

IPS 776

Web Data for Official Statistics – Methodology, Quality, Production and Community

Alexander Kowarik, Statistics Austria

8 October 2025

The Hague



Web Intelligence
Network



**Funded by
the European Union**

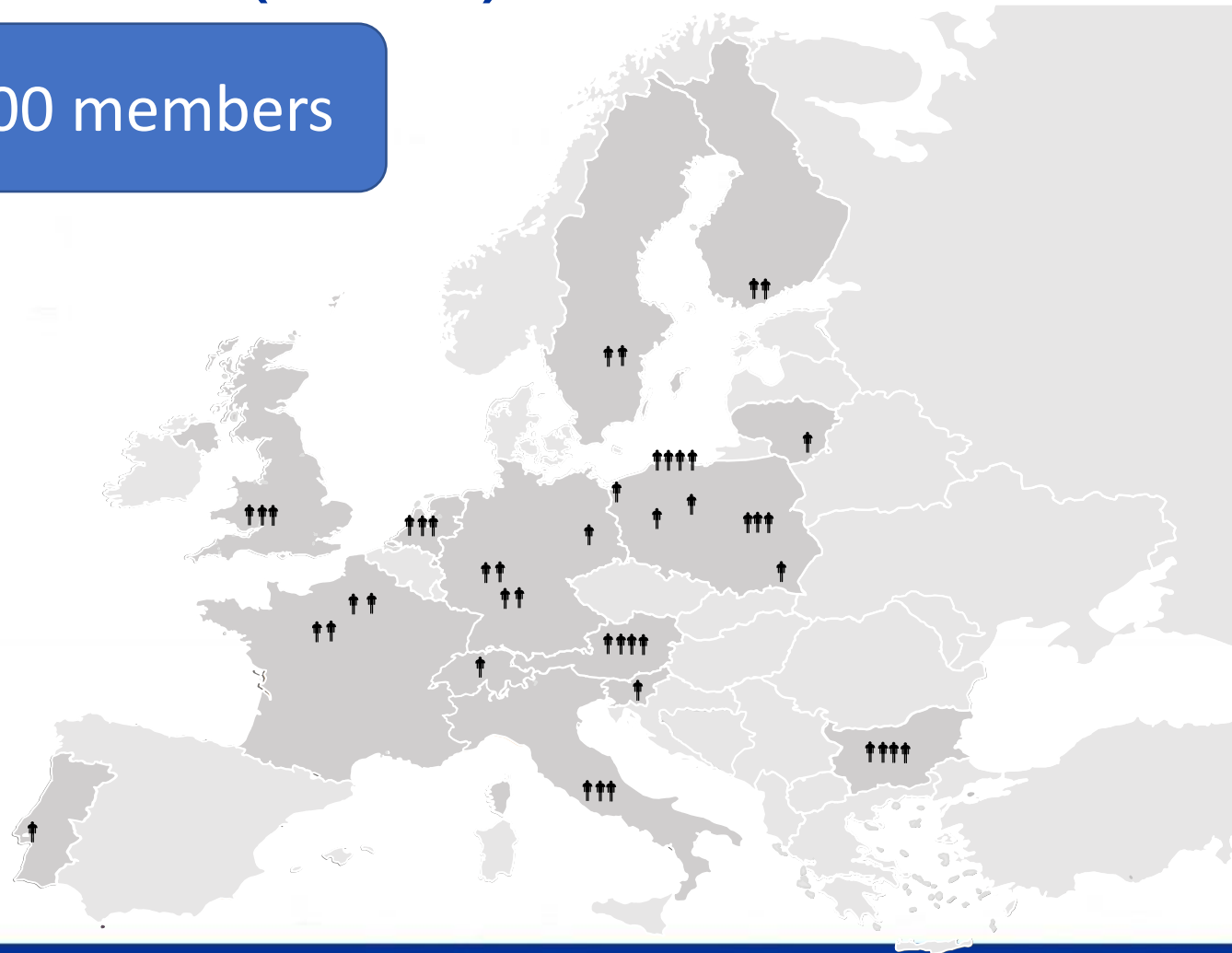
Web Intelligence Network (WIN)

14 countries, 17 organizations, ~100 members

Contribute to the development
of the Web Intelligence Hub

Reach out to **all ESS countries**

Use web data, use the WIH

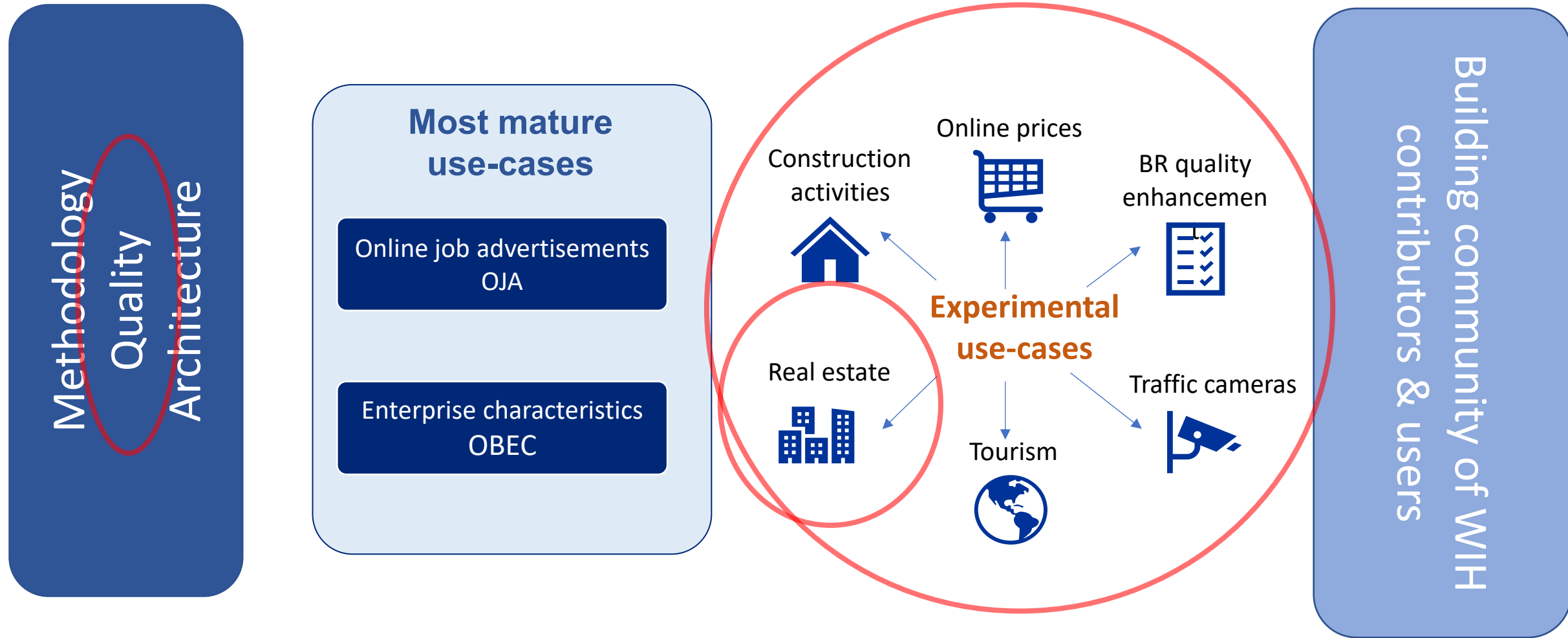


Web Intelligence
Network



**Funded by
the European Union**

Topics WIN was looking at



Web Intelligence
Network



**Funded by
the European Union**

Results for OJA and OBEC



- *Technical guidelines on the implementation of the software produced* - deliverable shows all necessary information to move OBEC and OJA usecases into the production using the WIH environment – it is a guide for accessing the system, reading the data, with examples of possible use
- *Suggested set of new metadata for the use by Eurostat with suggested set of tables to publish at Eurostat's website* - contains templates of tables and metadata, with data calculated by the use of the WIH platform and DataLab



Reports available from experimental work

- [Real estate data](#) - Experimental statistics for characteristics of the real estate market
- Report on methods and feasibility to track [construction activities](#) based on real estate web portals
- Reports on methodology and results [for online prices](#)
- Report on methods for analysing [hotel price data](#) and computing various indices of interest
- Report on methods to be used for [imputation of price data](#) in price statistics (travel-related expenditures)
- Reports on methodology and results to use [online data for business register](#) enhancement
- [Faster Economic Indicators](#) using new data sources - Reports on assessment of challenges and opportunities.



Web Intelligence
Network



**Funded by
the European Union**

Cross-cutting work – deliverables



- Quality

- [Quality Guidelines for acquiring and using web scraped data](#) – framework for identifying and selecting data sources for the web scraping process and guidelines on evaluating existing data set in terms of stability
- [Quality Assessment for the statistical use of webscraped data](#) – assessing the quality of data classification

- Methodology

- [Methodology report on using web scraped data](#) – sampling, web scraping process and causes of bias, methods specific for web scraped data
- [Minimal guidelines and recommendations for implementation](#) – guidelines for integrating webscraped data into official statistics

- Architecture

- [BREAL - Big Data REference Architecture and Layers for web scraped data](#) - “Extension and Enhancement (E&E)” of the BREAL framework



Presentations of today

- *Magdalena Six - Measuring the quality of official statistics based on web data*
- *Olav ten Bosch - Official statistics using web data: new use cases*
- *Klaudia Peszat - Real-estate statistics from web data*

Presentations



Web Intelligence
Network



**Funded by
the European Union**

Discussion

- How can we best address the methodological challenges arising from poor data quality in new data sources?
- To what extent is a purely web-based approach sufficient, and when does integration with established sources such as surveys or administrative data become indispensable?
- Should future projects prioritize solving concrete problems (e.g., developing fast labour market indicators) rather than focusing on specific data sources?
- How can multi-source data approaches – such as combining MNO data with registers or linking web data with tourism statistics – enhance accuracy and comprehensiveness?
- What are the opportunities and risks of applying Large Language Models (LLMs) for URL discovery, particularly in identifying and categorizing relevant web data sources?



Web Intelligence
Network



**Funded by
the European Union**

Lessons learned & conclusions

Immense knowledge-exchange, learning and added value from working together

New data sources
vs. official statistics
quality standards

Transparency &
control of the
production process

Role of domain
experts

Top-down vs.
bottom-up approach

National agenda and
country specificities

Governance set-up

Legal constraints

Added value for the
NSIs

WIH - highly innovative & extremely complex endeavor



- uRos2025 on-site event
24-26 November 2025

<https://www.urosconf.org/>

