Challenges and innovations in data production

Instructions: Click on the link to access each author's presentation.

Organiser: Peter Knizat

Chair: Peter Petko

Participants:

Christopher Sibley: New Data Sources & Trust in Offical Statistics: Central Statistics Office (Ireland)

Tobias Thomas: Innovation and its Impact on Official Statistics: Some experiences from Austria

Áron Kincses: Innovative use of new types of data sources: Some examples from Hungary

Peter Knizat: Challenges and innovations in data: new privately held data sources in official statistics and examples of innovations from Slovakia

IOS-ISI 2024
MEXICO CONFERENCE

New Data Sources & Trust in Offical Statistics

Central Statistics Office (Ireland)

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

Growing Need for Official Statistics

More **complex societal issues** requiring insights

Growing demand for better and **more timely insights**

Demand for more **granular data**

Increasing use of **secondary** (admin and privately held) data

Potential access to **privately held data** i.e. EU Regulation 223 revision

IOS-ISI 2024 MEXICO CONFERENCE

Commercially Held Data: Untapped Reservoir

Potentially more efficient, faster outcomes to help us to:

Understand **phenomena** like population movements, lifestyle changes, disease patterns, public transport and personal consumption.

Understand **increasingly complex and rapidly evolving subject matters**, such as the environment, epidemiology, public infrastructure and our global economy.

Lower the burden on companies and citizens by avoiding survey questionnaires.

Operating in a Changing Landscape....

RISE OF **MISINFORMATION AND FAKE NEWS**

CHANGING **SOCIAL ATTITUDES TO PRIVACY**

GROWING DISCUSSION AROUND **ETHICS**

LOSS OF **TRUST** IN THE ESTABLISHMENT

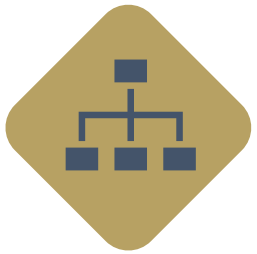
IOS-ISI 2024 MEXICO CONFERENCE

Trust is Key

“A confident relationship with the unknown”

Rachel Botsman

Trust is **earned** not built and can be lost very easily



Varies from organisation to organisation

NSIs work relies heavily on being **trusted**

To date we have achieved trust through **living and demonstrating our values**

IOS-ISI 2024 MEXICO CONFERENCE

Commercially Held Data & Trust

A legal framework for access is needed to earn trust…..

Sustainable flow of data that enables NSIs to produce comparable data over time.

No legal obligation means that commercial entities can **sell** valuable data to highest bidder.

Only National Statistical Institutes are able to produce high-quality, independent, objective, transparent, impartial insights that are free from political influence.

But

A Legal Framework is only a first step - More is required!

Canadian Experience

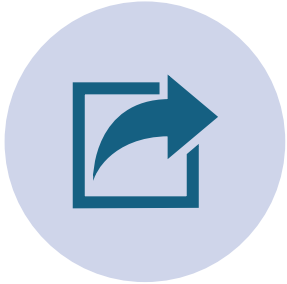
- Statistics Canada wanted to gain insight into personal spending using administrative data from private sector financial companies
- Media story- Statistics Canada planning to collect the banking information of 500,000 households without their knowledge or consent
- Over 100 complaints made, and the Office of the Privacy Commissioner identified significant privacy concerns despite their Statistics Act
- Event had a huge impact on trust and reputation for the NSI-
- Huge levels of negative media coverage
- Questions asked in parliament
- View presented-may be legal, but the action was not viewed as being ethical.

Response

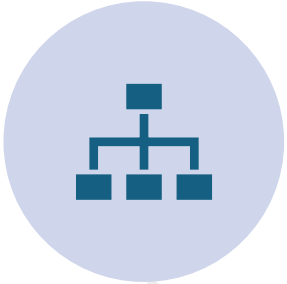
- Created a high-powered **Data Ethics Secretariat** within Stats Canada
- Adopted a **Necessity and Proportionality Framework**
- Every proposal for a new project or data acquisition must explain why it is important
- What the benefits are to Canadians?
- Who needs the information?
- Address **ethical considerations such as privacy, transparency, and fairness**.
- Developed a comprehensive **Trust Centre** on the website

Social Acceptance

It might be **legal BUT is it seen as right** or wrong and who decides this?



Linked to '**do no harm'** principle and to being for the **"public good"**



Need to **build ethics into our data processes**: Design, build, collection, processing, analysis and communication in both survey and administrative data

We also need to **listen, understand, and communicate clearly** what we are doing and why – we need communications and engagement strategies

Creating a Social Acceptance Strategy

Using Ethics as an Enabler

IOS-ISI 2024
MEXICO CONFERENCE


An Ethical Lens...
Internal
Ethics Culture, Structures and Processes inform work of CSO
External: Project Specific
Tailored communications & engagement strategies
External: Foundation Layer
High level communications & engagement strategy around earning trust

Be Proactive…

Consider **key audiences** e.g., data suppliers, the public, media, decision makers, Data Commissioner, special interested parties etc

Develop tailored comms & engagement strategy which considers **key messages:**

- **Media / General public: trust, societal benefit at individual & national level, legal compliance, confidentiality**
- **Data suppliers: assure, listen & respond, societal benefit**
- **Government & Decision-Makers**: support, decision-making power
- **Interest groups** e.g. civil liberties DP groups: legal compliance, anticipate, engage, incorporate messaging
- **Data commissioner**: transparent communication and compliance

Consider **focus groups / consultation processes** for sensitive projects…

IOS-ISI 2024 MEXICO CONFERENCE

Be Transparent...

Develop dedicated **Trust Section on website** which explains:

- What NSIs do and why
- Where data comes from
- Confidentiality
- GDPR
- Legal Framework
- Details of specific Projects.

Remember to sell the societal benefit of why we need the data!

IOS-ISI 2024
MEXICO CONFERENCE

Thank you

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

IOS-ISI 2024
MEXICO CONFERENCE

Innovation and its Impact on Official Statistics

Some experiences from Austria

Prof. Dr. Tobias Thomas
Director General, Statistics Austria

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

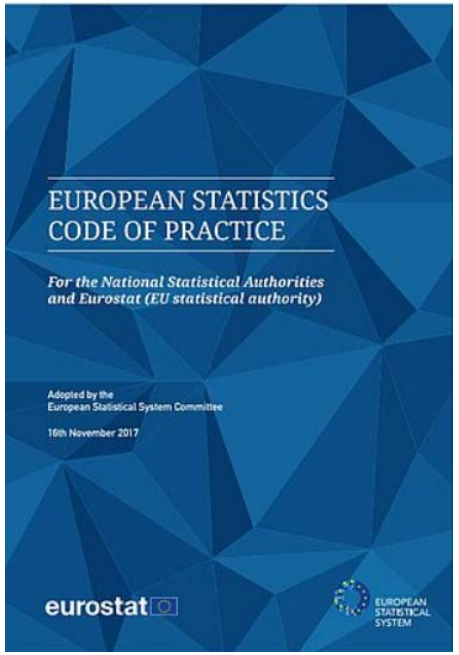
Agenda

- Innovation and its impact on quality
- Emerging innovations in Official Statistics
- AI and machine learning
- Frameworks to foster innovation

IOS-ISI 2024
MEXICO CONFERENCE

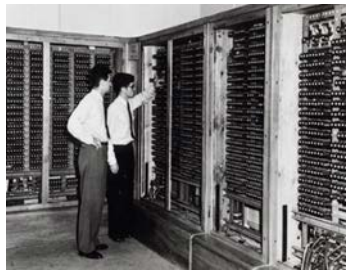
Achieving standards with innovation

- **Independence**
- **Objectivity**
- **Quality**
- **Confidentiality**
- **Accessibility**



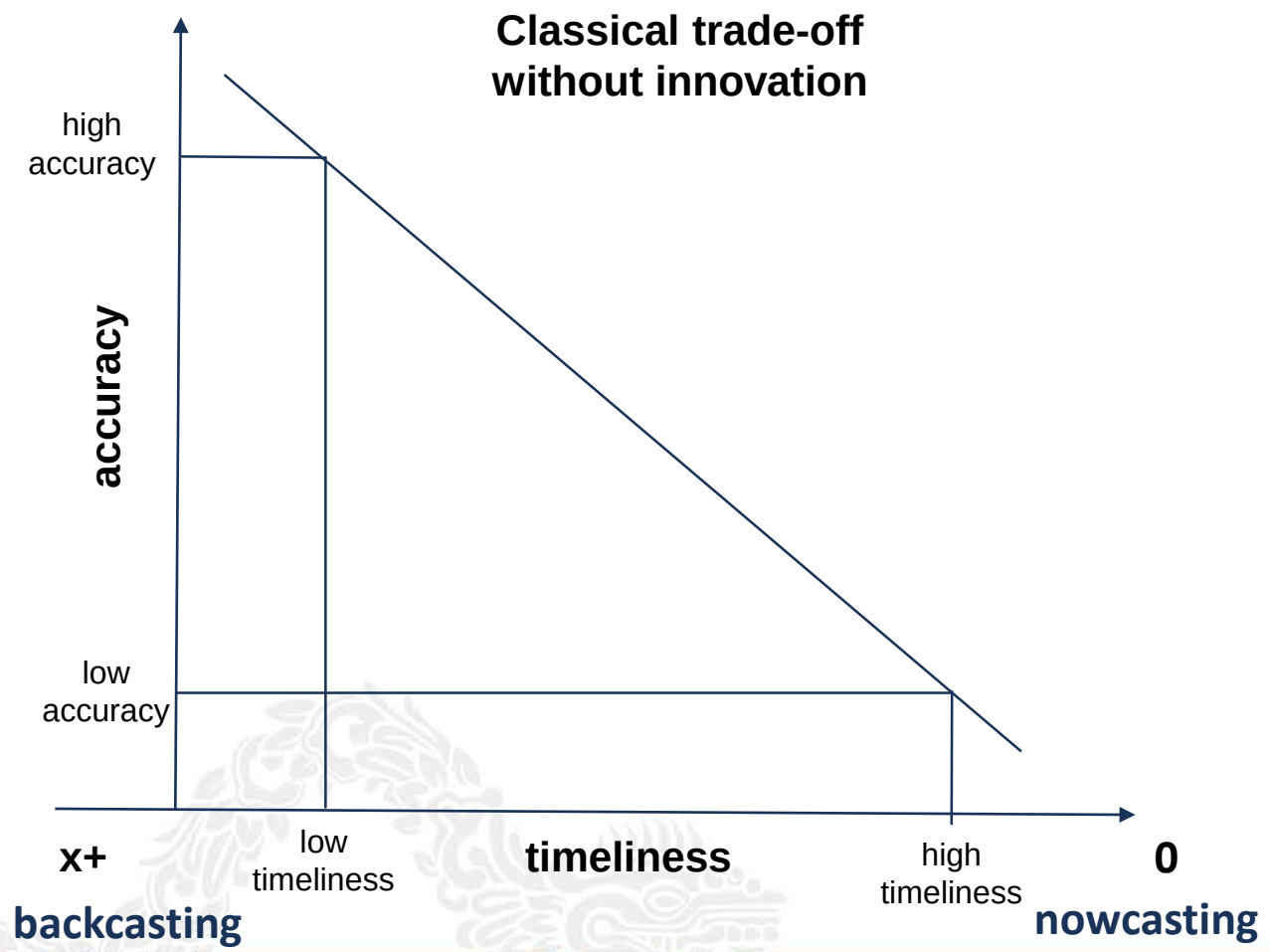
IOS-ISI 2024
MEXICO CONFERENCE

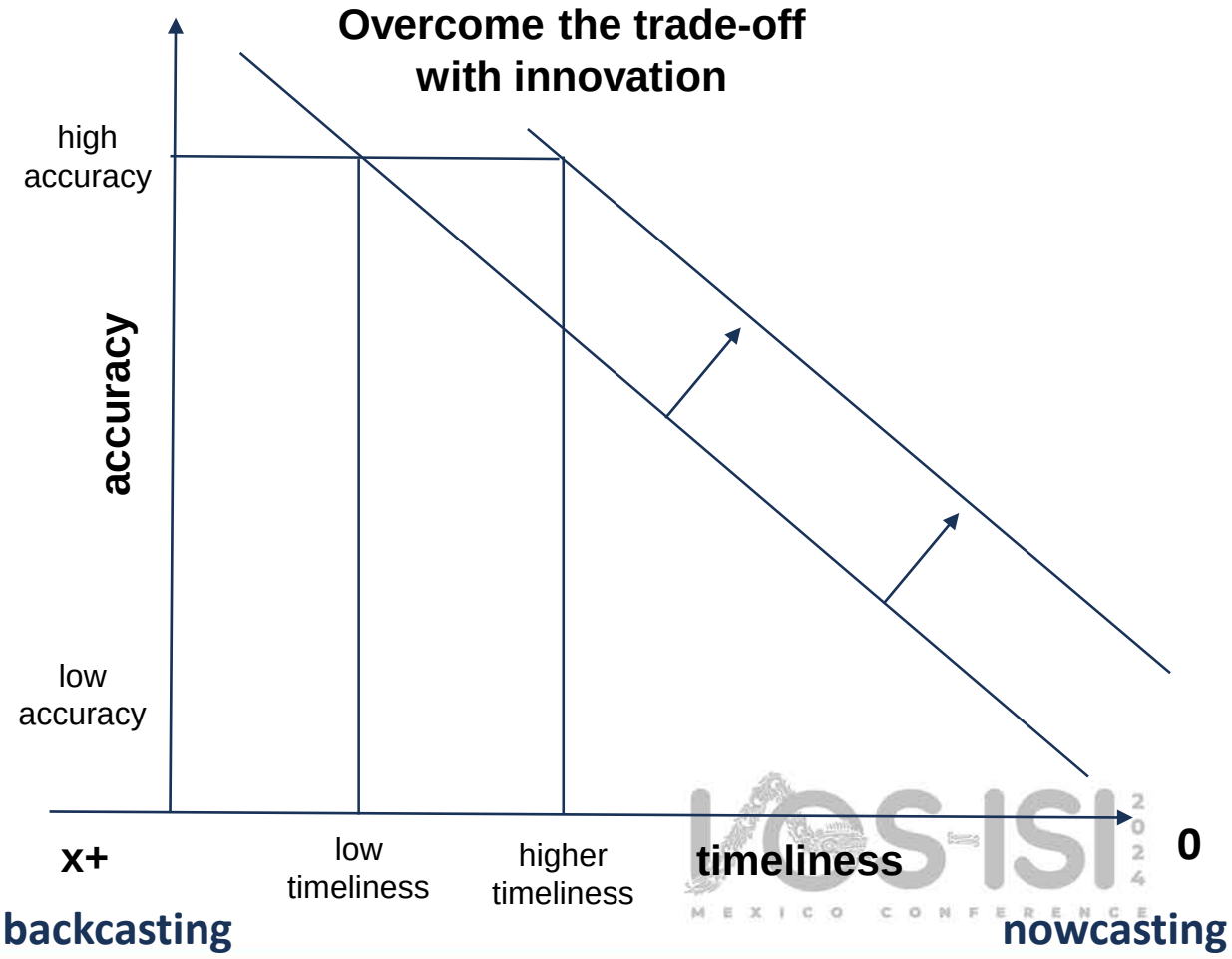
Innovation is a constant companion of Official Statistics...

before 1960s	since 1960s	since 1990s	since 2010	the future
				
no use of IT	**IT use**	**broader use of administrative sources**	**new data sources**	**IoT and trusted smart statistics**
We talk about years	Continuous improvement but still often long periods before statistics are published	„t + x" with continuously decreasing "x" ("flash estimates")	Huge data sets often available at once	Real time statistics

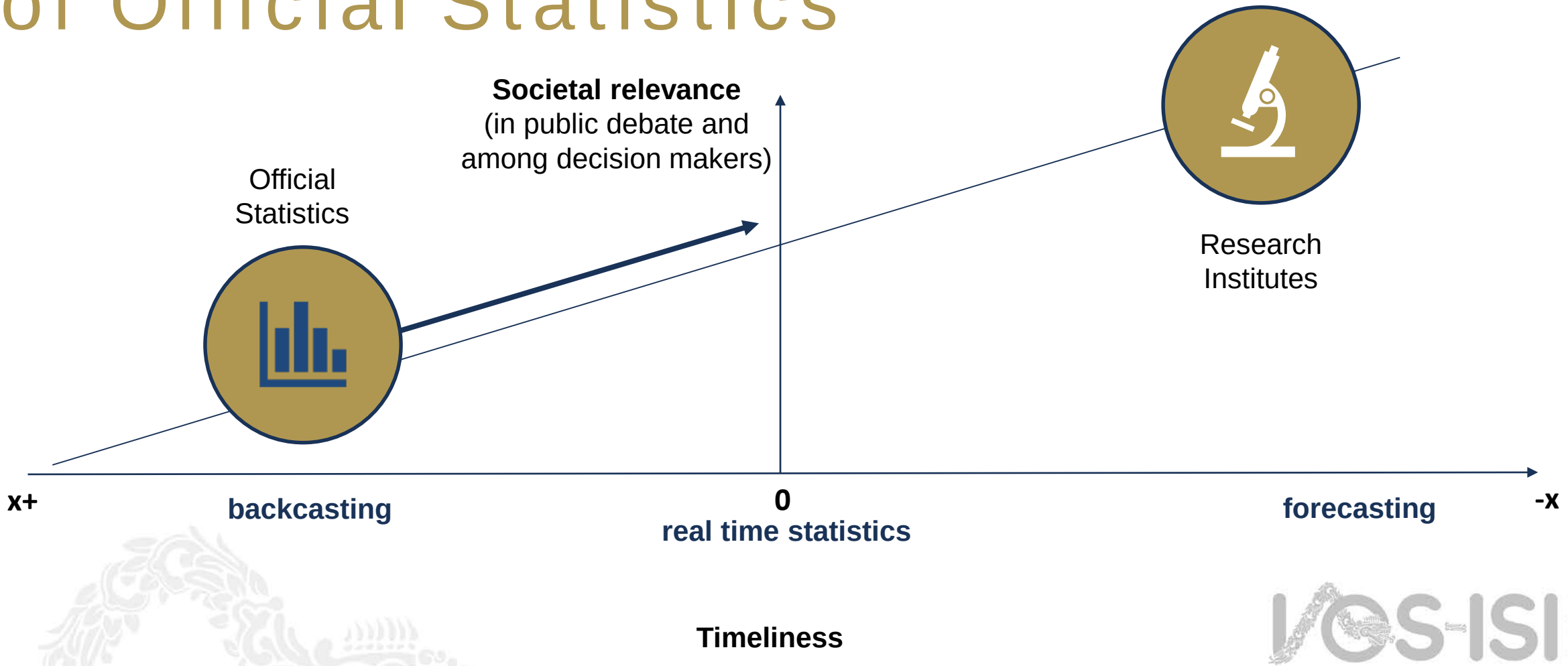
IOS-ISI 2024 MEXICO CONFERENCE

...and helps to overcome traditional trade offs between quality dimensions





Innovation helps to ensure relevance of Official Statistics



New data sources provide new opportunities for Official Statistics

Making an optimized use of all information available

- Comprehensive access to administrative data sources (inclusive overview)
- New digital data sources (web scraped data, earth observation data)
- Privately held data (data of corporates)

IOS-ISI 2024 MEXICO CONFERENCE

New data sources used in Austria
(selection)

Business Statistics

- Use of balance sheets („Saldenliste“)
- Data transfer via webservice interface

Consumer Price Index (CPI)

- New regulation governs the provision of **scanner data**
- Use of Machine Learning Methods

Agricultural Statistics

- Experimental project using Earth Observation data
- Agreement with EODC on using cloud based services

New innovative methods are increasingly applied

Keeping up methodological soundness

- Broader use of Machine Learning
- More applications for Micro Simulation
- Improved nowcasting methods
- Standardized metadata-driven solutions
- Advanced use of real time data

IOS-ISI 2024 MEXICO CONFERENCE

New methods used in Austria
(selection)

Machine Learning

- Classifying free text
- Imputation
- Synthetic data generation

Micro Simulation

- Population projection
- Projection for certain groups (students etc.)

Improved Nowcast Methods

- Quick estimation of Consumer Price Index
- More in depth analysis of data revisions

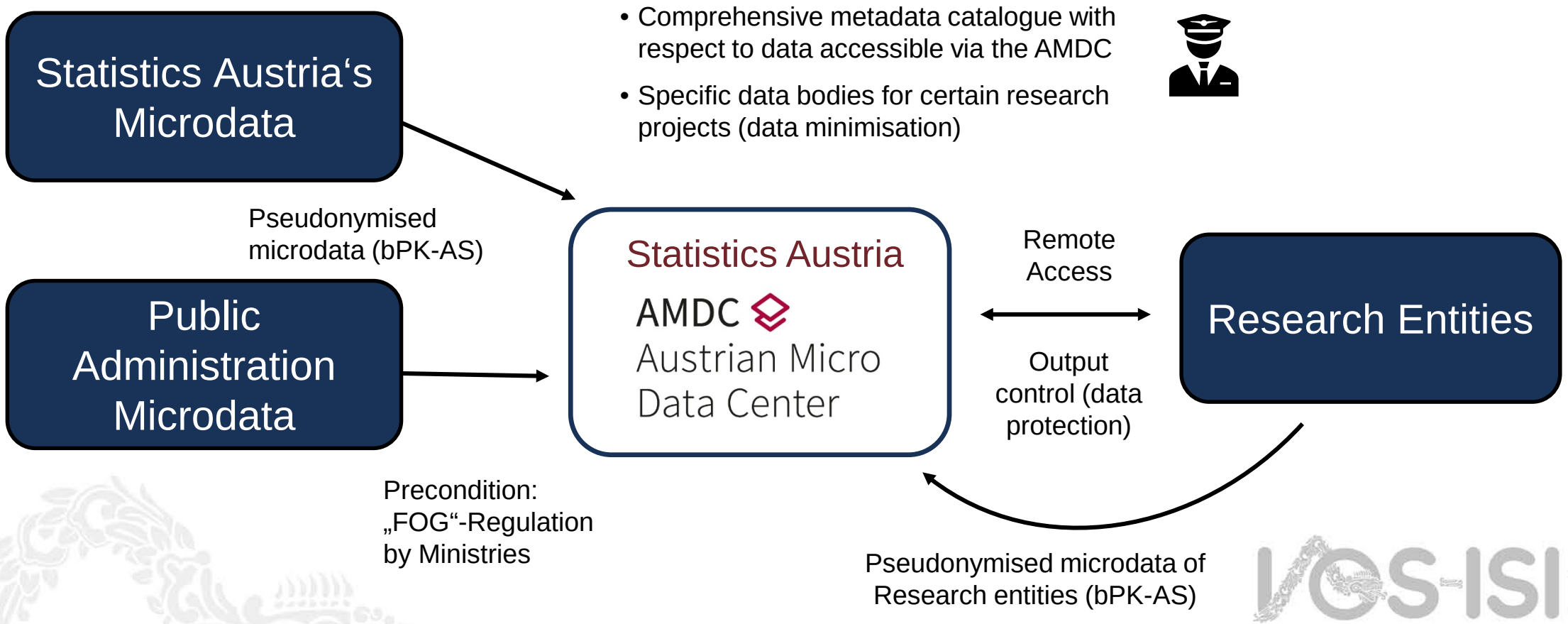
Dissemination of statistical results breaks new ground

Reacting to new demands in accessing information

- More Social Media presence
- Higher accessibility via API
- Improved visualisation
- Better access to microdata for research
- Extending Geographic Information Systems (GIS)
- Contributing to Data Spaces

IOS-ISI 2024 MEXICO CONFERENCE

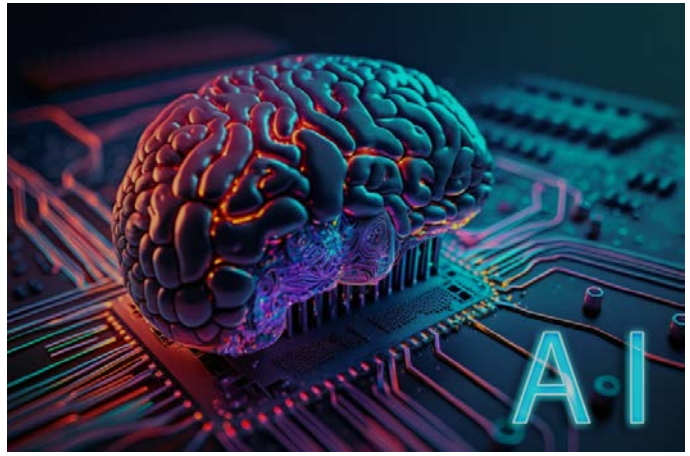
Austrian Micro Data Center creates new access to research data



vbPK-AS = Encrypted sector-specific personal identifier Official Statistics

AI & Machine Learning pave their way

Artificial Intelligence (defined by Microsoft): “An **artificially intelligent computer system** makes **predictions** or takes **actions** based on **patterns in existing data** and can then **learn from its errors** to **increase its accuracy**.”



Machine learning has made great progress in recent years, especially due to **increasing availability of large amounts of data** and **high computing power**. In machine learning processes, an algorithm learns to complete a task independently through repetition. Unlike conventional algorithms, no solution path is modeled.

IOS-ISI 2024
MEXICO CONFERENCE

AI-Top 10 challenges require action

Challenge	Need to act by an NSI
Implementation strategy	Clear roadmap
Unclear goals	Strategic approach
Lack of trust	Communication to users
Lack of data	Improving accessibility to sources
Lack of understanding	Investment in know-how
Privacy concerns	Enhance confidentiality methods
Unreliable results	Careful assessment of results
Bias in algorithms	Enhancing transparency
Processing power requirements	Investment in IT
Technical difficulties	Investment in infrastructure

Inspired by: Dataconomy (https://dataconomy.com/)

AI utilized at Statistics Austria

Machine Learning

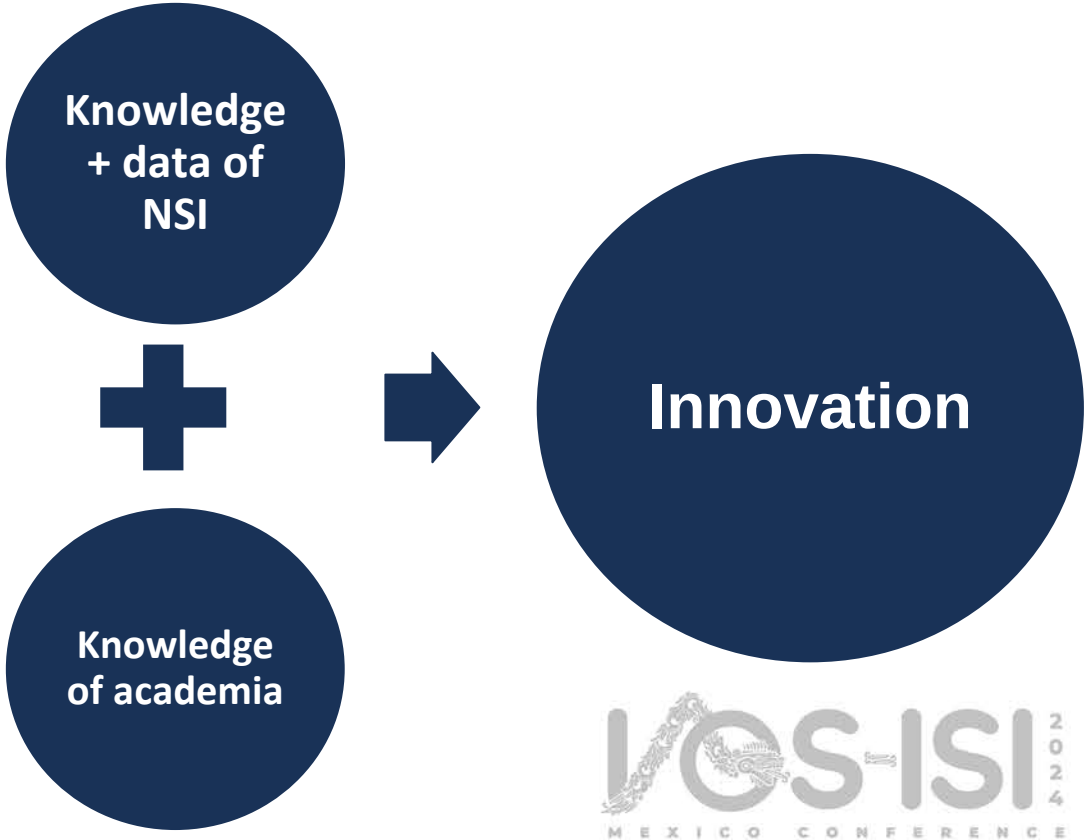
- Using **web data** for statistical purposes
- Imputation for the whole population frame in order to optimize sample designs
- Classifying land use by having **earth observation** images as input
- Synthetic data generation by using **ML-models**

AI beyond ML

- Development of R Code by **ChatGPT**
- AI-based tool for observing media appearence
- **DeepL** for translating texts

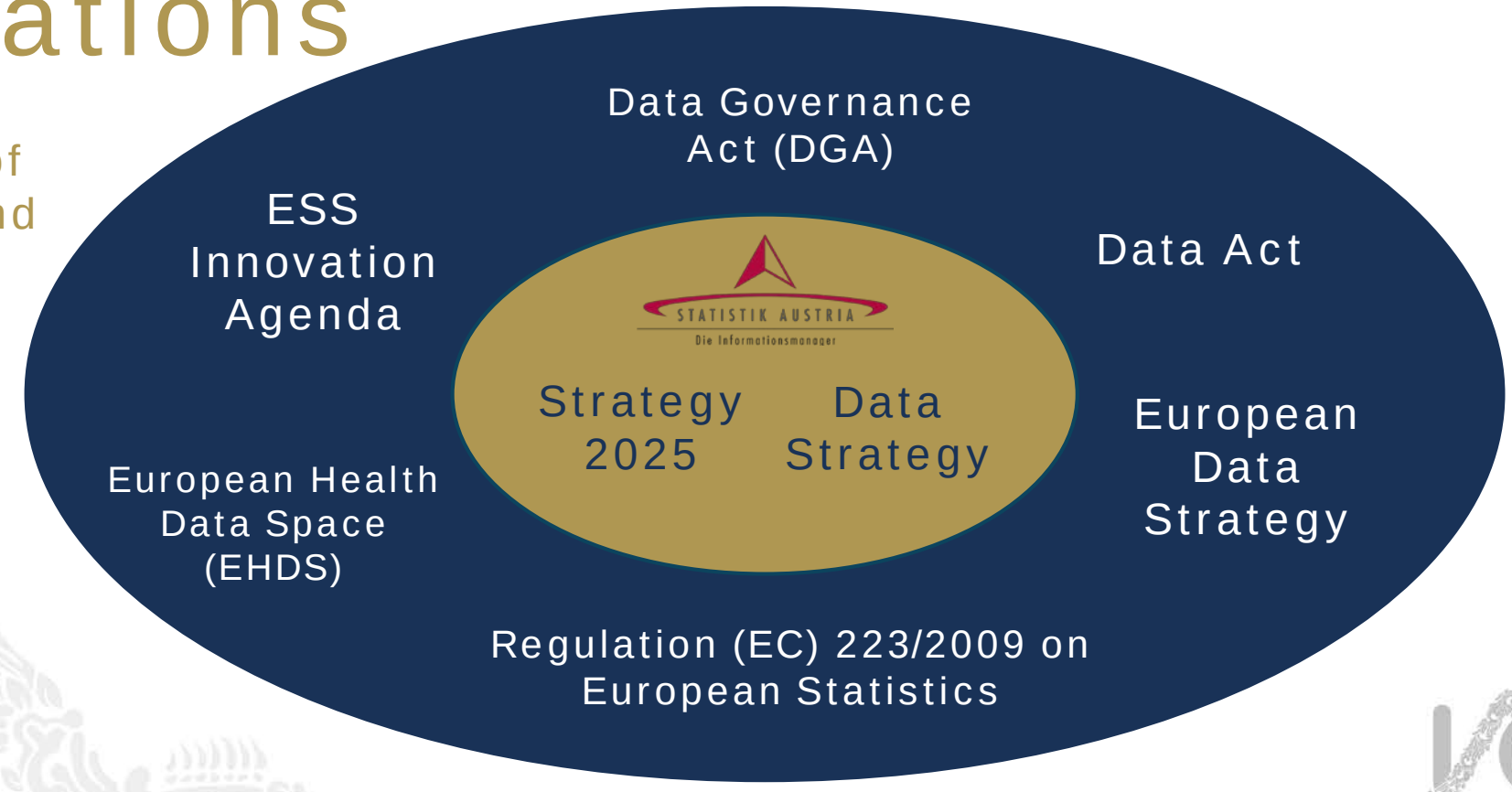
Cooperation with the scientific community accelerates innovation

- Common research projects
- Research infrastructures
- Job rotation between NSIs and academia
- Use of Micro data
- ...



Innovation takes place within the framework of national and EU regulations

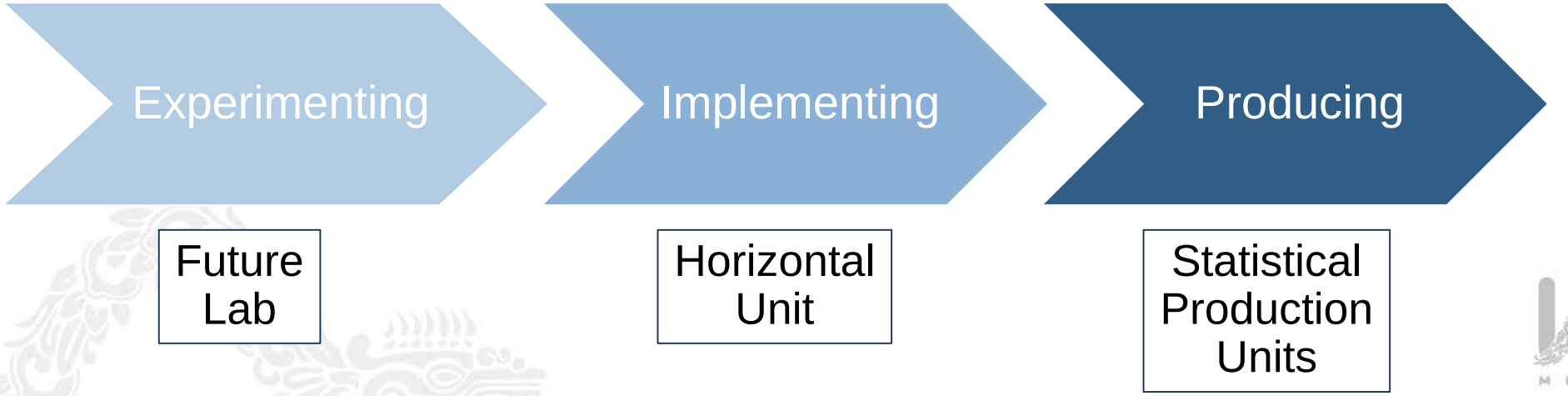
A selection of strategies and legal texts...



IOS-ISI 2024 MEXICO CONFERENCE

Future labs can foster innovation within NSIs (food for thought)

- Future lab as a flexible think tank
- Right to fail
- Taking on board new developments
- Sandboxing
- Experimental projects
- Implementation of innovation in the production units is challenging



IOS-ISI 2024
MEXICO CONFERENCE

The best is yet to come!

INEGI

International Statistical Institute

IAOS
IMPROVING OFFICIAL STATISTICS

IAOS-ISI 2024
MEXICO CONFERENCE

Innovative use of new types of data sources

Some examples from Hungary

Dr. Áron Kincses
President, Hungarian Central Statistical Office

INEGI International Statistical Institute IAOS IMPROVING OFFICIAL STATISTICS

Agenda

1. Data strategy
2. Online cash register data in statistical data production
3. Project on interlinkage of statistical and admin databases

IOS-ISI 2024 MEXICO CONFERENCE

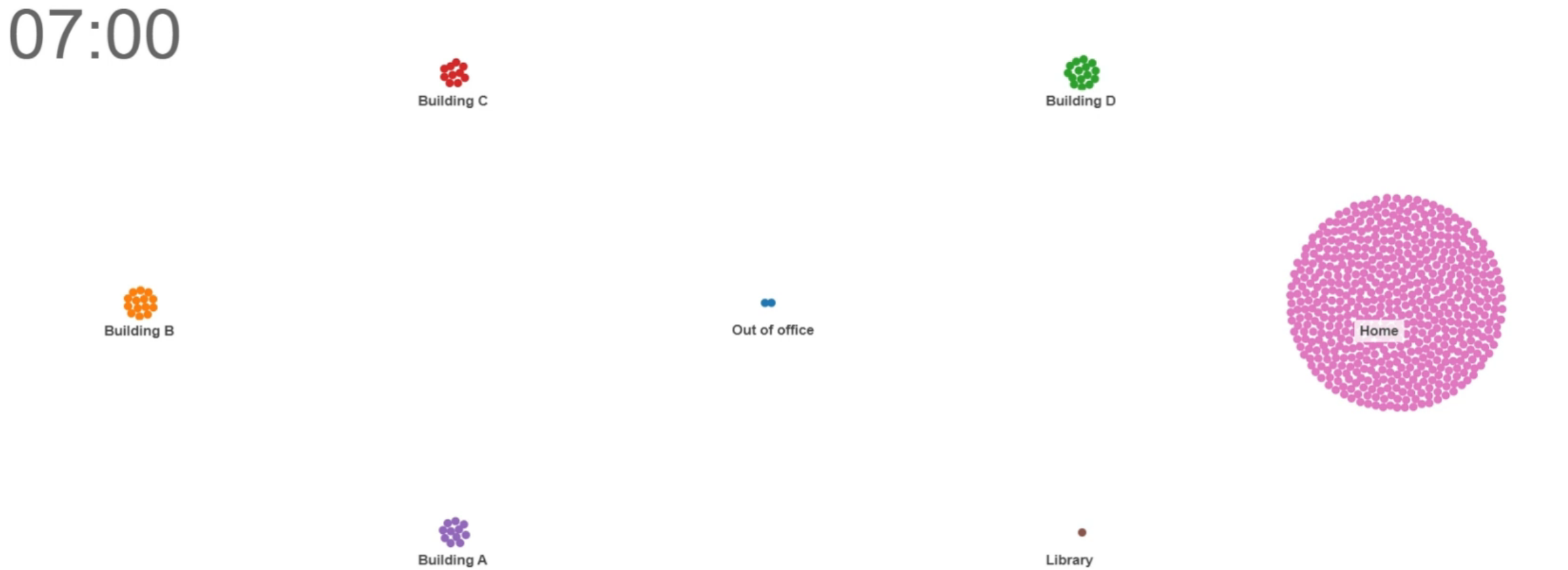
Statistics at crossroad

need for a new data strategy

- Increase relevance of statistics
- Reduce time of dissemination
- Enhance importance of official statistics for policy and economic decision making
- Use more administrative and alternative data instead of traditional, collected data
- Improve efficiency and timeliness through machine learning and artificial intelligence
- Collaborate at international level, and with universities and academia

IOS-ISI 2024 MEXICO CONFERENCE

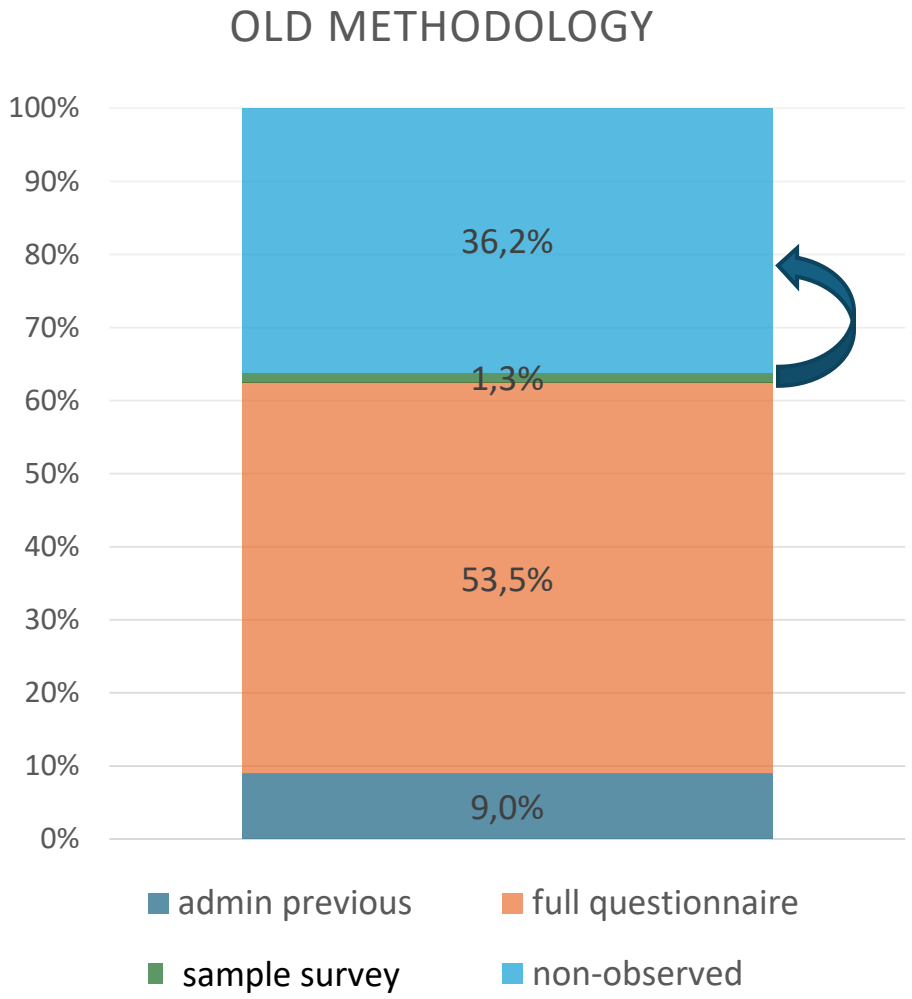
One working day of my colleagues

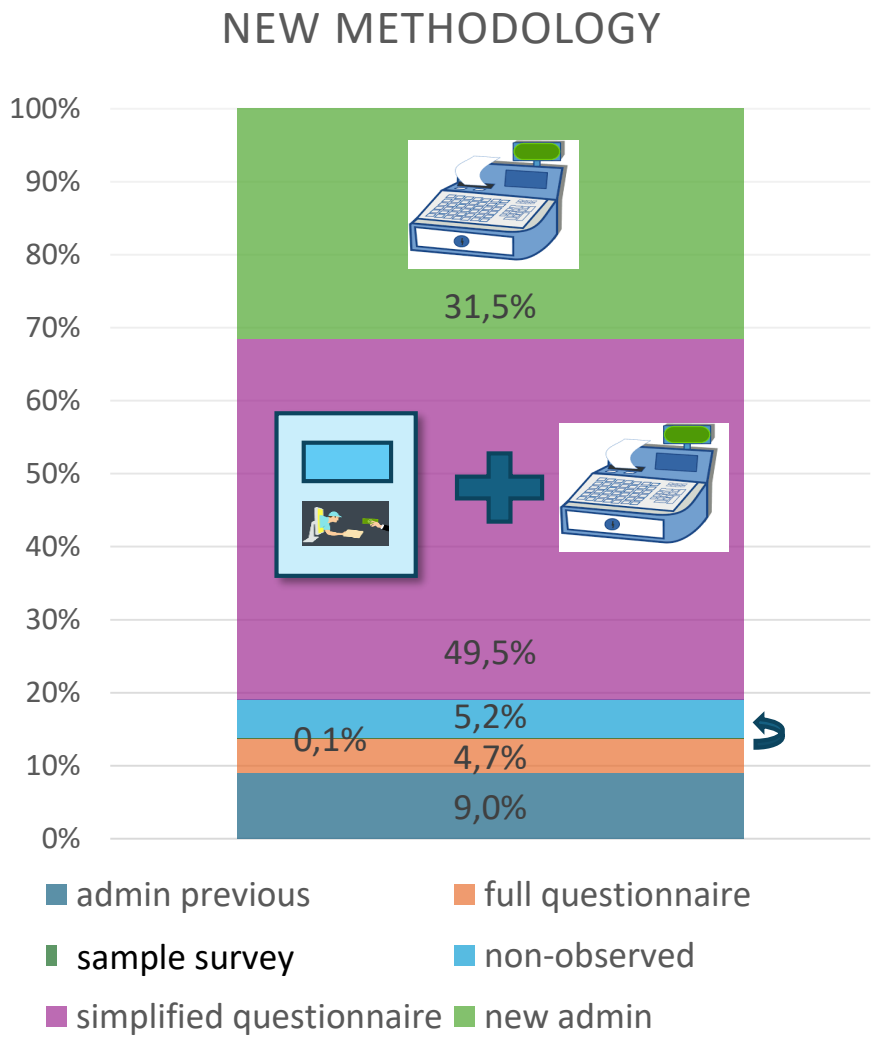


Use of online cash register data to estimate retail turnover

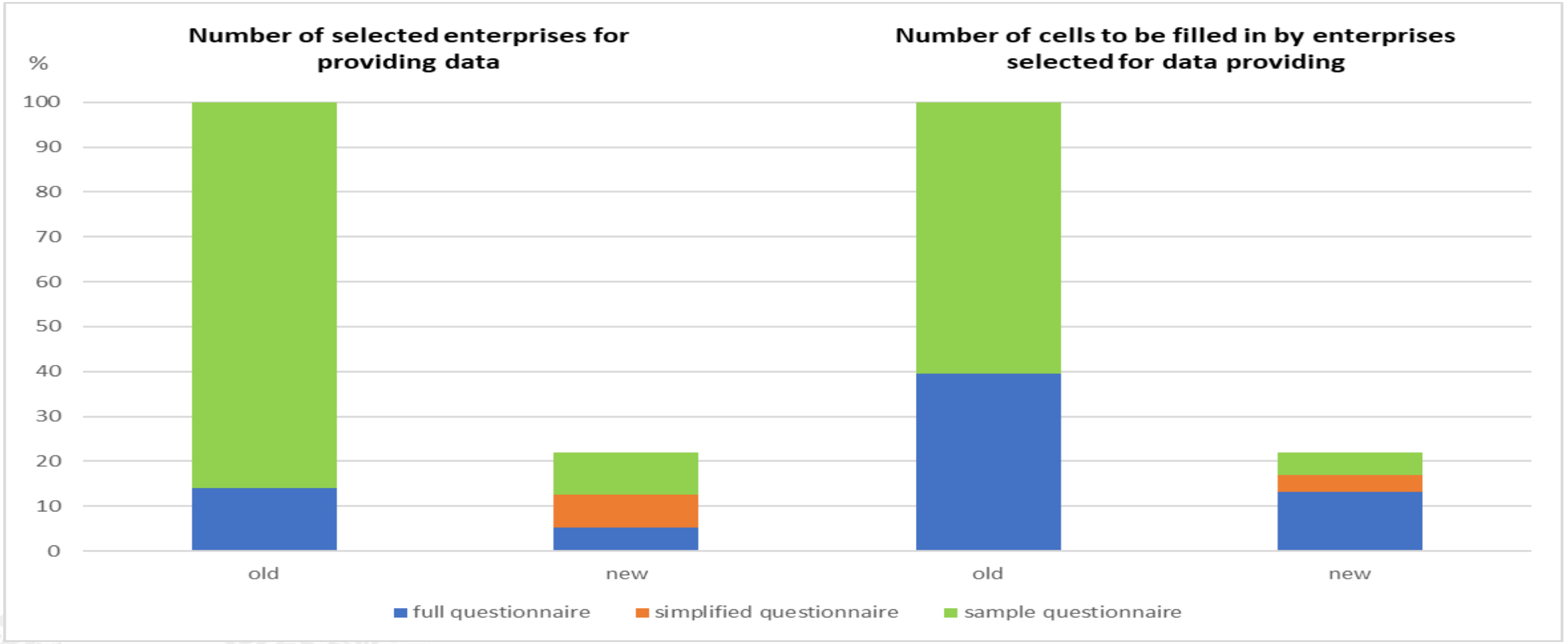
- From 2014 legislation requires businesses operating cash registers to provide real-time sales data to the National Tax and Customs Administration (NAV)
- Access to real-time retail sales data from the NAV database

Population and sampling

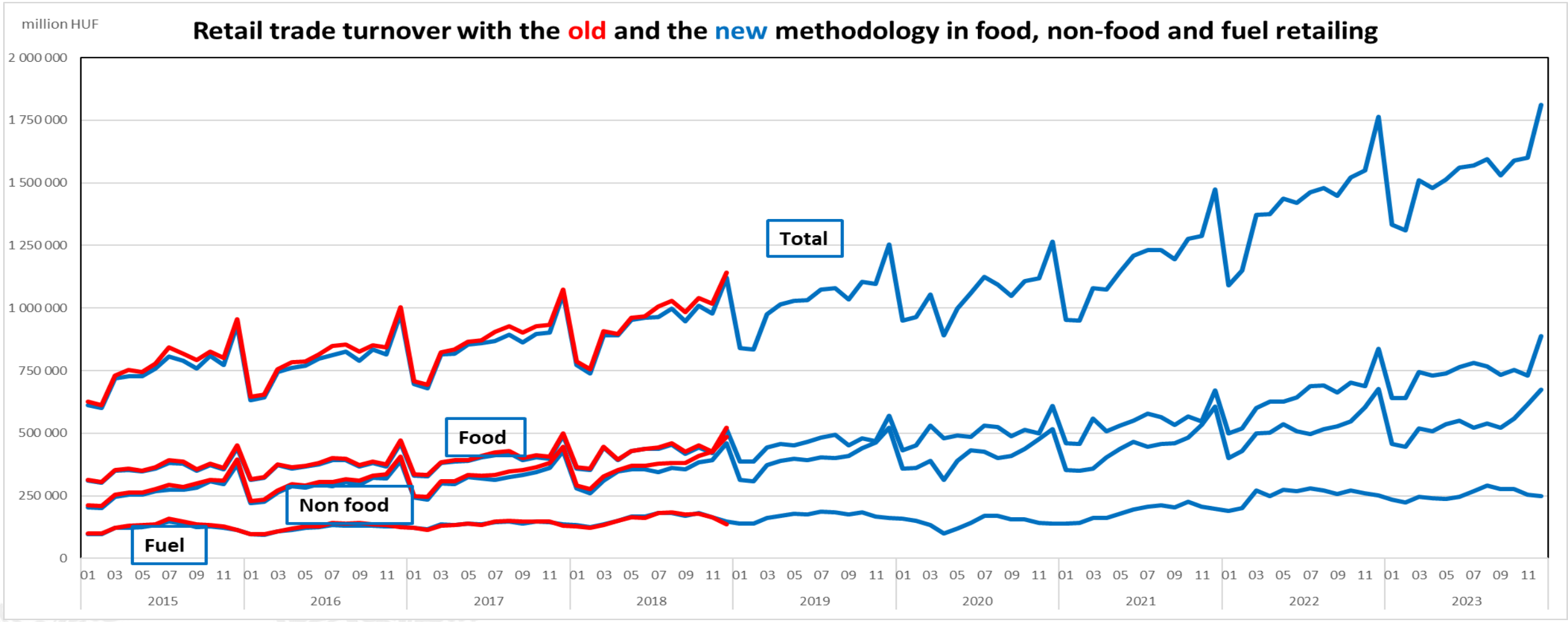




Respondents' burden

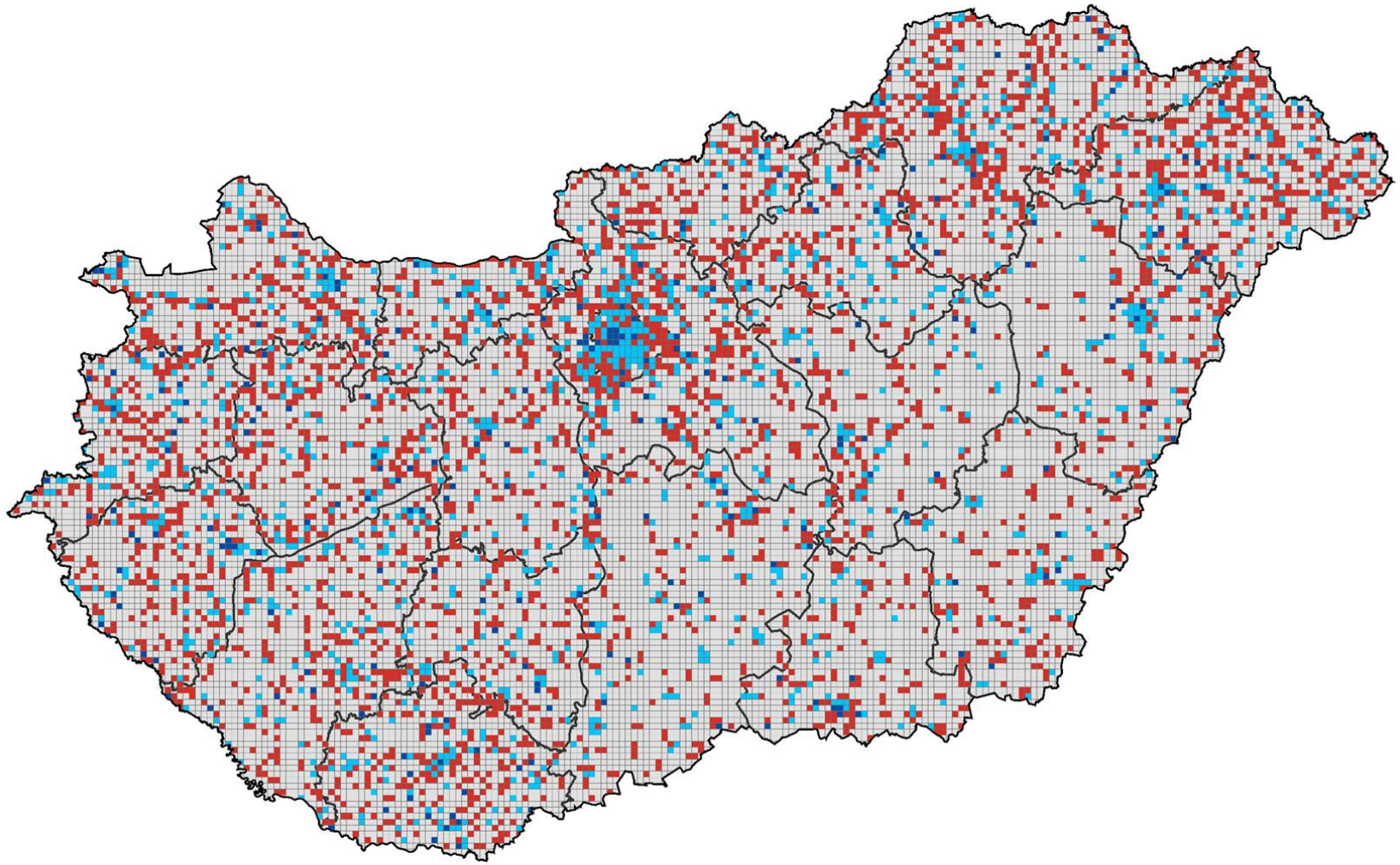


Results


million HUF
Retail trade turnover with the old and the new methodology in food, non-food and fuel retailing
2 000 000
1 750 000
1 500 000
1 250 000
1 000 000
750 000
500 000
250 000
0
Total
Food
Non food
Fuel
01 03 05 07 09 11
2015
2016
2017
2018
2019
2020
2021
2022
2023

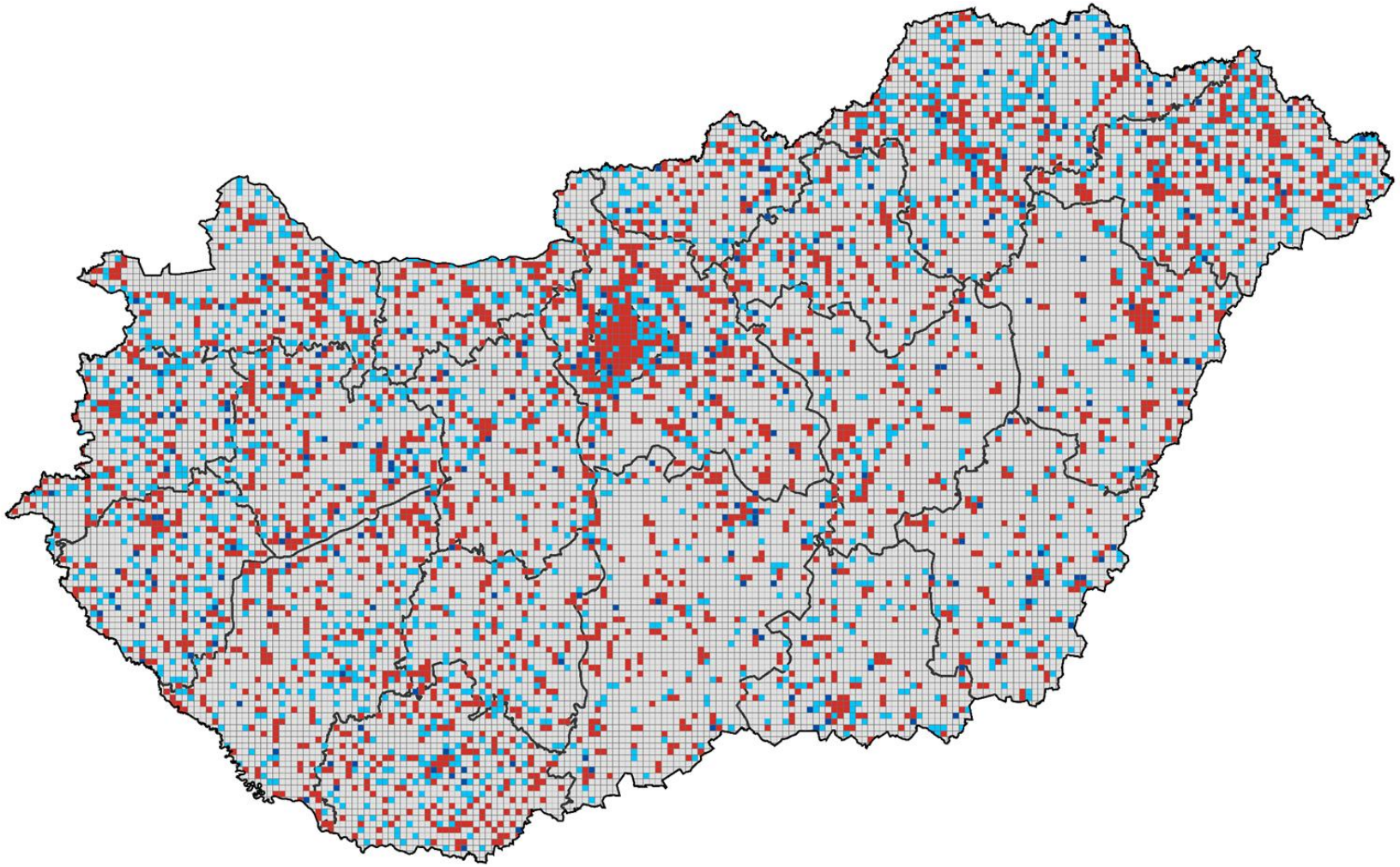
Value index of retail trade in Hungary, April 2020

(same period of previous year=100)



Value index of retail trade in Hungary, April 2021

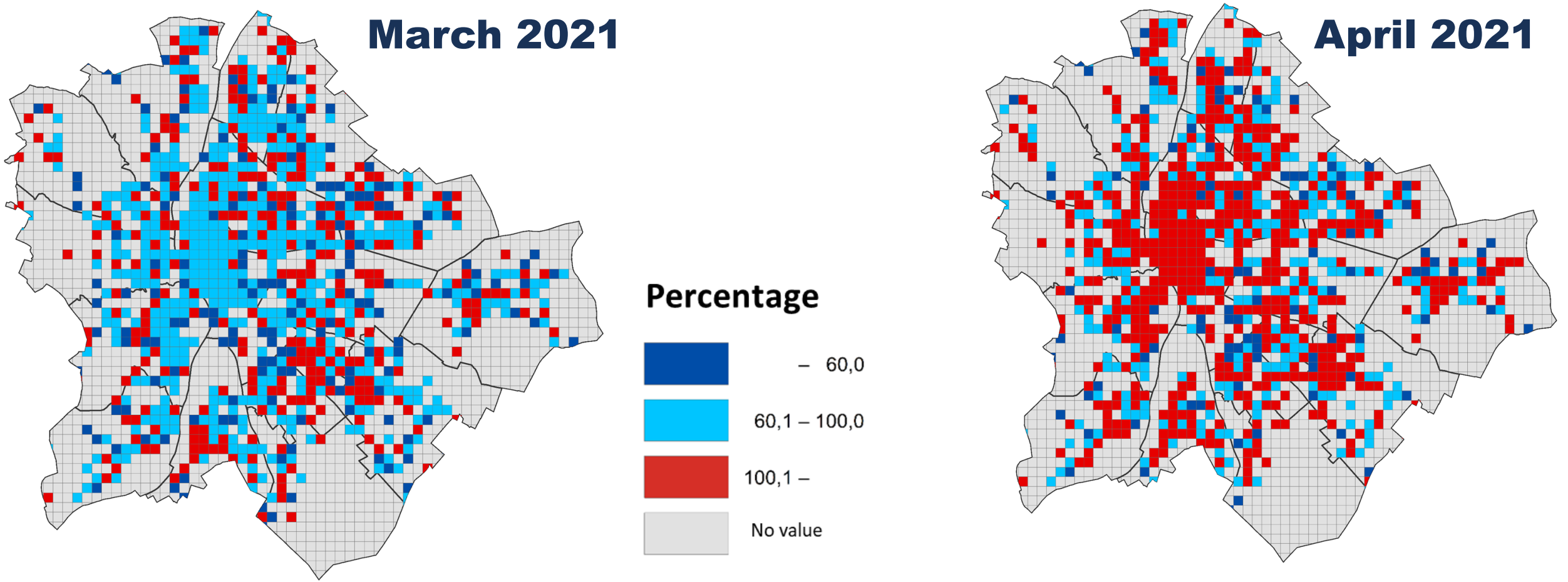
(same period of previous year=100)



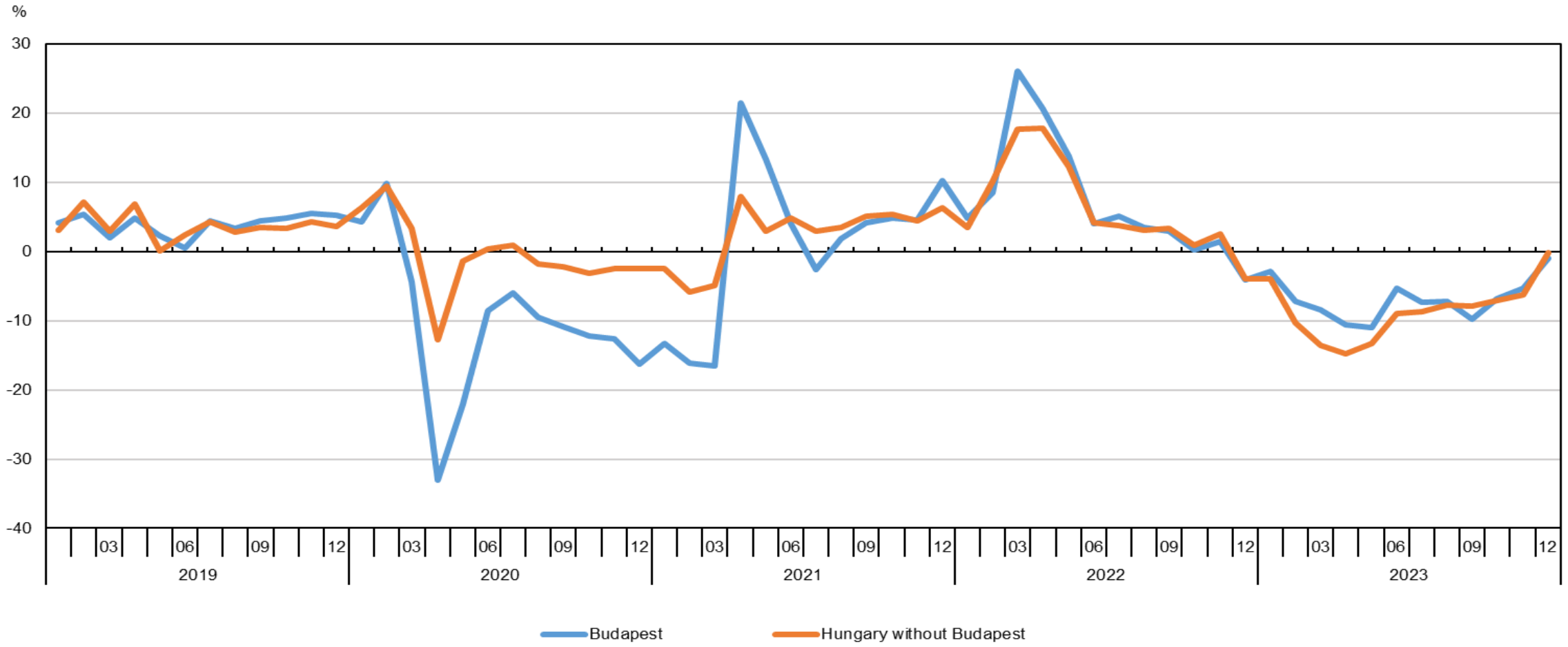
IOS-ISI 2024
MEXICO CONFERENCE

Value index of retail trade in Budapest

(same period of previous year=100)

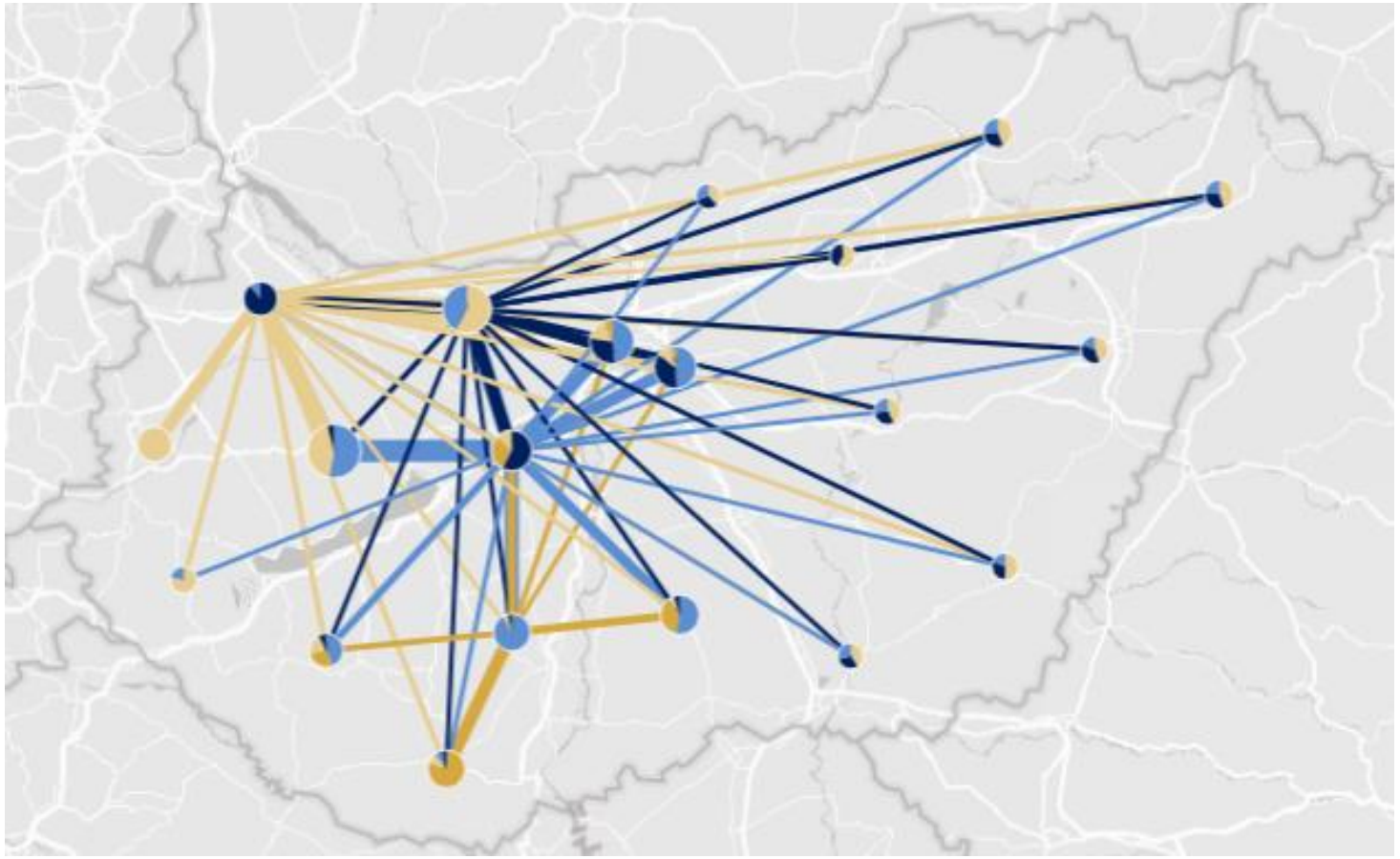


Monthly change of volume of retail trade in Budapest and the rest of Hungary (same period of previous year=100)



Virtual Hungary (VIMA) project

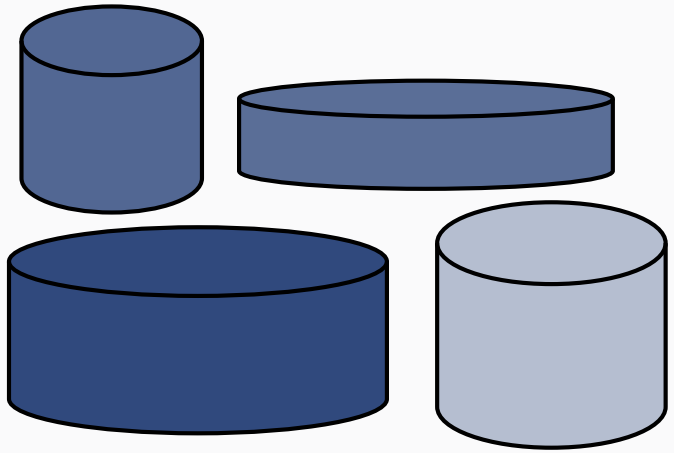
- linking available databases, administrative data at individual level
- new possibilities for analysis, information exploration
- complex access to new contexts and phenomena



Virtual Hungary (VIMA) project

SEPARATE DATABASES

- censuses (population census)
- administrative data
  (tax office, social care…)
- registers (Business register)
- large-sample data collections
  (labour force survey, income and living conditions...)



Anonymization

Linking

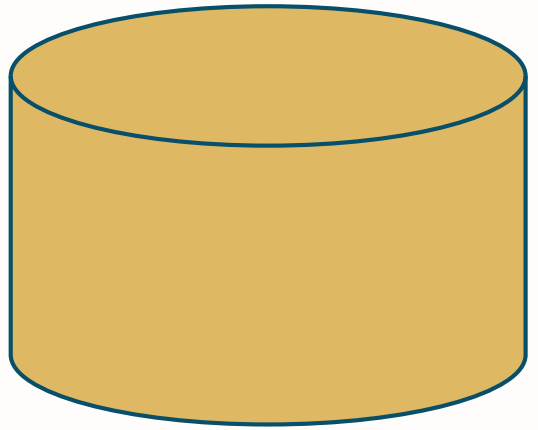
Imputation, extrapolation

Modern techniques

Statistical matching, Automatization, Micro-simulation, Machine learning, Deep learning

LINKED DATABASE

- anonymized database
- individual or low aggregation level
- continuously updated data

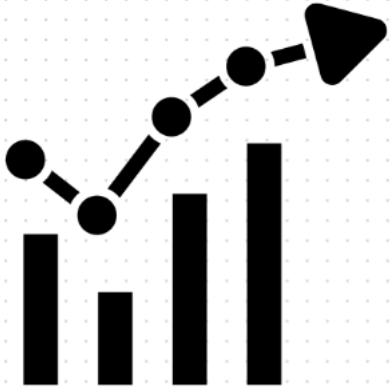


Virtual Hungary (VIMA) project

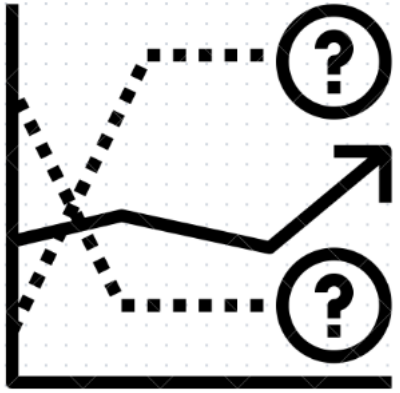
- Up-to-date

- Granularity

- Predictive ability

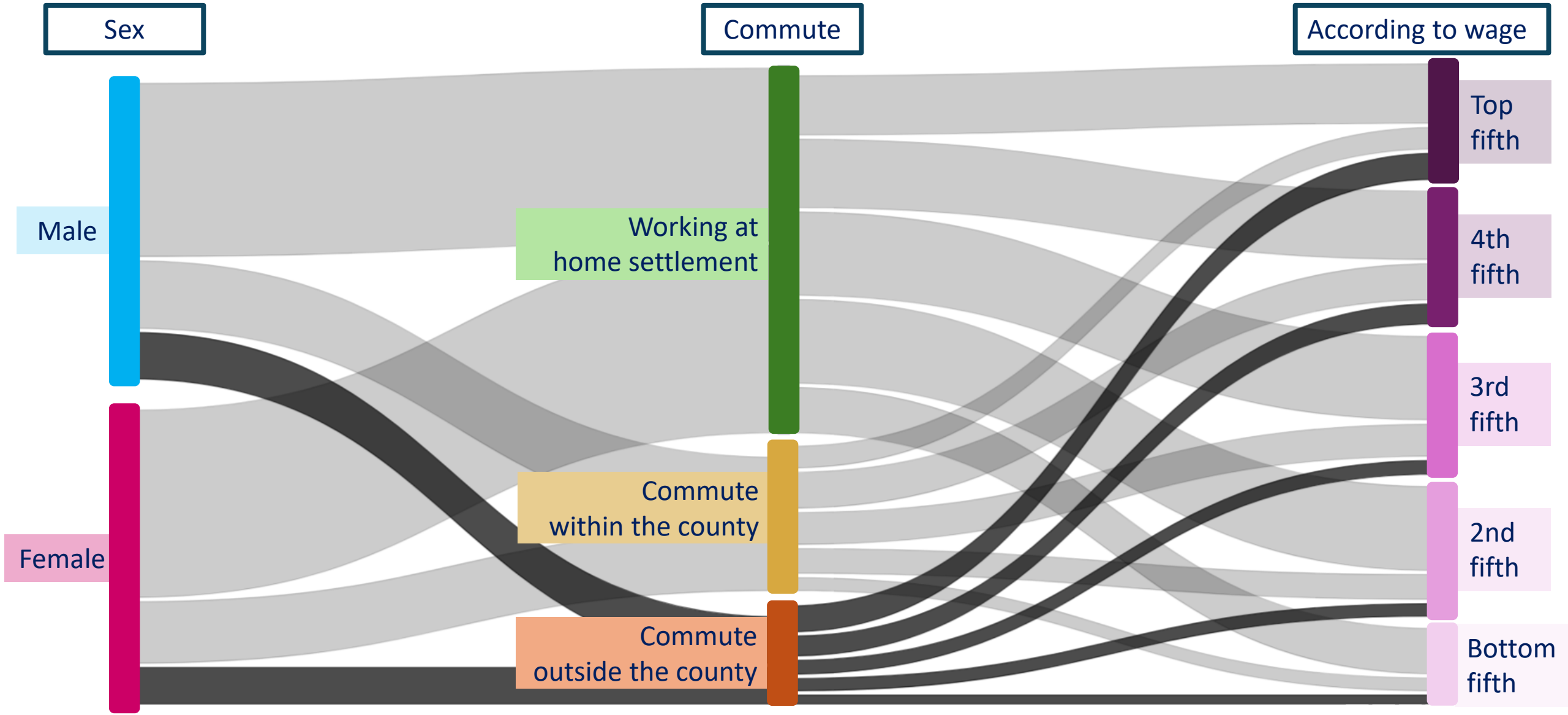


- Scenario analysis

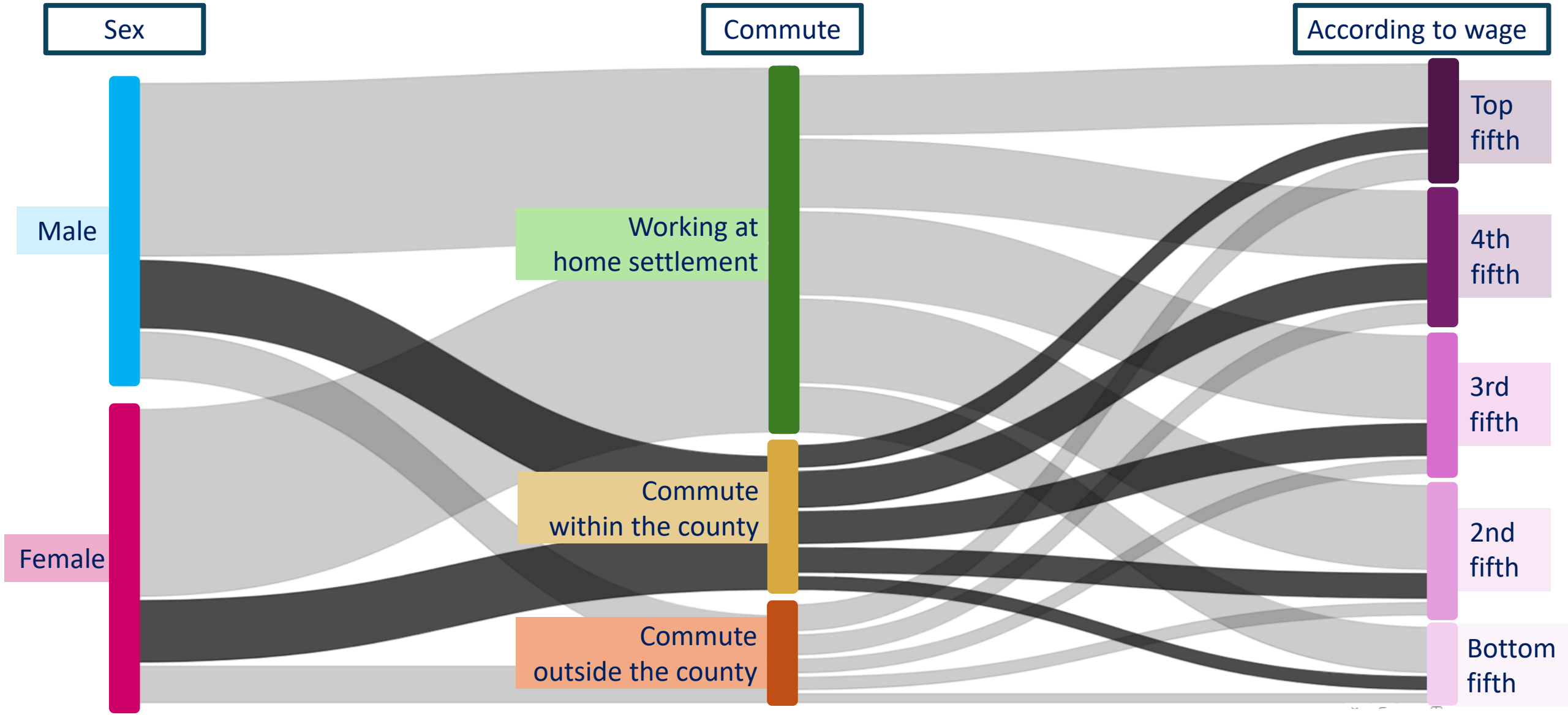


By linking databases, significantly more information can be revealed than the sum of the information contained in the original databases.

Interlinkage of variables from different databases



Interlinkage of variables from different databases



Conclusions

- Benefits of alternative data sources:
  - Reduce the administrative burden on enterprises
  - Increase accuracy of official statistics
- Enhance relevance and fill data gaps with new data sources:
  - Interlink previously separate statistical domains
  - Increase frequency of statistics
  - Provide statistics on new domains
- Global Data Strategy:
  - Emphasize need for effective data strategy at both the Member States and international level

IOS-ISI 2024 MEXICO CONFERENCE

IOS-ISI 2024
MEXICO CONFERENCE

Thank you!

Dr. Áron Kincses
President, Hungarian Central Statistical Office

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

IOS-ISI 2024
MEXICO CONFERENCE

Challenges and innovations in data:

new privately-held data sources in official statistics and examples of innovations from Slovakia

Statistical Office of the Slovak Republic

INEGI

International Statistical Institute

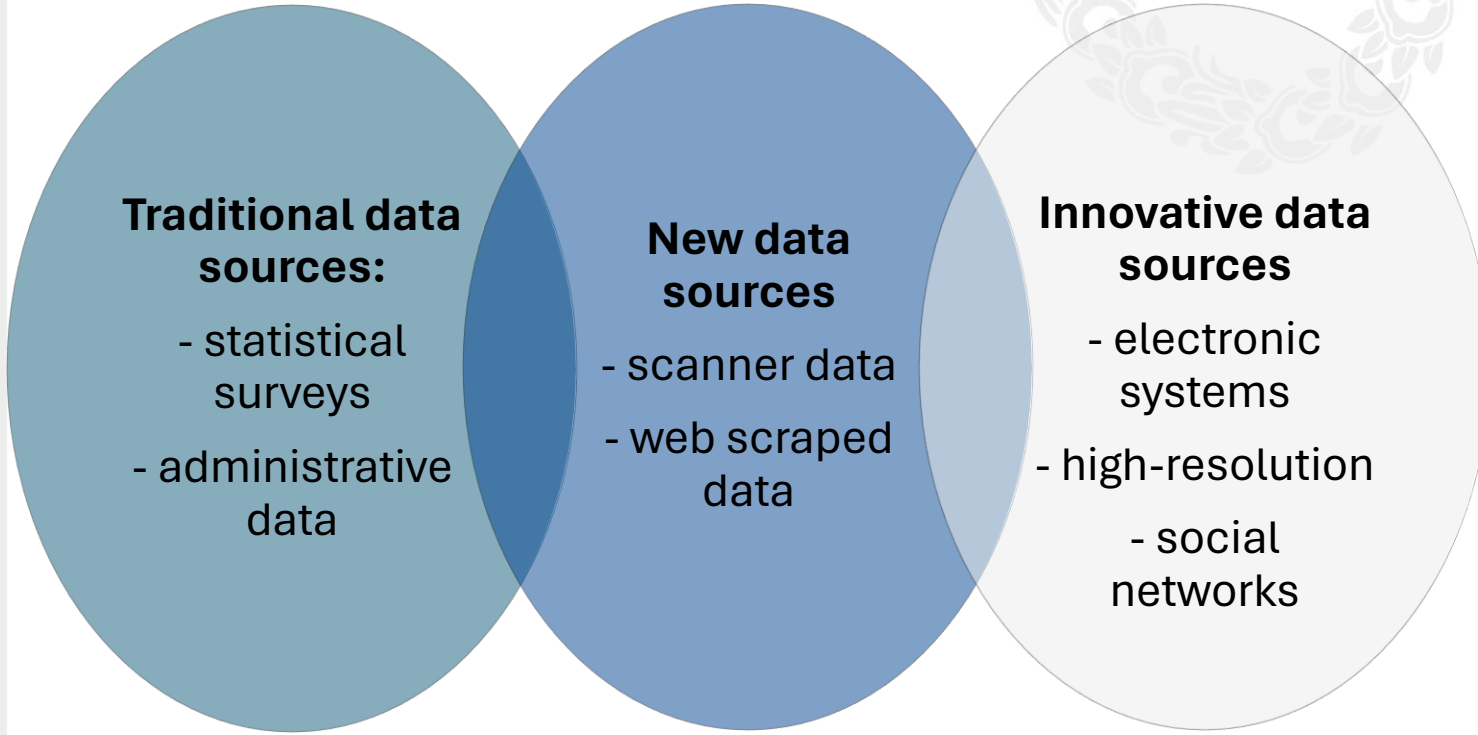
IAOS
IMPROVING OFFICIAL STATISTICS

Agenda

- New data sources and transformation of official statistics
- Innovative statistics – project from Slovakia
- Artificial Intelligence – who is the owner of data?
- Concluding remarks

IOS-ISI 2024
MEXICO CONFERENCE

Data sources in official statistics:

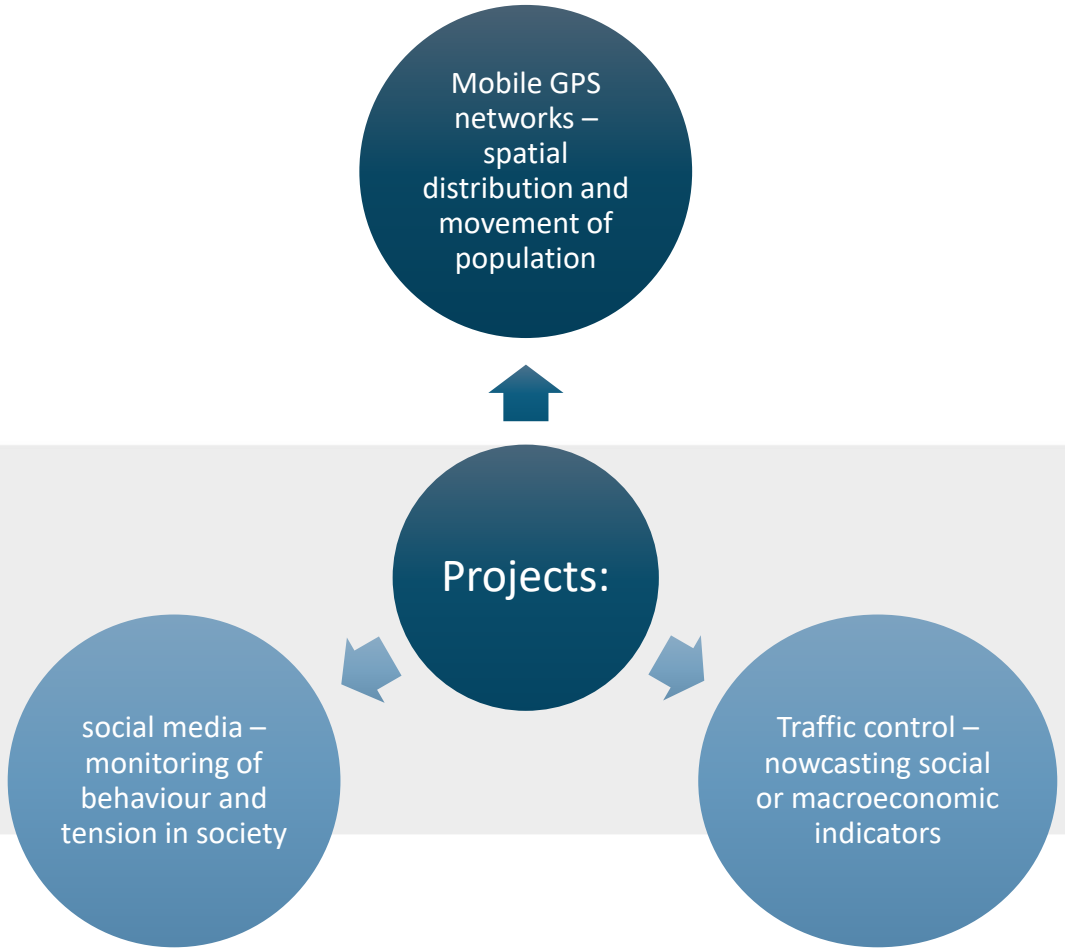


IOS-ISI 2024
MEXICO CONFERENCE

Transformation of official statistics

„We truly believe that new data sources, which are usually owned by private holders, will lead to tremendous changes in dissemination of information and statistics. If National Statistical Institutes (NSIs) do not follow the trend in innovations, the other private institutions could take a role of the leader in producing new reliable, even official, statistics. NSIs must become more agile by researching and implementing new data sources into official and experimental statistics, where an active collaboration with a research community, academia and private sector is of the utmost importance.„

Innovative statistics – projects from Statistics Slovakia



Main objectives of innovative projects (using privately-held data):

- strategic planning for future
- experimental and official statistics
- collaborate with research and academic community
- building relationship with privately-held data owners

- Natural language processing (NLP)
- Machine learning
- Complex statistical models

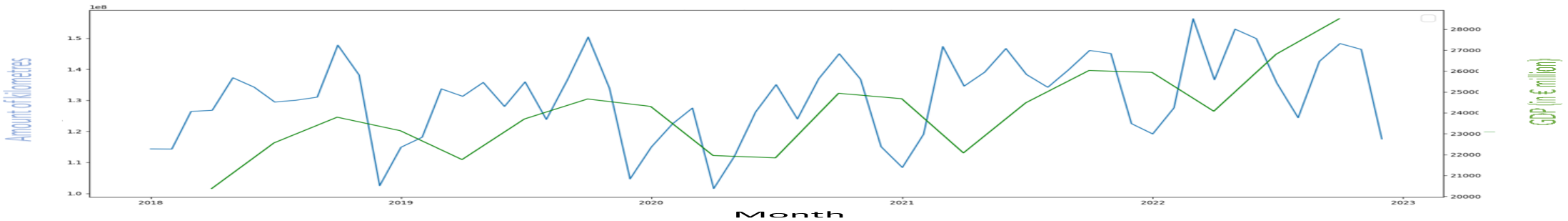
Data sources:
- Electronic systems
- Web-scrapping
- GPS systems

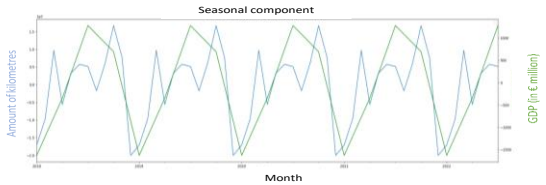
IOS-ISI 2024 MEXICO CONFERENCE

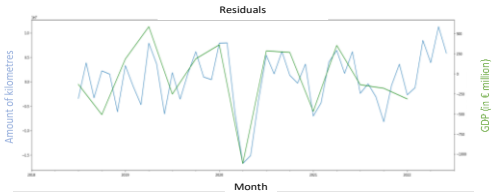
Innovative statistics – Toll data vs Gross Domestic Product

Toll data can be used for nowcasting macroeconomic variables and economic cycles:

- A transport of goods and commodities through highways is usually a strong indicator of the economic performance
- A monthly change in the transport can show a downturn or upturn in the overall economy or outputs of industries that rely on the road transport





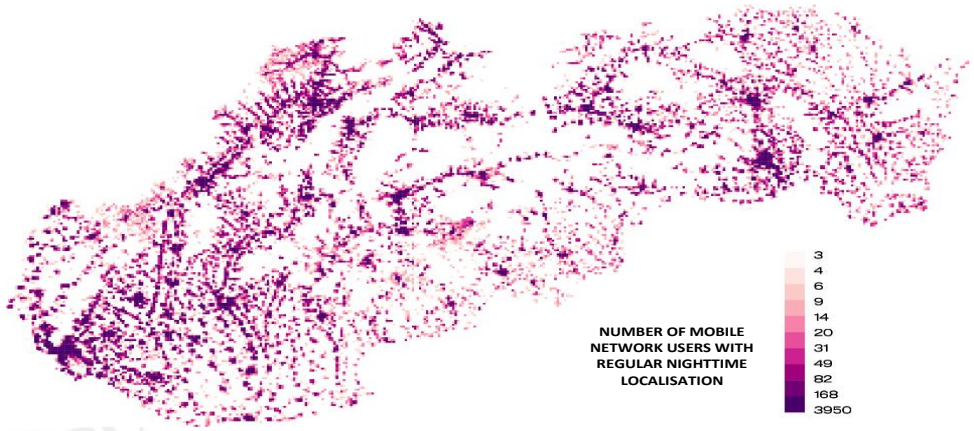


Innovative statistics – Mobile network operators

MNO data have a great potential utilisation for experimental statistics:

- MNO data can provide information on flows of people, goods and information – the public administration needs accurate and detailed data on the behaviour of the population in time and space
- MNO data cannot replace conventional population census data, but can serve as a supplement that tracks other concepts such as daytime and nighttime movements of population

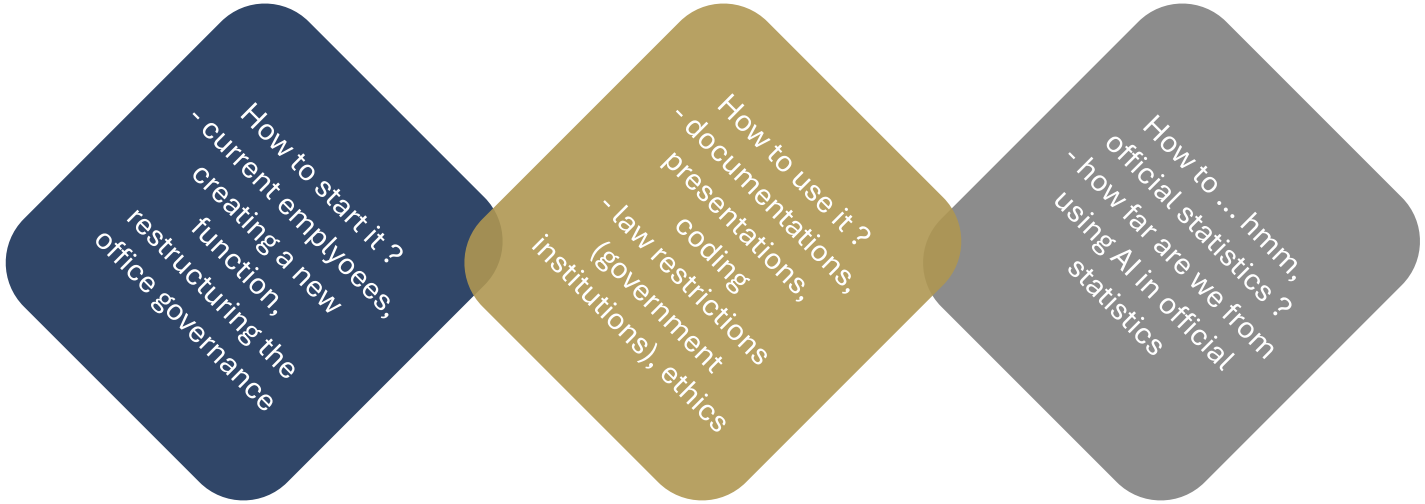
Spatial distribution of mobile network users with regular nighttime localisation on a grid 1x1 km – Jamuary 2023, Slovakia



Visualisation of movement flows between regular nghttime and daytime locations of mobile network users – colors of the flow correspond to the destination city

IOS-ISI 2024 MEXICO CONFERENCE

Artificial Intelligence in official statistics – who is the owner of data?



EU Legislation in Progress – Artificial Intelligence Act (June 2023)

AI regulatory approach in the world: While the USA had initially taken a lenient approach towards AI, calls for regulation have recently been mounting. The Cyberspace Administration of China is also consulting on a proposal to regulate AI, while the UK is working on a set of pro-innovation regulatory principles. At international level, the OECD adopted a (non-binding) Recommendation on AI in 2019, UNESCO adopted Recommendations on the Ethics of AI in 2021, and the Council of Europe is currently working on an international convention on AI. Furthermore, in the context of the newly established EU-US tech partnership (the Trade and Technology Council), the EU and USA are seeking to develop a mutual understanding on the principles underlining trustworthy and responsible AI. EU lawmakers issued a joint statement in May 2023 urging President Biden and European Commission President Ursula von der Leyen to convene a summit to find ways to control the development of advanced AI systems such as ChatGPT.

https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

IOS-ISI 2024 MEXICO CONFERENCE

Concluding remarks and questions

IOS-ISI 2024
MEXICO CONFERENCE

Thank you and see you soon

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS