



# ESSnet on BD II WPC Implementation – Enterprise Characteristic

Galya Stateva  
Bulgarian National Statistical Institute



# Achieved results from ESSnet on BD I

- Logical Reference Architecture of webscraping of Enterprises' websites
- Methods
- IT solutions - developed software available at:  
[https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WP2\\_Links](https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WP2_Links)
- Legal Framework & Netiquette
- Output indicators



# The Actors: WPC Participants

- SA - Austria
- BNSI – Bulgaria (Coordinator)
- Destatis – Germany
- SF - Finland
- Istat – Italy
- CBS – Netherlands
- GUS – Poland
- ONS – United Kingdom



# Enterprises Characteristics: Objectives

## Main objectives:

- Improve or update existing information (ICT usage in enterprises and BR)
- Maximize the quality and quantity of the statistical outputs
- To achieve important economies of scale: opportunities for sharing of resources at ESS-level



# Deliverables and work organization

Tasks	Timetable	Participants
ESS web-scraping policies	June 2019	BG, UK, WPB team
Methodological Framework/Guidelines	October 2019, v. 1 October 2020, v. 2	All
Experimental Statistics, including reference metadata	October 2019 October 2020	AT, BG, IT, NL, PL, UK
Starter Kit for NSIs	June 2020, v. 1 October 2020, v. 2	BG, IT, PL
Quality template for statistical outputs	August 2020	AT, BG, FI, PL



# Task 1: ESS web-scraping policies

Activities	Timetable	Participants
Developing of the Draft ESS web-scraping policy on the base of WP2 Deliverable 2.1	February 2019	UK and BG
Internal discussion and feedback on the Draft	March – May 2019	All
WP B Contribution and feedback	April – May 2019	UK, BG and WPB team
Consultation with Eurostat (e.g. TF Big data, B1 Methodology; Innovation in official statistics)	April – May 2019	UK, BG and WPB team
Final ESS web-scraping policy	June 2019	UK and BG



## Task 2: Methodological Framework

Activities	Timetable	Participants
Identification of statistical production processes in terms of four phases of the BD life cycle and possible statistical outputs at national level	February 2019	All
Standardization of the WP2 use-cases and selection of potential new use cases	February 2019	All
Development and test of functional prototypes	September 2019, ver. 1 September 2020, ver. 2	All
Definition of the implementation requirements of prototypes	July – September 2019, ver. 1 July – September 2020, ver. 2	All
Data and application architecture for BD production	September 2019, ver. 1 September 2020, ver. 2	All

Storage

Quality

Metadata

Searching

Crawling

Scraping

...

Preprocessing

Cleaning

...

Machine Learning

Deterministic  
decision rules

...

Visualization

SBR update

...

Collect  
(Internet access)

Process

Analyse

Disseminate

Four Phases of Big Data Life Cycle





## Task 3: Experimental Statistics

Activities	Timetable	Participants
Calculation of the statistical indicators defined by the Methodological Framework	September 2019 September 2020	AT, BG, IT, NL, PL, UK
Evaluation of the quality indicators for all statistical outputs (e.g. accuracy, sensitivity, specificity)	September 2019 September 2020	AT, BG, IT, NL, PL, UK
Reference metadata for the dissemination of experimental statistics	September 2019 September 2020	AT, BG, IT, NL, PL, UK
Dissemination of experimental statistics on a wiki platform and dedicated page of Eurostat	October 2019 October 2020	AT, BG, IT, NL, PL, UK

Recommended period for web-scraping activity? (ICT usage in enterprises, e.g. March-April)



## Task 4: Starter Kit for NSIs

Activities	Timetable	Participants
Selection of software solution <b>by use-cases</b>	December 2019	BG, IT, PL
Making of Starter Kit packages	June 2020, v.1 October 2020, v.2	BG, IT, PL
Preparing of guidelines and testing procedures	May 2020, v. 1 September 2020, v.2	BG, IT, PL



# Task 5: Quality template for statistical outputs

Activities	Timetable	Participants
Investigation of the UNECE Framework for the Quality of Big data and SGA-2 WP8	December 2019	AT, BG, FI, PL
Definition of quality concepts for experimental statistics on enterprise characteristics	June 2020	AT, BG, FI, PL
Quality template design (MSD)	August 2020	AT, BG, FI, PL



# Other issues

- WPC webex meetings
- F2F meetings
- Project Wiki and GitHub
- Economies of scale



**THANK YOU FOR YOUR ATTENTION!**

