



Web Intelligence Network Conference

From Web to Data

Conference Opening

Gdansk, 4 February 2025

Albrecht WIRTHMANN, Eurostat

Official Statistics paradigm shift

- From a **data-scarce** world ...
 - where statistical authorities concentrated on **collecting data** using surveys/census and administrative sources as primary sources
- ... to a **data-rich** world
 - where statistical authorities should **distil reliable information** from multiple data sources building on their assets: traditional sources, privacy preservation, quality and sound methodologies (calibration, integration, ...)



European Data Strategy

Create a single European data space in a way that:

Data can flow within the EU and across sectors for the benefit of businesses, researchers and public administrations.

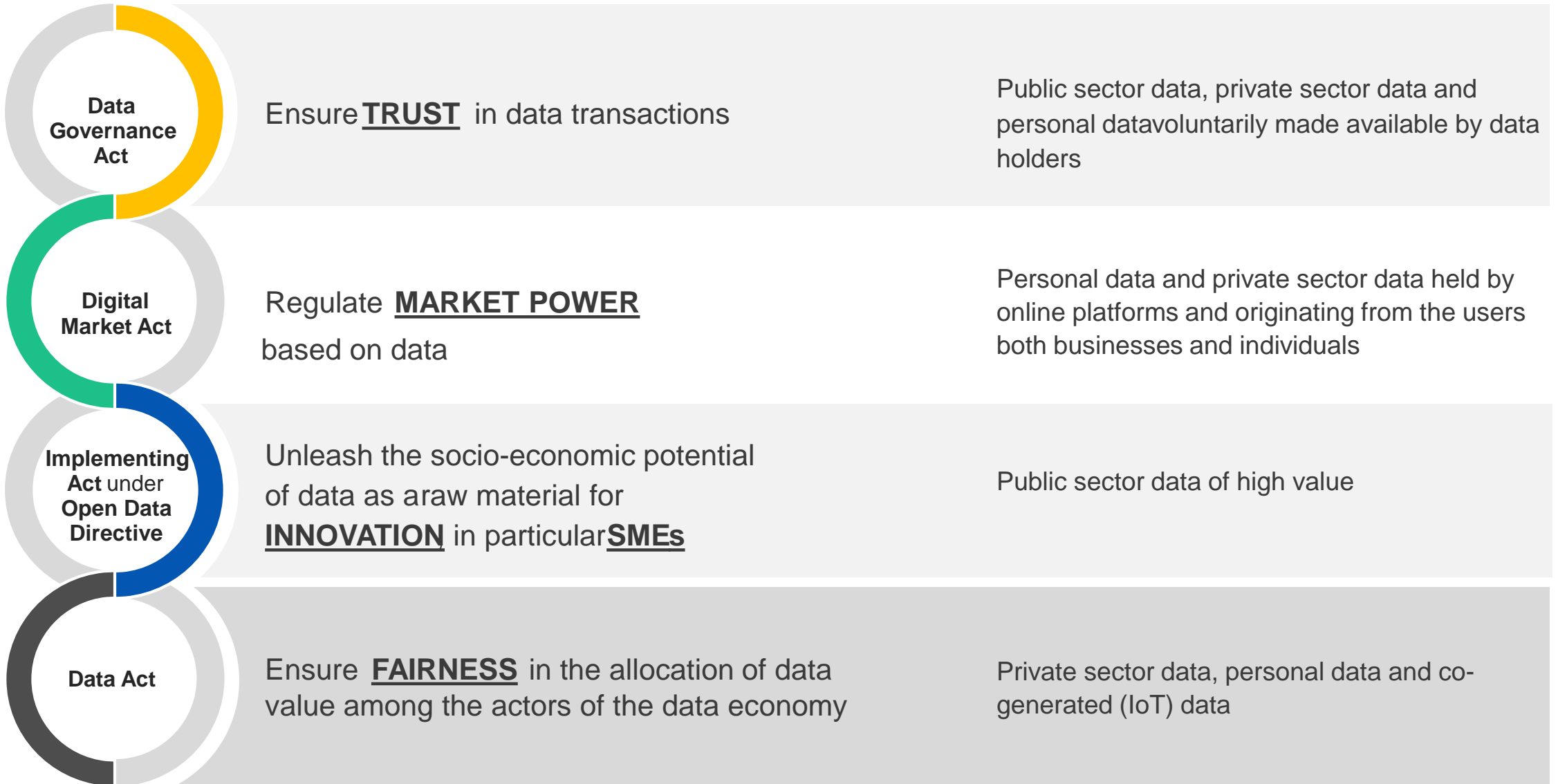
Fair, practical and **clear rules apply** for access to and use of data.

There are clear and trustworthy **data governance mechanisms**.

European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected.

Next generation **standards, tools and infrastructures** to store and process data is invested in.

European Strategy for Data: 4 Key Instruments



Common European Data Spaces

**Rich pool of data
(varying degree of
accessibility)**

**Free flow of data
across sectors and
countries**

Full respect of GDPR

**Horizontal
framework for data
governance and
data access**



Health



Industrial &
Manufacturing



Agriculture



Finance



Mobility



Green Deal



Energy



Public
Administration



Skills

- Technical tools for data pooling and sharing
- Standards & interoperability (technical, semantic)
- Sectoral Data Governance (contracts, licenses, access rights, usage rights)
- IT capacity, including cloud storage, processing and services

Data infrastructures and governance go hand in hand

The reply of the ESS

Official Statistics is important for the society.
And evolves with it ...

“Knowledge is power; statistics is democracy.”

*Olavi Niitamo,
former DG of Statistics Finland*

*“Official Statistics is about representing Life.
As Life changes, so must do Official Statistics.”*

*Mariana Kotzeva
Rome 4.7.2018
13° Conferenza Nazionale di Statistica ISTAT*

Strategic priorities for the next Decade

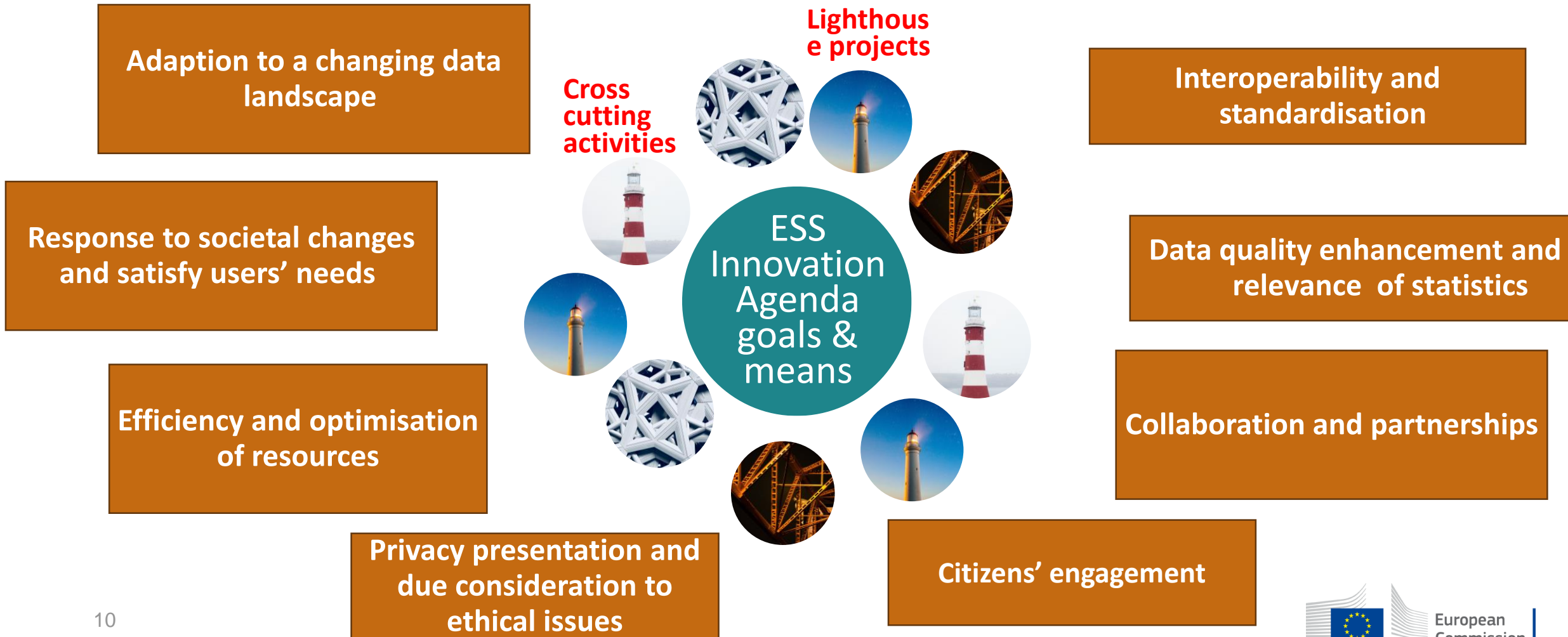
Keep Relevance, Trustworthiness and effectiveness in response to evolving statistical information needs

- Ensure continuous relevance of European statistics
- Chartering new frontiers for European statistics
- Quality of European Statistics
- Increase trust in the ESS, data security, protection and ethics

Leverage new data sources and digital technologies

- Unlock full potential of new data sources, administrative data and data integration
- New technologies for developing, producing and disseminating European statistics
- International statistical cooperation and partnerships

European Statistical System Innovation Agenda - goals



Innovation agenda portfolio of projects

22 factsheets available on CROS for use and trans

One-Stop-Shop for Artificial Intelligence and Machine Learning for Official Statistics (AI/ML4OS)

Towards a comprehensive suite of resources for developing and implementing AI/ML solutions for official statistics



The project

New advances in artificial intelligence (AI) and machine learning (ML) present promising opportunities for the production of official statistics, with applications across the production pipeline including improvements in efficiency, unlocking potential of novel data sources, enhancing data quality and delivering more user-friendly outputs.

Harnessing this potential is a key strategic aim of the European Statistical System (ESS). Through its Innovation Agenda, the ESS is pursuing a coordinated approach and has established a One-Stop-Shop for Artificial Intelligence and Machine Learning in Official Statistics (AI/ML4OS). This is a four-year project, which includes the participation of National Statistical Institutes (NSIs) from 16 countries.

The overarching aim of the project is to produce a single point of entry for ESS staff to facilitate accessing and deploying AI/ML solutions.

The motivation

Expertise on new analytical tools for statistical production is dispersed across the ESS, with each NSI having established its own competencies in specific areas. Yet many challenges and opportunities are common to all Member States. As coordinating the delivery of its outputs, the remaining seven WP are use cases, each of which explores a specific tool or theme within AI/ML in the context of official statistics. Each WP is led by organisations. A collaborative approach to developing solutions is therefore desirable. It is further hoped that this collective effort will lead to thriving communities of ESS staff for ongoing mutual support beyond the timeline of the AI/ML4OS project.


The methodology

The AI/ML4OS project is composed of 13 work packages (WP) that are separate but interlinked. The first six WPs are oriented towards supporting the operation of the project and opportunities are common to all Member States. The remaining seven WPs are use cases, each of which explores a specific tool or theme within AI/ML in the context of official statistics. Each WP is led by organisations. A collaborative approach to developing solutions is therefore desirable. It is further hoped that this collective effort will lead to thriving communities of ESS staff for ongoing mutual support beyond the timeline of the AI/ML4OS project.

eurostat

Web Intelligence Hub

Modernising the data collection process within the European Statistical System (ESS)



The project

The Web Intelligence Hub (WIH) is a toolbox for high-quality statistics from web content, aiming to integrate smart technologies and innovative data sources within the European Statistical System (ESS).

Implemented by Eurostat and the ESS as part of the ESS Innovation Agenda, the WIH focuses on creating new data pipelines and using cutting-edge technologies such as artificial intelligence (AI), machine learning (ML) and natural language processing.

The WIH applies its tools and methodologies to web content across multiple domains. Currently, the WIH most developed use cases are:

- online job advertisements (OJA);
- online-based enterprise characteristics (DBEC);
- multinational enterprises (MNE)

Future plans include expanding to additional domains. The main goals of the WIH are to:

- develop new data pipelines for web scraping;
- create infrastructure to harvest and extract website content;
- produce microdata for statistical purposes

The motivation

The project is essential for modernising the data collection process within the ESS. By leveraging web scraping and natural language processing technologies, the WIH aims to provide up-to-date and detailed data, and enhance the accuracy and relevance of European statistics to support policymakers with timely and comprehensive insights.

Moreover, the WIH seeks to make use of web content to support the production of high-quality statistics. In today's digital age, web content is more than just an additional data source; it is a crucial, detailed and up-to-date resource. The WIH taps into this vast web content, ensuring that the data is of the highest quality and relevance. By complementing traditional data sources like surveys and administrative data, the WIH enables the production of more timely and detailed European statistics, with less burden on data providers.


The projects outputs, which include methodologies and data for several use cases, offer better insights into topics such as labour market trends and business activities, establishing an infrastructure for continuous and efficient content, and data extraction for statistical production.

In turn, these high-quality statistics can be used to improve decision-making capabilities for policymakers and generally support the public good, given that they are more reflective of the wealth of information available on the web.

eurostat

Statistical Literacy

Empowering people and educators



The project

Statistical literacy has never been more important. This is especially the case in the digital age, where data is becoming increasingly central to boost statistical literacy.

The ESS Innovation Agenda, Eurostat's project on improving statistical literacy is aimed at both young people and educators.

The motivation

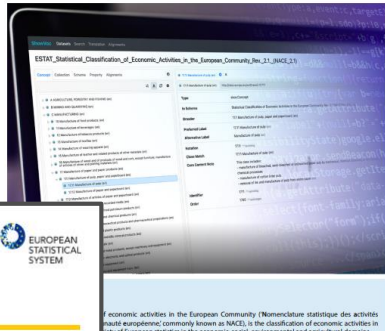
As new technologies and sources of information spread, it is increasingly important for everyone to understand data in most fields. Making official statistics easier to grasp is a key step towards improving statistical literacy in society.

Recognising the need to improve statistical literacy in the education sector, Eurostat is committed to creating resources that make the learning and understanding of statistics accessible at all levels. Through these efforts, Eurostat aims to equip educators and students alike with the skills necessary to navigate and interpret the data-rich world we live in.

eurostat

Machine learning for automated classification according to NACE

Updating the classification of economic activities in the EU



The project

Economic activities in the European Community (Nomenclature statistique des activités économiques européennes) commonly known as NACE, is the classification of economic activities in every European statistics in the economic, social, environmental and agricultural domains. As the economy evolves, it has become increasingly important to update the NACE classification to meet the needs of the ESS.

The motivation

In today's fast-changing data world, the European Statistical System (ESS) Innovation Agenda aims to meet the fast-growing demand for new, more timely and detailed statistics that people need and trust. As part of this agenda, Trusted Smart Surveys can revolutionise data collection by combining traditional survey techniques with modern technological advancements, keeping the respondent at the heart of the data they submit.

These surveys intelligently combine self-report questions with smart features collected via sensor-enabled devices such as smartphones, wearables and other devices, aiming to enhance data quality, reduce burden on participants, and produce more timely and granular data.

The Trusted Smart Surveys project seeks to develop this concept by showcasing a complete, end-to-end data collection process.

The motivation

In official statistics, adopting a strategic approach has become increasingly essential for addressing new informational needs such as globalisation, innovation and the green transition, while minimising the statistical burden on respondents. Microdata linking (MDL) is one such approach to meet these new needs and challenges. MDL is a method used to create new statistics and indicators by combining existing data. This approach helps uncover valuable insights and information. Contributing to the European Statistical System (ESS) Innovation Agenda, the aim of this project is to produce new data on enterprises belonging to multinational enterprise (MNE) groups and to enhance the use of national business registers (NBR) and the EuroGis Register (EGR) for statistical purposes. In this way, MDL can be used to replace or complement methods to produce statistics.

The motivation

Globalisation and the growing complexity of multinational enterprise group activities present ongoing challenges for accurately capturing and analysing enterprise groups within European statistics. This project aims to develop new statistical indicators that offer deeper insights into the performance and characteristics of enterprises that belong either to a domestic or to a multinational enterprise group, be them domestically or foreign controlled. Central to this approach is MDL, which relies on unique identification numbers across statistical domains to enable microdata linkage. In business statistics, the national business registers as well as EGR form the foundation of most MDL initiatives, supporting a cohesive view of enterprise activities across borders.

eurostat

European Space Economy Thematic Account

Assessing the economic impact of the space industry in Europe



The project

The European Space Economy Thematic Account (ESETA) project aims to provide macroeconomic statistics for the EU space sector. It will facilitate a wide range of new analyses, and enhance the understanding of the space sector's contribution to the overall economy in Europe.

In line with the European Statistical System (ESS) Innovation Agenda, the project aims to develop a thematic account for the European space activities and products. The production of space-relevant goods and services will be traced, and their economic effects quantified. The project also addresses the need to monitor the outcomes of public investments in the space economy.

The motivation

The main goals of the ESETA project are to provide a comprehensive view of the space sector in terms of size and value, and to clarify the sector's role in the economy by measuring both supply (production) and demand (consumption). Given the relatively small and dispersed nature of the space economy, its products are spread across various economic activities. A clear definition of the space economy in the form of a list of statistical codes is consequently needed. The ESETA will then enable the measurement and monitoring of the direct and indirect contributions of the space-related industries to the European Gross Domestic Product (GDP) and employment, using official statistics. It will also provide insights into the economic and social benefits of the space economy and its global value chains.

eurostat

Use of microdata linking techniques (MDL) for getting insights into enterprises belonging to multinational enterprise groups

Producing statistics with innovative methods



The project

In today's fast-changing data world, the European Statistical System (ESS) Innovation Agenda aims to meet the fast-growing demand for new, more timely and detailed statistics that people need and trust. As part of this agenda, Trusted Smart Surveys can revolutionise data collection by combining traditional survey techniques with modern technological advancements, keeping the respondent at the heart of the data they submit.

These surveys intelligently combine self-report questions with smart features collected via sensor-enabled devices such as smartphones, wearables and other devices, aiming to enhance data quality, reduce burden on participants, and produce more timely and granular data.

The Trusted Smart Surveys project seeks to develop this concept by showcasing a complete, end-to-end data collection process.

The motivation

In official statistics, adopting a strategic approach has become increasingly essential for addressing new informational needs such as globalisation, innovation and the green transition, while minimising the statistical burden on respondents. Microdata linking (MDL) is one such approach to meet these new needs and challenges. MDL is a method used to create new statistics and indicators by combining existing data. This approach helps uncover valuable insights and information. Contributing to the European Statistical System (ESS) Innovation Agenda, the aim of this project is to produce new data on enterprises belonging to multinational enterprise (MNE) groups and to enhance the use of national business registers (NBR) and the EuroGis Register (EGR) for statistical purposes. In this way, MDL can be used to replace or complement methods to produce statistics.

The motivation

Globalisation and the growing complexity of multinational enterprise group activities present ongoing challenges for accurately capturing and analysing enterprise groups within European statistics. This project aims to develop new statistical indicators that offer deeper insights into the performance and characteristics of enterprises that belong either to a domestic or to a multinational enterprise group, be them domestically or foreign controlled. Central to this approach is MDL, which relies on unique identification numbers across statistical domains to enable microdata linkage. In business statistics, the national business registers as well as EGR form the foundation of most MDL initiatives, supporting a cohesive view of enterprise activities across borders.

eurostat

Trusted Smart Surveys

Revolutionising data collection with modern survey technology



The project

In today's fast-changing data world, the European Statistical System (ESS) Innovation Agenda aims to meet the fast-growing demand for new, more timely and detailed statistics that people need and trust. As part of this agenda, Trusted Smart Surveys can revolutionise data collection by combining traditional survey techniques with modern technological advancements, keeping the respondent at the heart of the data they submit.

These surveys intelligently combine self-report questions with smart features collected via sensor-enabled devices such as smartphones, wearables and other devices, aiming to enhance data quality, reduce burden on participants, and produce more timely and granular data.

The Trusted Smart Surveys project seeks to develop this concept by showcasing a complete, end-to-end data collection process.

The motivation

In official statistics, adopting a strategic approach has become increasingly essential for addressing new informational needs such as globalisation, innovation and the green transition, while minimising the statistical burden on respondents. Microdata linking (MDL) is one such approach to meet these new needs and challenges. MDL is a method used to create new statistics and indicators by combining existing data. This approach helps uncover valuable insights and information. Contributing to the European Statistical System (ESS) Innovation Agenda, the aim of this project is to produce new data on enterprises belonging to multinational enterprise (MNE) groups and to enhance the use of national business registers (NBR) and the EuroGis Register (EGR) for statistical purposes. In this way, MDL can be used to replace or complement methods to produce statistics.

The motivation

Globalisation and the growing complexity of multinational enterprise group activities present ongoing challenges for accurately capturing and analysing enterprise groups within European statistics. This project aims to develop new statistical indicators that offer deeper insights into the performance and characteristics of enterprises that belong either to a domestic or to a multinational enterprise group, be them domestically or foreign controlled. Central to this approach is MDL, which relies on unique identification numbers across statistical domains to enable microdata linkage. In business statistics, the national business registers as well as EGR form the foundation of most MDL initiatives, supporting a cohesive view of enterprise activities across borders.

eurostat



Regulation (EC) 223/2009 on European Statistics

General objectives of the revision

Tap the full potential of digital data sources and new technologies

Enable the ESS to become more pro-active and produce statistics more efficiently

Foster innovation and the development of new statistics

Key changes introduced by the revision

**Statistical
response to
crisis
situations**

**Sustainable
access to new
data sources**

**Innovation and
experimental
statistics**

**New roles in
data
governance
frameworks**

Data sharing

**Statistical
governance**

Implementation of the revised Regulation

Stands and actions

Reuse administrative and publicly available data for developing and producing European statistics

Use of privately held data for statistical purposes

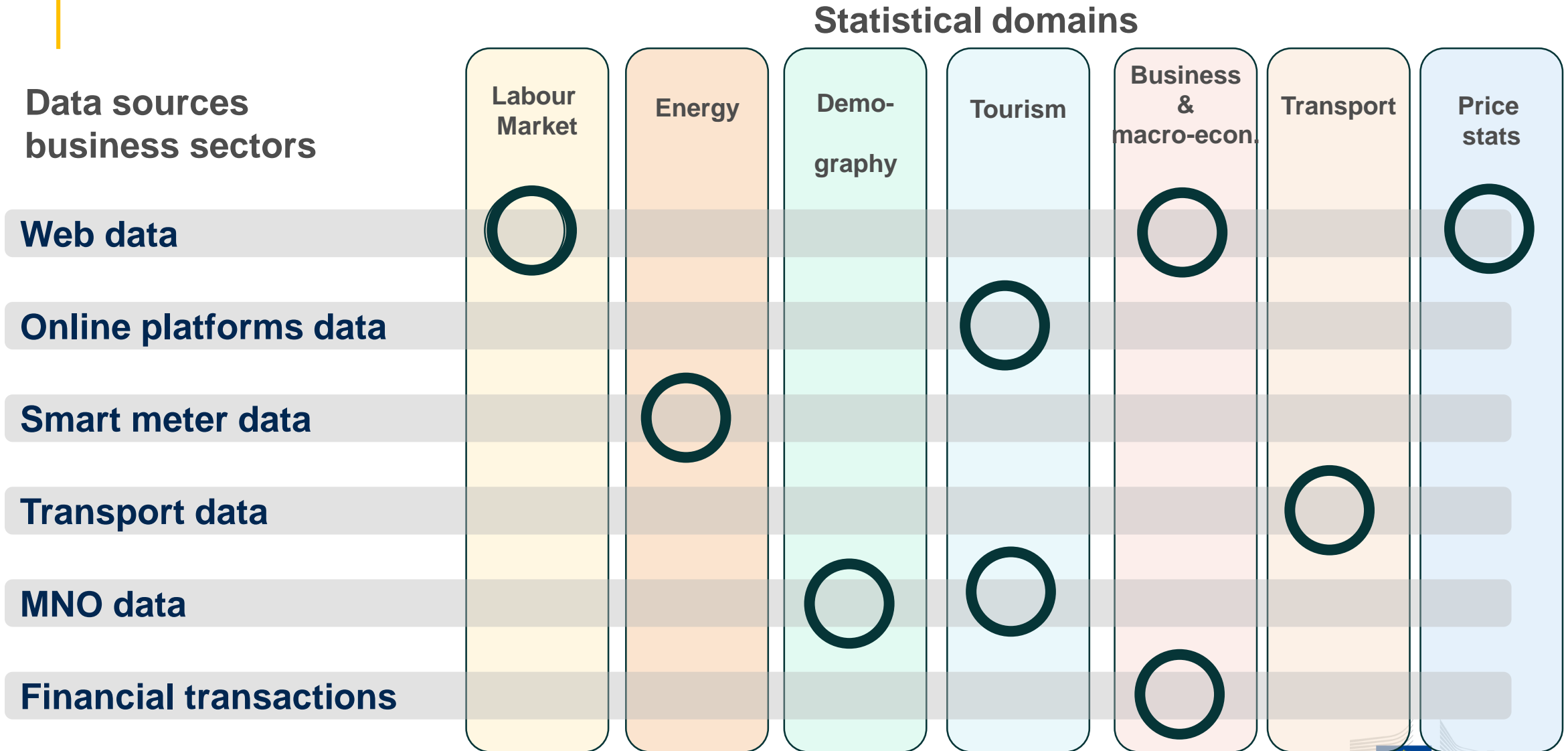
Development of European statistics

New roles for NSIs in national data governance frameworks

Statistical response to urgent policy needs in crisis situations

Foster data sharing within the ESS and with the ESCB

Thematic Domains and Data Sources



Key Aspects for implementation of 223

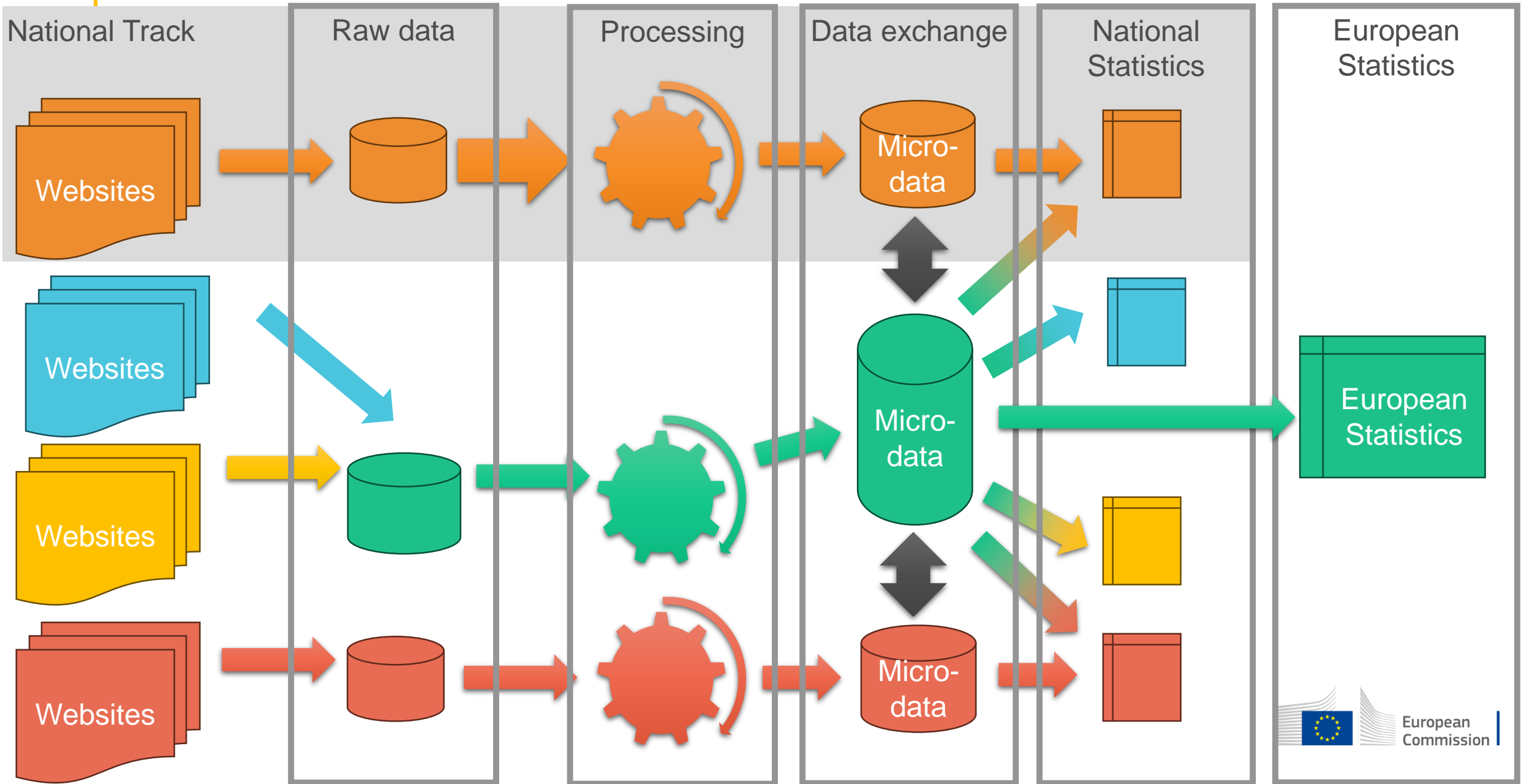
- Added value of new data sources from user perspective
- Impact on current business model for statistical production
 - Development, collection, processing, validation, dissemination
 - Technical modalities
 - Role of ESS members
 - Data for research
 - Costs and resource requirements
 - Need for developing methodology, infrastructure, tools
 - Legal compliance
- Sustainable partnership model (for ESS and for data holders!)
- Education and training
- Data protection
- Public communication and acceptance
- ESS preparedness

Web Intelligence



- Support to capability building at national level
- Creation and maintaining of a community of experts
- Further development of use cases
 - Quality improvements
 - Further elaboration and definition of statistical indicators
- Elaboration and implementation of a business model
 - Access to data sources
 - Data processing
 - Data use
- Application of revised European statistics regulation

Web Data production pipeline



Thank you

Contact: Albrecht.Wirthmann@ec.europa.eu



© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.