

BY-COVID Spring 2024 Baseline Use Case Workshop:

Integration of individual-level socioeconomic data for infectious diseases research and prevention in Europe

Workshop participants

Please add your name, role/job title & photo to this board



Enrique Bernal-Delgado
MD, MPH, PhD
Senior scientist, IACS



Nina Van Goethem
PhD
Epidemiologist, Sciensano



Marjan Meurisse
Scientist / PhD student
Sciensano / UCLouvain



Simon Salliner
Research Data Management
Specialist, DANS



Eduardo A. Bracho Montes de Oca
PhD researcher at UCLouvain, UGent, and Sciensano
Statistician-to-be



Francisco Estupiñán-Romero MD, MPH,
Senior scientist, IACS



Francesco Messina PhD
Bioinformatic scientist,
INMI



Lorenz Dolanski Aghamanoukjan
Data Scientist,
Health Expert,
GDS



Vittorio Grieco
MD
Preventive Medicine resident,
University of Catania

Topic: Exploring socioeconomic data sources in Europe: availability, limitations, and mobilization

Participants: Eduardo Antonio Bracho Montes de Oca, Melisane Lima, Ritu Rani, Lorenz Dolanski-Aghamanoukjan, Christian Ohmann, Vittorio Grieco

Chair: Vasso Kalaitzi

Step 1: Reflect

Spend some time reflecting on each question. Write down your answers to the prompts. If someone's feedback resonates with you, feel free to add a thumbs-up emoji.

Which challenges did you experience when mobilizing and linking sensitive social science data for health research? How have you faced those challenges?	How can we improve the availability and accessibility of sensitive socioeconomic data for EU-level health research?	What are limitations of existing socioeconomic data sources? How can we deal with these limitations in the future?
<p>★</p> <ul style="list-style-type: none"> E.g.: Long data access procedures. We generated synthetic data to allow coding without access to real data. E.g. Use of area-level open data. Lack of use of semantic standards across data sources, partic. in cross border as a non expert in the SSH field, difficult to decide on which indicator(s) and data sources to use Complex request procedures and long procedures / might take months to get data Making choices on which variables are really important data linkage: ID-System in AT is good on a technical level, but processes/governance across institutions is not "up to speed" Data governance does not prioritize research use cases <p>Privacy / data protection issues</p> <p>Solutions usually involve limiting yourself to aggregation</p> <p>Issues with aligning administrative areas for which socioeconomic information is publicly available with healthcare administrative areas (i.e., making sure the populations in the administrative area and the healthcare area match)</p> <p>Data timeliness not always match the rest of the data sources</p> <p>Establishing maximum response times from institutions regarding both the processing a data application up until acceptance of the data request and from acceptance to actual access to data</p> <p>Policies that will define what data researchers need</p> <p>By creating a EU universal secure VRE</p> <p>Perhaps proposing important national project where actors holding the data are involved and agree on sharing in advance.</p> <p>Liaising with the different data governance structures</p> <p>data governance barriers (allowed uses, only aggregated data, ...)</p>	<ul style="list-style-type: none"> E.g.: Standardization Standardization of the data application processes across institutions 	<p>★</p> <ul style="list-style-type: none"> E.g.: Representativeness Accuracy of self-reported data Unclear policies of the owner organisation wrt anonymisation/ps eudonymisation only aggregated data (e.g. municipality level) Solution: Extra admin support for researchers Scattered/fragmented Timeliness of updates (surveys) -> Use of continuously updated administrative sources Siloed information systems and data sources <p>Level of granularity meaningless for the research of our interest</p> <p>Usually cover only limited populations</p> <p>codebooks often missing (discoverability of the data sources)</p> <p>Different owners / holders of different variables</p> <p>methodology on the socioeconomic data capture difficult to find or to access before data request</p> <p>reusing statistics imply methodological dependencies that may limit meaningful research <p>Technical interoperability; metadata schema</p> <p>None existence of official detailed documentation on collection / missing metadata</p> <p>Sampling methods are not relevant/ or may reduce representativeness to usual questions health services and policy research</p> <p>Clear documentation on which procedures can be done when linking data</p> <p>lack of data documentation: e.g. details on exactly what tools of income are reflected in the data and which are not</p> </p>