Web Intelligence Hub (WIH) – <b>Mátyás Mészáros</b></li> <li>4. Providing online based enterprise characteristics with the Web Intelligence Hub - <b>Jacek Maślankowski</b></li> <li>5. Firms innovation capabilities and corporate websites: evidence on Italian SMEs - <b>Caterina Liberati</b></li> </ol>
12:30-13:30	<i>Lunch break</i>
13:30-15:10	<b>SESSION II. OJA USE CASE</b> <b>Chair: Fernando Reis</b> <ol style="list-style-type: none"> <li>1. Leveraging online job advertisements for green skills analysis in France - <b>Emiline Roger</b></li> <li>2. Development of a labour shortage indicator by occupation from OJA data - <b>Annalisa Lucarelli</b></li> <li>3. Combining online job advertisements with probability sample data for enhanced small area estimation of job vacancies - <b>Donatas Šlevinskas</b></li> <li>4. Assessment of classifiers using pre-defined data source - <b>Vladimir Kvetan</b></li> <li>5. Online job advertisements classification using encoder-like large language model - <b>Mikołaj Tym</b></li> <li>6. Using language models for extracting regions of employment from online job vacancies - <b>Adam Tsakalidis</b></li> </ol>
15:10-15:40	<i>Coffee break</i>
15:40-17:10	<b>SESSION III. OBEC USE CASE</b> <b>Chair: Olav ten Bosch</b> <ol style="list-style-type: none"> <li>1. Evaluating the completeness of business databases: a comparison with official records using web scraping techniques - <b>Josep Domenech</b></li> <li>2. Use of dedicated business website to enhance the statistical business register in the Netherlands - <b>Arnout Van Delden</b></li> </ol>



	<ol style="list-style-type: none"><li>3. Applying survey sampling theory to web-scraped data: an analysis of OBEC data using the IPW estimator - <b>Vilma Nekrašaitė-Liegė</b></li><li>4. Online based enterprise characteristics (OBEC) in Statistics Poland - <b>Ewelina Niewiadomska</b></li><li>5. Trade links: estimating interregional trade using weblinks - <b>Juergen Amann</b></li></ol>
<b>Wednesday, 5 February 2025</b>	
9:00-10:40	<b>SESSION IV. NEW USE CASES</b> <b>Chair: Klaudia Peszat</b> <ol style="list-style-type: none"><li>1. New use-cases of web data for official statistics - <b>Olav ten Bosch</b></li><li>2. Measuring construction activities using advertisements from real estate portals. ESSnet WIN Work Package 3, Use Case 2 - <b>Tobias Gramlich</b></li><li>3. Analysing housing market in Tricity Metropolitan Area in Poland - <b>Olgun Aydin</b></li><li>4. Constructing a Hedonic House Price Index for Poland using listings data from 1996-2024 - <b>Radosław Trojanek</b></li><li>5. Using web data for energy statistics: methodology and key lessons - <b>Herbeth Sandrine</b></li></ol>
10:40-11:10	<i>Coffee break</i>
11:10-12:50	<b>SESSION V. QUALITY OF WEB DATA</b> <b>Chair: Ciprian Alexandru</b> <ol style="list-style-type: none"><li>1. Web content based statistics: the challenges ahead - <b>Fernando Reis</b></li><li>2. Exploiting the web presence of enterprises to improve NACE code classification - <b>Johannes Gussenbauer</b></li><li>3. Assessing the quality of enterprise characteristics and online job advertisement classifications derived from web data - <b>Ville Auno</b></li><li>4. Quality guidelines for acquiring and using web scraped data - <b>Magdalena Six</b></li><li>5. A specialised architectural framework for web data: the BREAL extension and enhancement - <b>Giuseppina Ruocco</b></li></ol>
12:50-13:50	<i>Lunch break</i>
13:50-15:00	<b>SESSION VI. METHODOLOGY ON USING WEB DATA</b> <b>Chair: Jacek Maślankowski</b> <ol style="list-style-type: none"><li>1. Selective scraping, sampling and other methods to minimize known causes of biases of web data - <b>Alexander Kowarik</b></li><li>2. Online job advertisements deduplication using large language model - <b>Jakub Żerebecki</b></li><li>3. Finding the Goldilocks data collection frequency for the Consumer Price Index - <b>Luigi Palumbo</b></li><li>4. Integrating big data and administrative sources for estimating vehicle mileage and analyzing road traffic accidents - <b>Marco Broccoli</b></li></ol>
15:00-15:15	<b>CLOSING SESSION</b>
15:30-18:00	ESSnet WIN WP leads and Eurostat meeting