UNECE High-level Group for the Modernisation of Official Statistics

Input Privacy-preserving Techniques

This business case was prepared by ISTAT and Statistics Netherlands and is submitted to the HLG-MOS for their approval.

Type of Activity			
☒	New project	☐	New activity
☐	Extension of existing project	☐	Extension of existing activity
Projects are undertaken by separate project teams. Projects are expected to produce a significant contribution to achieving the HLG-MOS vision		*Activities are undertaken by Modernisation Groups. These activities produce smaller, more detailed outputs to help achieve the HLG-MOS vision*	

Purpose

Statistical organizations are more and more investing on becoming part of a data ecosystem where they acquire and integrate data from multiple sources and provide richer statistical products.
In this scenario, the issue of privacy preservation is particularly relevant: the more sources are acquired and integrated, the higher are the risks of disclosing information violating individual privacy rights. Hence, from a legislative perspective there are indications to take privacy into account throughout the whole data treatment process, through the **'privacy by design'** concept.
National Statistical Organizations (NSOs) are used to apply techniques for enforcing privacy by design on the **output side**, i.e. when publishing aggregated statistical data for dissemination purposes and when sharing microdata for research purposes with statistical disclosure control (SDC) and other output privacy-preserving techniques.
However, NSOs have still to invest on dealing with privacy protection on the **input side**, in a complementary but distinct way with respect to output privacy preservation investments[1].
Different classes of techniques can be used to deal with input privacy[2]. Among them **Secure Multiparty Computation (SMC)** and **Homomorphic Encryption (HE)** play a relevant role.

These methods are particularly suitable for use in a non-trusted environments such as access to private data, interconnectivity of highly sensitive data for the purpose of scientific research, data analytics in Cloud and AI. The goal of this project is to investigate statistical use cases that require protection on the input side, assess and determine applicability of selected classes of techniques for main scenarios, identify opportunities for sharing across statistical community and create community of practice across statistical organizations and external partners (academia, private sector).

Description of the activity

The project is divided into four work packages. The approach is iterative and modular in a way that more mature techniques can be tested with PoCs to speed up their adoption and additional techniques could be added as new work packages and strengthen each other if we do them jointly.

WP1. Documenting statistical use-cases relevant for application of privacy-preserving techniques

1 F. Ricciato, A. Bujnowska, A. Wirthmann, M. Hahn, E. Barredo-Capelot, A reflection on privacy and data confidentiality in Official Statistics, ISI 2019.

2 UN Handbook on Privacy-Preserving Computation Techniques, http://publications.officialstatistics.org/handbooks/privacy-preserving-techniques-handbook/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf

The first step is to investigate and document statistical use-cases where input privacy-preserving techniques can provide (part of) solution. While theoretical descriptions of existing techniques in relation to (mostly academic) use-cases are available the real application of these techniques in typical statistical scenarios is at the moment still grey area.
Part of this WP is also to establish criteria for assessing maturity and readiness of techniques that will be tested in specific modules (WP2 and WP3).

WP2. Secure Multiparty Computation (SMC) methods
SMC is a class of methods based on the principle of secret sharing. The secret shares are produced in a way that does not reveal anything about the input source data to the individual compute parties but allows to compute exactly the correct output that would be obtained by a direct computation on the clear input.

WP2 will investigate and test SMC methods; their maturity and applicability for statistical scenarios. For methods identified as mature the assessment would be supported with practical PoC in partnership with academic or private organization.

WP3. Homomorphic Encryption (HE) methods

Homomorphic Encryption is a class of methods with a special algebraic structure that allows computations to be performed directly on encrypted data without requiring a decryption key.
WP3 will investigate and test HE methods; their maturity and applicability for statistical scenarios. For methods identified as mature the assessment would be supported with PoC or workable prototype in partnership with academic or private organization.

WP4. Identify opportunities for operationalization of methods and sharing of solutions
This work package will identify opportunities for operationalization of mature methods in generic scenarios where practical solutions could be reused and shared across statistical community.

It will also enable exchanging best practices, sharing knowledge and tools among NSIs. Results of the project will be disseminated through the web and organizing seminars in UNECE countries.

Most of the WP work can largely be done be through virtual meetings, but up to two face-to-face meetings would be needed for fast headstart of the project and to create statistical community of practice.

Alternatives considered

Alternative to standalone project could be separate activities for certain techniques undertaken by groups of interested countries or allocated to other Modernisation Groups but in that case the progress would be considerably slower. Use-cases often require combination of privacy-preserving techniques (for example SMC in combination to HE) so it is beneficial to consider applicability of various techniques within the same activity which would be more difficult to achieve without project. Similarly separate activities would be less effective in building international community of practice.

How does it relate to the HLG-MOS vision and other activities under the HLG-MOS?

This project will enable statistical organizations to unlock insights from the data that currently can't be acquired or integrated because of privacy constraints like in scenarios of sharing public-private data or linking privacy sensitive datasets internationally (for example border statistics).
Project will build on results of HLG Data Architecture project and UN Handbook for Privacy-preserving Techniques and contribute to future data-driven activities such as Data-driven decision making support at the local level,

Proposed start and end dates

Start: *January 2020*	**End:** *December 2020*