

MAIN OUTCOMES OF THE INEXDA WORKING GROUP ON STATISTICAL DISCLOSURE CONTROL (SDC)

Ana Esteban

Head of Bank of Spain Data Laboratory (BELab) Unit

12TH BIENNIAL IFC CONFERENCE “STATISTICS AND BEYOND: NEW DATA FOR DECISION MAKING IN CENTRAL BANKS”

Basel

22 August 2024





International Network of Exchanging Experiences on
Statistical Handling of Granular Data

General misión

- Promote data sharing and data access
- Promote G20 Data Gaps Initiative II, in particular recommendation 20, addressing the accessibility of granular data.
- Acknowledging and supporting the work on data sharing of the Irving Fisher Committee on Central Bank Statistics

Organisation

- Current chair: Stefan Bender of the Deutsche Bundesbank.



Why create this WG in INEXDA?



Statistical disclosure control (SDC) is a major challenge faced by research data centers (RDC) when making sensitive microdata available to external researchers.



The goal of SDC methods is to prevent the disclosure of sensitive information of individual units (respondents). This can be achieved with: (1) microdata anonymization, and (2) output control.



Microdata anonymization aims to protect the original microdata before making it available to researchers.



Output control procedures ensure that individual agents cannot be re-identified in the results published outside the research data center

INEXDA WG ON STATISTICAL DISCLOSURE CONTROL

Participants, goals and tasks

Participants



Goals



- Identify the current SDC needs, procedures, and tools used among INEXDA members
- Foster harmonization in the area of SDC within INEXDA
- Identify open challenges in the area of SDC and define a plan to address them.

Tasks



- Conduct a survey among INEXDA members to identify current SDC needs,
- Conduct a technical session to review specific use cases and discuss implementation details.
- Define a plan to address open challenges.

Primary Anonymization



Secondary Anonymization

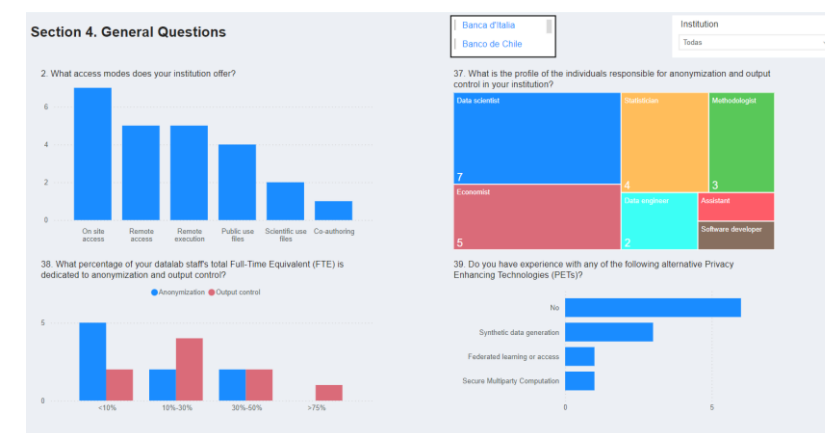


SURVEY CONTENT

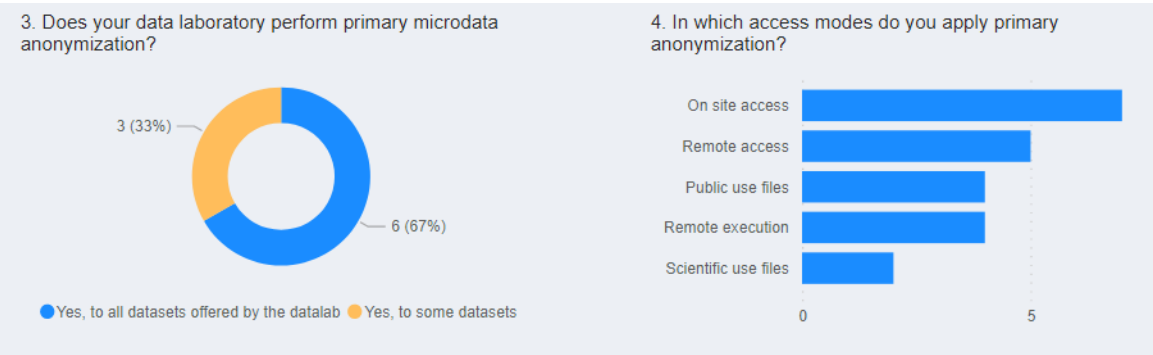
Output Control



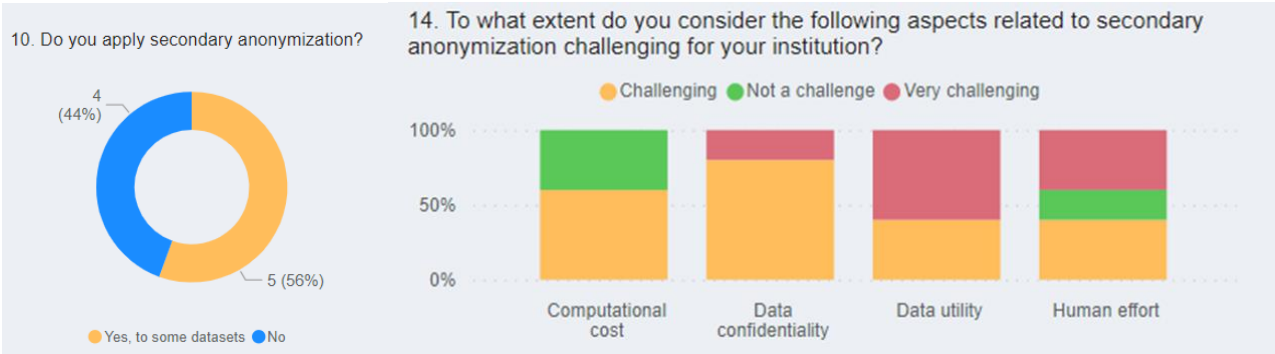
General Questions



Primary Anonymization

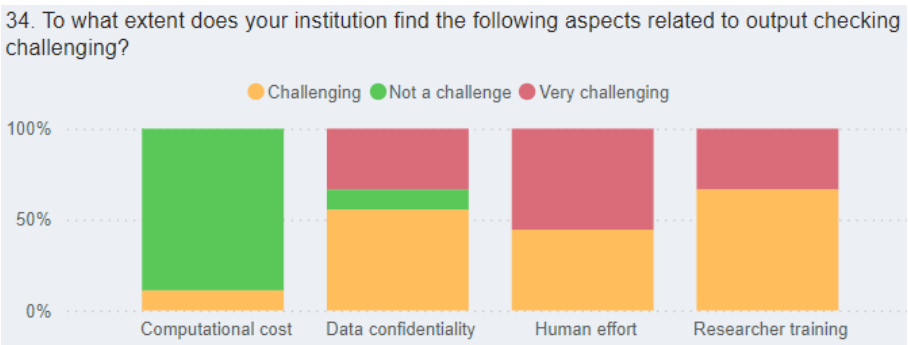
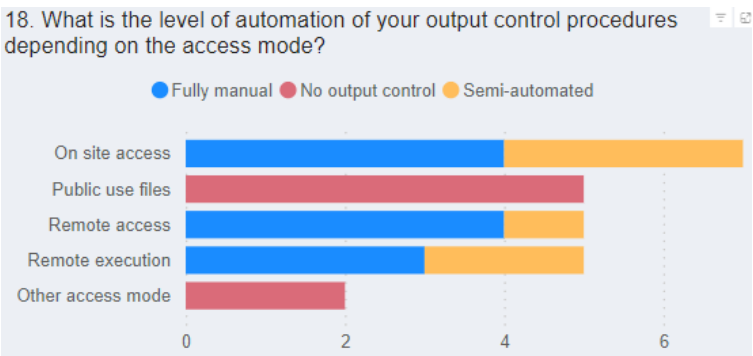


Secondary Anonymization



SURVEY: MAIN RESULTS

Output Control



Eleven SDC use cases were presented by the members of the WG in Madrid, classified into the following topics:



ANONYMIZATION

- Disclosure avoidance in the Spanish Survey of Household Finances (*Cristina Barceló, Banco de España*)
- Anonymization Algorithm for trade transactions (*Claudia Velázquez, Banco de México*)
- Joint analysis of categorical and continuous key variables for secondary anonymization of sensitive microdata (*Eugenia Koblents, Banco de España*)
- Bank of Italy remote execution system (*Daniele Piras, Banca d'Italia*)
- BPLIM's approach to protecting sensitive data (*Joana Pimentel, Banco de Portugal*)



OUTPUT CONTROL

- Tools and resources for output checking at the RDSC of Bundesbank (*Hariolf Merkle and Christian Hirsch, Bundesbank*)
- Automated checking of research output (ACRO) (*Marco Stocchi, Eurostat*)
- Measures to ease code reproducibility for output control (*Ricardo Arcos and Emma Pérez, Banco de España*)
- Output control practices in the Spanish Survey of Household Finances (*Cristina Barceló, Banco de España*)



DATA SHARING AND PRIVACY ENHANCING TECHNOLOGIES (PET)

- Utility and confidentiality assessment of synthetic financial data-Pilot in collaboration with the European Commission (*Eugenia Koblents, Banco de España*)
- Towards a shared European Statistical System infrastructure for collaborative confidential computing (*Fabio Ricciato, Eurostat*)



Questions that a new RDC may inquire ask in relation to SDC and example **answers**, based on the experience of the participants in this WG.

SAFE DATA



- What techniques can RDCs use to maintain data confidentiality?
- Can different anonymization techniques be applied depending on the mode of access to the data?
- Is it necessary to guarantee that data cannot be indirectly re-identified?
- ...

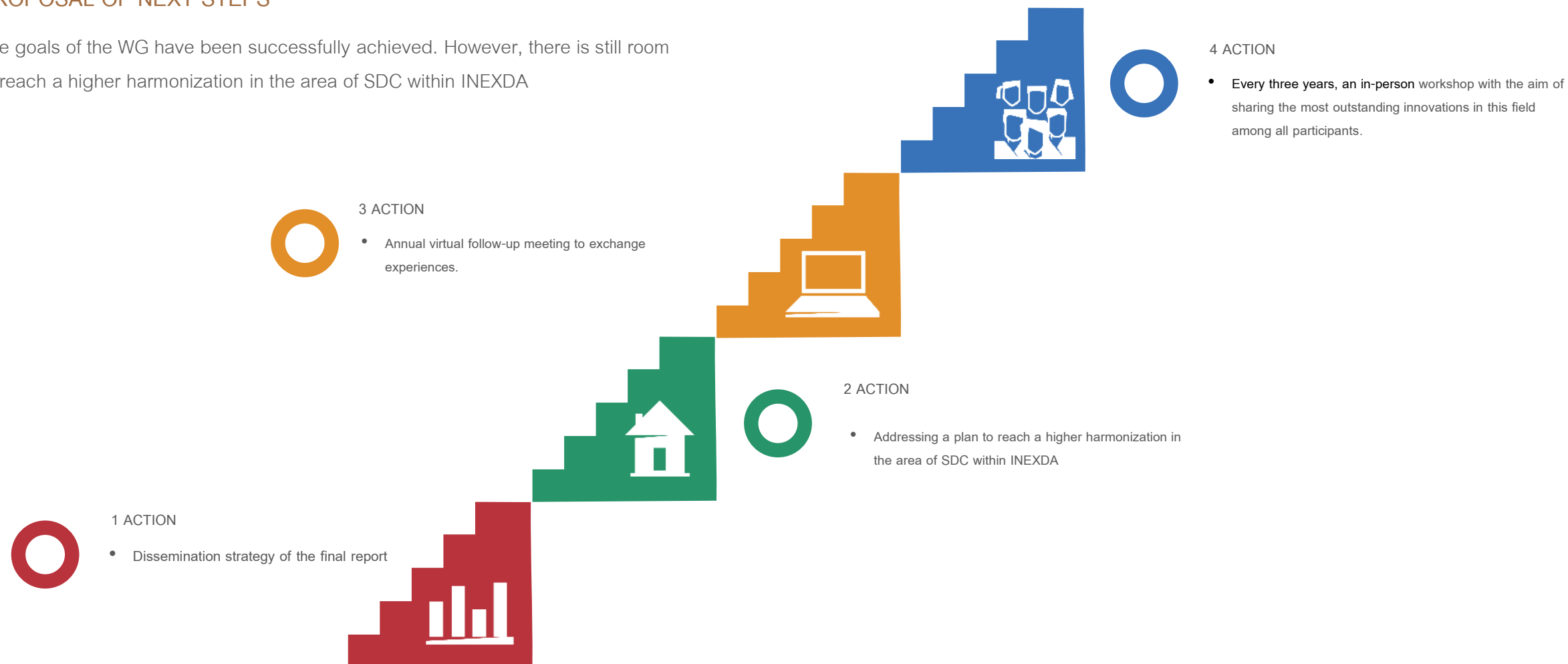
SAFE OUTPUT



- What requirements does a good output have to meet to facilitate the work of the output checker?
- Can output checking be fully automated?
- What rules are required to ensure that the released output is safe?
- ...

PROPOSAL OF NEXT STEPS

The goals of the WG have been successfully achieved. However, there is still room to reach a higher harmonization in the area of SDC within INEXDA



THANK YOU FOR YOUR ATTENTION

