ESSnet TSS-METH-TOO

Trusted Smart Statistics: methodological developments based on new data sources

Expectations from Eurostat

Fabio Ricciato
Eurostat - Unit A5 'Methodology; Innovation in Official Statistics'

ESSnet TSS-METH-TOO Kick-off meeting
Rome, 5. December 2023

Why this ESSnet project is important?

- *Focus on MNO data, but results will **show the way** for all other "big data" sources (!)*

- *Focus on methodological work content-wise, but results have clear **strategic implications** context-wise*
  - **How NSIs will approach "Big Data" sources**
  - **How Private Data Holders and NSI will position to each other**

What is found **inside**
will have an impact **outside**

Source: https://www.pexels.com by Nicole Michalou

European Commission

Mind the inter-dependencies
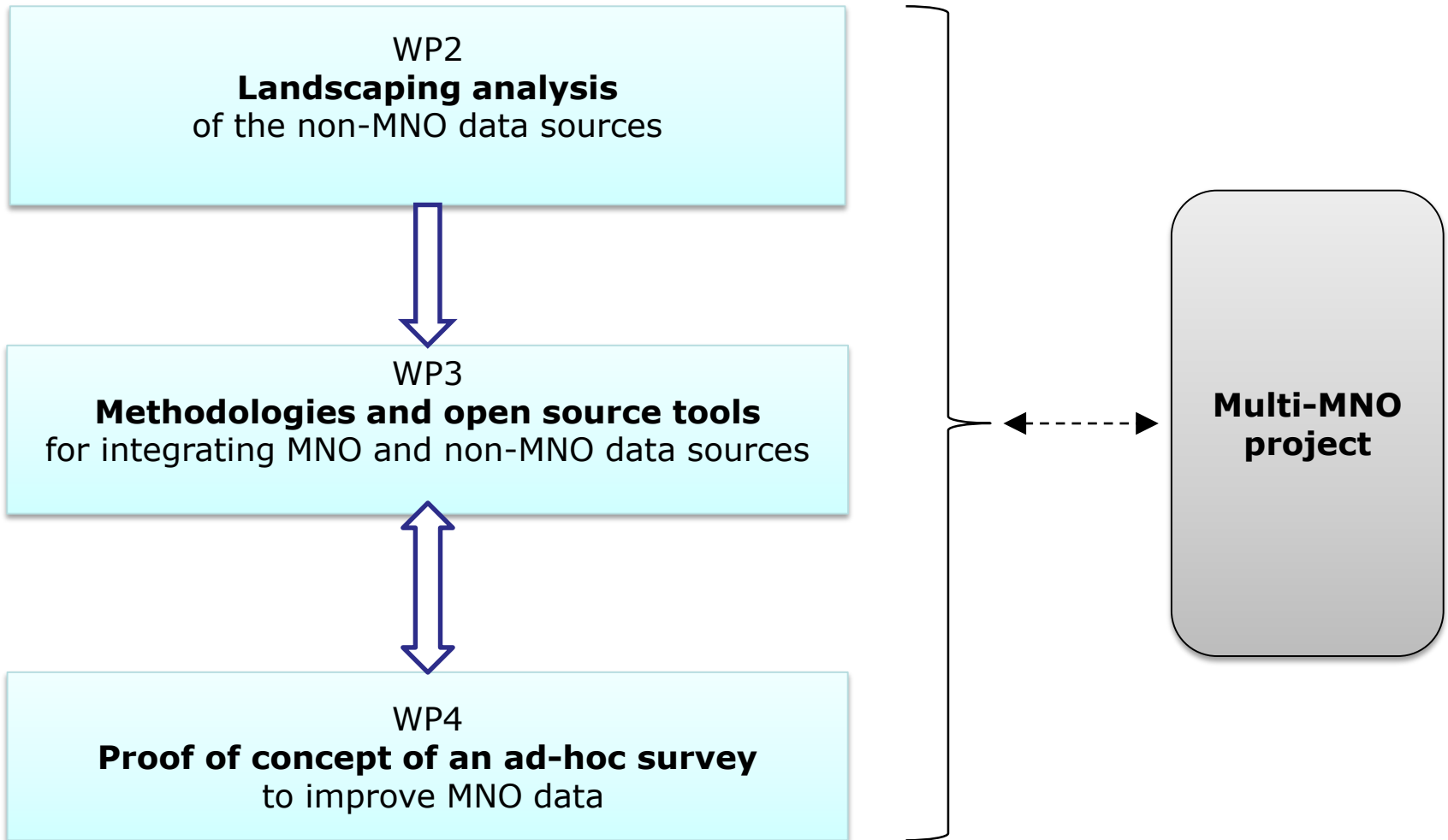
WP2 – Landscaping analysis … (1/2)

- *Two large categories*
  - **Statistical data**, collected by NSI primarily for statistics (census, surveys)
  - **Non-statistical data**, collected primarily other purposes and then reused for statistics: adminstrative data, "big data", etc.

- *Consider feasibility and costs*
  - Costs of **acquisition** (from external data holders) and <u>marginal</u> cost of **collection** (for statistical data)
  - Cost vs quality trade-offs
  - Consider sustainability of costs for <u>regular production</u> of official statistics, not merely experimental stats

- *Mind the European scope*
  - Consider potential availability and costs in <u>all</u> EU countries
  - Availability ≠ accessibility

WP2 – Landscaping analysis … (2/2)

- *Qualitative vs quantitative assessment*
  - When quantitative assessment is not feasible, a qualitative assessment might suffice (e.g., for non-statistical data)
  - Quantitative assessment is expected for statistical data, and particularly for ad-hoc survey (WP4) it is **required**

- *Prioritise Breadth over Depth*
  - Considering as wide as possible set of dimensions (breadth) is more important than going deep into fine-grained assessment in each and every dimensions (depth)
  - Staged approach:
    preliminary analysis → short-list → deeper analysis → shorter list …

- *Is assessment use-case dependent?*
  - Point of research, to be clarified in the project
  - Liaise with Multi-MNO project to ensure choices are coordinated between the two projects

European Commission

WP 3 – Methods and open-source tools (1/2)

- *Ensure consistency within WP3: open-source tools in WP3 that implement the methods developed in WP3*

- *Ensure consistency with WP2: develop methods/tools that work with the non-MNO data selected/short-listed in WP2*

- *Ensure consistency with WP4: develop methods/tools that can be applied to the ad-hoc survey data formulated in WP4*
  - NB: the interdependency is bidirectional: ad-hoc survey in WP4 should take advantage of the methods made available in WP3

- *Ensure consistency with Multi-MNO project*
  - The kind of "MNO data" that are considered in the ESSnet should be aligned to the methodological choices made in Multi-MNO

European Commission

WP 3 – Methods and open-source tools (2/2)

Consider the following classes of methods:

- *Type-I) Methods that require **linking** micro-data between MNO data and non-MNO data*
- *Type-II) Methods that are based solely on **aggregated** data and do not require micro-data linkage*
- *Type-III) **Hybrid** methods, where micro-data linkage can be performed only for a fraction of statistical units but not for all (e.g., based on voluntary consent given by respondents)*

- *WP3 is expected to **cover both Type-I and Type-II** and assess quantitatively the potential benefit of Type-I over Type-II*
  - **Micro-data linkage reduce uncertainty by 5% or 50% or 95% ?**
  - **That matters also for data protection compliance**
  - **Assessment based on formal models, simulations, synthetic data, else? Point for discussion?**
- *Ideally WP3 will develop also hybrid methods of Type-III*
  - More challenging, but also more scientifically intriguing!

European Commission

WP 4 – proof-of-concept of ad-hoc survey

- *Mind the EU Scope*
  - **Design a survey that is prospectively conducted across all EU countries**
  - **If national variations are needed (e.g. due to national specifities in mobile usage) justify and provide guidelines**
- *Outcome #1: survey design*
  - **Cover all design and testing aspect, output should be a ready-to-be-launched sample survey**
  - **Which questions ? Which frame/sample? How to conduct it?**
  - **May consider hybrid approaches, e.g. mix of traditional survey vs mobile app**
- *Outcome #2: quantification of costs €€*
  - **Together with assesment of accuracy gain (to be quantified in WP3) will serve as input for deciding about actually launching it**
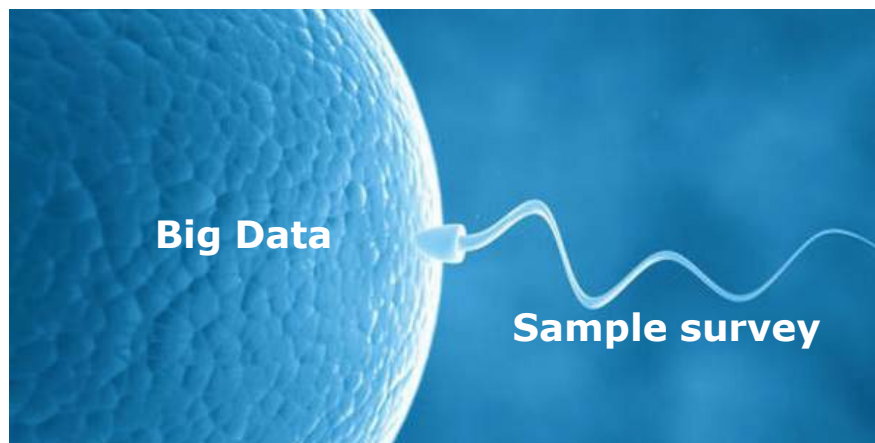
Strategic implications

> **Non-statistical "Big Data"**
> \+ statistical data from purposedly design **sample survey**
> **= high-quality statistics** (that could not be generated otherwise)

- The project will clarify whether the above conjecture holds true or not (for MNO data) and will quantify costs and benefits
- The outcome has impact on NSI-MNO mutual relationship: one-way provider/consumer vs. more balanced partnership based on data contributions from both sides.
- It has also impact on the mutual positioning of survey data and "big data" (replacement or complement?)

European Commission