Shari Stehrenberg, Konstantinos Giannakouris

Trusted Smart Surveys: Solutions for the European Statistical system - An overview of the objectives and the main challenges

Shari Stehrenberg[1]; Konstantinos Giannakouris [2]

[1] Destatis (Statistisches Bundesamt Deutschland)
[2] Eurostat

Abstract:

Smart surveys and trusted smart surveys offer opportunities to extend the content, improve the quality and reduce the response burden by combining traditional survey data collection methods with non-traditional and new forms of data collections.
Smart Surveys are surveys in which respondents are asked to employ smart devices (e.g. smartphones, tablets, activity trackers) to collect data. They combine (inter)active data provided explicitly by the respondent together with passive data collected in the background by sensors.
Trusted smart surveys refer to an augmentation of the smart survey concept by technological solutions aimed at increasing the degree of trustworthiness, hence promoting public acceptance and participation. Constituent elements of a trusted smart survey are the strong protection of personal data based on privacy-preserving computation solutions, full transparency and auditability of processing algorithms.
To employ the potential of (trusted) smart surveys, both methodology and architecture of statistical processes need to be further developed, extended and modified. With the ESSnet Smart Surveys[1] twelve NSIs started in 2020 a project focusing on the opportunities to 'smartify' surveys. The ESSnet takes a first step towards the development of a European platform to share Trusted Smart Survey solutions and components, by defining a conceptual framework including the methodological, technical and legal dimensions of such surveys.
In this paper, we aim at providing a general overview on the objectives and main challenges of Trusted Smart Surveys as solutions for the European Statistical system. Moreover, an outlook on future work related to trusted smart surveys should be given.

Keywords:

Data collection; trusted smart statistics; trusted smart surveys; digital devices; sensor data; privacy enhancing technologies, combining traditional and non-traditional data

[1] The ESSnet Smart Surveys is financed by the European Commission under the Grant agreement number 899365 - 2019-DE-SmartStat

1. Introduction:

Surveys and administrative data have been the only kinds of input data available for official statistics for decades. Therefore, any aspect of the statistical production, including data processing methodologies, data governance models, regulations, organizational practices, all described in the Generic Statistical Business Process Model (GSBPM)[2], have been tailored to such kinds of data. In the last two decades, smart devices, electronic networks and constant generation of data on all aspects of life and the environment have become an integrative component of our societies and economies. [1]

In the recent years several research activities and pilot projects, such as the ESSnet Big Data I and II[3] – launched by the European Statistical System in 2016 and 2018 - have demonstrated the potential of exploiting such new big data sources for official statistics. Some of the results have been published as experimental statistics.[4]

However, the potential gains that might come with these non-traditional, big data sources, involve several important challenges to be addressed today.

Big data are fundamentally different along multiple dimensions, and a systematic critical review of such differences reveals that adopting new types of data for official statistics requires changes in almost every aspect of the statistical system: processing methodologies, computation paradigms, data access models, regulations, organizational aspects, communication and disseminations approaches, and so forth. [1] The term “Trusted Smart Statistics” (TSS) was put forward by Eurostat to signify this evolution [2] and officially adopted by the European Statistical System (ESS) in 2018 in the so-called Bucharest memorandum.[3] Trusted smart surveys are part of the European Trusted Smart Statistics endeavour and form a bridge between traditional and non-traditional data sources. In this paper, we aim at providing a general overview on the objectives and main challenges of Trusted Smart Surveys as solutions for the European Statistical system.

The term smart surveys has been widely used to refer to data collection based on smart personal devices. Typically, smart surveys involve (continuous, low-intensity) interaction with the respondent and with her personal device(s). They combine (inter)active data provided explicitly by the respondent (such as responses to queries, or shared images) together with passive data collected in the background by the sensors (e.g. accelerometer, GPS) of the same device or by other devices within the personal sphere of the respondent. The term *trusted smart surveys* refers to an augmentation of the smart survey concept by technological solutions aimed at increasing the degree of trustworthiness, hence promoting public acceptance and participation. Constituent elements of a trusted smart survey are the strong protection of personal data based on privacy-preserving computation solutions, full transparency and auditability of processing algorithms. [1][4]

Several National Statistical Institutes, within and outside the European Statistical System (ESS), have been modernising their data collection by exploiting the enhanced possibilities offered by (trusted) smart surveys. Eurostat maintains an inventory[5] of tools, sources and information on projects developed within the European Economic Area, exploiting the capabilities of smart surveys in the specific domains of HETUS[6] and HBS. The inventory contains up-to-date information on circa 15 tools for HETUS and 10 tools for HBS.

It is expected that the maturity of these tools will increase over time. Subsequently, we may anticipate a continuation of the current trend of harmonising the tools among the different countries and developing the same or similar functionalities at a conceptual level following the

[2] Generic Statistical Business Process Model: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1
[3] ESSnet Big Data II: https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1_en
[4] Experimental statistics: https://ec.europa.eu/eurostat/web/experimental-statistics/
[5] Inventory of innovative tools and sources for HBS and TUS. Accessed at : https://webgate.ec.europa.eu/fpfis/wikis/display/ISTLCS/INVENTORY
[6] HETUS: Harmonised European Time Use Survey: https://ec.europa.eu/eurostat/web/time-use-surveys; HBS: Household Budget Survey; https://ec.europa.eu/eurostat/web/household-budget-surveys.

recommendations of the Common Statistical Production Architecture (CSPA)[7]. Moreover, this development will fully comply with the General Data Protection Regulation (GDPR)[8] and other relevant privacy legislation. Nevertheless, these important and innovative developments remain fragmented initiatives that take place within the individual Member States.
Therefore, the development of a European platform for trusted smart surveys aims at paving the way to innovative data collections, ensuring that all countries have access to reliable components based on the latest technologies for conducting trusted smart surveys within a privacy-preserving framework.

2. Towards a European platform for Trusted Smart Surveys

Trusted smart surveys build upon the potential of digital devices and sensors that are connected to the internet and generate data that can be embedded in the statistical data collection and production systems. Thus, a part of data capturing, analysis and processing are envisioned as embedded in activities that generate and simultaneously analyse part of the data on the edge. Seen from the ESS perspective, this approach is part of an incremental innovation trend that transforms data collection and eventually the statistical production systems of official statistics.
However, trusted smart surveys bring a disruptive element of change with respect to acceptance by the citizens, as data providers and users of smart devices. Smart surveys may run the risk of not being very successful in terms of adoption and acceptance, due to concerns about privacy by respondents. Mitigating this risk, a carefully designed and cleverly communicated trusted smart surveys system has better chances to be adopted and accepted. This should be considered on the top of data quality that is crucial to gain user engagement and acceptance of the smart devices and sensors paradigm and services.
We envision a European endeavour for developing a flexible European platform for trusted smart surveys, implementing a set of common (horizontal) functions and configurable services that can be used to build particular instances of trusted smart surveys for a wide range of application domains. Such a platform should be modular, evolvable, extensible and agnostic to the specific application domains. Its development involves features that are complex to design and implement, and this adds to the motivation to tackle the challenge at the European level.
Essentially, the idea is to provide ESS members with capabilities to instantiate and run new surveys, fully configurable to meet the specific needs of different surveys and national needs. These capabilities should be enhanced with strong build-in privacy and confidentiality guarantees delivered through privacy-preserving techniques, while avoiding duplication of development costs.
In collaboration with a consortium of twelve NSIs, Eurostat has launched a European project (ESSnet) on Trusted Smart Surveys. This project is organised in three work packages:

1. Communication and coordination (lead beneficiary: Destatis),
2. Development and execution of pilot projects (lead beneficiary: CBS), and
3. Conceptual framework and specifications for a European platform for trusted smart surveys (lead beneficiary: Istat).

The ESSnet aims at taking a first step towards the future development of a European platform for trusted smart surveys. It will cover the pre-development stage: collection of requirements, identification of design principles, development of a conceptual framework and a reference architecture, and formulation of specifications. In doing so, the project reuses, to the greatest extent possible, the tools and experience gained in previous European activities, most prominently related to innovative activities concerning the HETUS and the HBS. The outcome of this ESSnet is intended to serve as a direct input to the future development of such a platform in subsequent projects.

[7] https://statswiki.unece.org/display/CSPA/
[8] https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04, GDPR came into force in May 2018

3. **Overview of objectives of Trusted Smart Surveys**

The overarching objectives, on which the current Trusted Smart Surveys project focuses its efforts, can be summarised as follows:

- Promoting the development of trusted smart surveys for data collection across multiple countries and different statistical domains. Thereby, developments in the use of innovative tools for data collections in one country can benefit when cross-fertilised with architectures, experiences and developments of smart surveys from other countries.
- Developing a general framework, a reference architecture and concrete specifications for a European platform providing the necessary functional and technical environment for conducting trusted smart surveys.
- Developing technological and methodological solutions for processing the input personal data using privacy-preserving technologies (e.g. Secure Multi-Party Computation [5]), with no need to centralise the data at a single entity and in a way that provides solid guarantees in terms of privacy protection, full auditability and complete transparency of the processing methods applied to the data. Embedding survey-specific incentive strategies to foster participation by citizens, through a coherent combination of public communication and individualised incentives, including but not limited to personalised feedback, gamification or public rewards.

4. **Main challenges of Trusted Smart Surveys**

Smart surveys are based on data generated by connected devices through unique addressing schemes that create a pervasive environment (taking into account that a person can interact at any time with the digital world and the physical world). Subsequently, beyond the specific technological challenges related to the collection and processing of highly detailed behavioural data through sensors and smart devices, the main challenge of this European endeavour is developing a solution architecture related to a personal data ecosystem, the governance of the data, privacy preservation and data security.

The concept of trusted smart surveys builds on the concept of involving new relations to citizens as co-producers of data production platforms [6] ideally developing a participatory approach. Data collection is complemented by factual detailed private data produced directly by the respondents' activities, generated by sensors and partially processed on the edge. This framework requires a strong engagement of data respondents in the data collection and processing of data for official statistics. Citizens should be involved in the development of new devices of data production as well as informed of the possibilities of the devices' deployment with respect to collecting their own data, and subsequently rewarded with respect to their demands, interests and contributions. As such, the concept of trusted smart surveys builds once more on the premises of matters of concern, related to data governance and privacy preservation at the front end of any architectural design or software development.

These issues are not new and are part of the wider range of concerns "towards an internet of trusted data" [7] that have been preoccupying our datafied societies. Building trust by placing data privacy and data security at the centre of a European platform for trusted smart surveys is a strong requirement and a challenge.

Initial analysis showed that the challenges that trusted smart surveys have to face are mainly related to the requirements of the GDPR[8]. Pursuant to Article 35 (1) of the GDPR, a data protection impact assessment (DPIA) has to be carried out '*Where a type of processing […] is likely to result in a high risk to the rights and freedoms of natural persons…'* Trusted smart surveys are concerned due to collecting highly detailed behavioural personal data, and from a practical point of view, no extensive experience is yet available on DPIA, as it is a new process under the GDPR for the NSIs.

Currently the ESSnet looks into the issue from a wider perspective. In the most typical case, the NSIs themselves also act as data processors. However, data processors may also be separate private legal entities (e.g. application service providers) which act on behalf of an

NSI. Consequently, a data processing agreement becomes mandatory (GDPR, Art.28). The processors have to provide sufficient guarantees to implement appropriate technical and organisational measures in a way that processing will meet the GDPR requirements and ensure the protection of the rights of the data subject. Therefore, some NSIs require the processors to be certified (for example according to ISO/IEC27001 on how to manage information security).
While all countries have adopted the GDPR, its interpretation and implementation varies within the ESS. In practice, this means that countries may impose supplementary requirements in terms of privacy preservation and data security. In Germany, for example, in the context of smart surveys, an application service provider must also meet the requirements of the Federal Cyber Security Authority of Germany.
More generally, (some) NSIs may perceive collecting and processing personal data via apps too risky from a data privacy or ethical point of view. In July 2020, the European Court of Justice declared the EU-US Privacy Shield[9] as invalid. This judgment is likely to influence the admissibility of app-assisted data collection. Researchers have recently noted that 'Google Play Services component of [these] apps is extremely troubling from a privacy viewpoint.'[8]
The development of a European platform for trusted smart surveys can be seen as an opportunity to address collectively at the ESS level these challenges related to the use of innovative data collections while preserving data privacy at a European level.

5. Work in progress - Outlook

This paper outlines the objectives and challenges in developing a European platform for trusted smart surveys, as part of the European Trusted Smart Statistics. Within this framework, work has already started and concrete actions financed by Eurostat are integrated within the action plans of the NSIs. It is obvious that considerable efforts will be spent in the definition of technical solutions and frameworks for trusted smart surveys in particular related to the architectures that engage data respondents, preserve their privacy and ensure data protection.
In order to tackle the challenges related to data protection, the members of the project of the ESSnet Smart Surveys decided to establish a special working group dedicated to legal-ethical issues. In this way, the ESSnet Smart Surveys also makes an important contribution to further shaping and defining the key notion of "Trust" in Trusted Smart Statistics.
Given its multiple components, trusted smart surveys involve a wide range of stakeholders from the policy, science, statistics, and business domain, and most importantly from the public. In parallel to the development of technical solutions, trusted smart surveys should therefore invest further in communication – most importantly in preparing proposals that seek to engage with the data subjects (respondents) as co-producers of data production systems.

[9] EU-US Privacy Shield is a framework for regulating transatlantic exchanges of personal data between the EU and the USA. This judgement declared that this framework is no longer a valid mechanism to transfer personal data from the EU to the USA.

References

[1] F. Ricciato, A. Wirthmann, M. Hahn (2020). Trusted Smart Statistics: How new data will change official statistics. Cambridge University Press, Data & Policy Volume 2, 2020, e7 https://doi.org/10.1017/dap.2020.7

[2] F. Ricciato, A. Wirthmann, K. Giannakouris, F. Reis, M. Skaliotis (2019) Trusted smart statistics: Motivations and principles. Statistical Journal of the IAOS 35(4), 1–17.

[3] European Statistical System Committee (2018) Bucharest Memorandum on Official Statistics in a Datafied Society (Trusted Smart Statistics). Available at https://ec.europa.eu/eurostat/web/ess/-/dgins2018-bucharest-memorandum-adopted (accessed 27 May 2021).

[4] F. Ricciato, K. Giannakouris, A. Wirthmann, M. Hahn, Trusted Smart Surveys: a possible application of Privacy Enhancing Technologies in Official Statistics, 2020 Societa' Italiana di Statistica (SIS 2020)

[5] D. Archer et al. From keys to databases (2018) – Real-world applications of secure multi-party computation. The Computer Journal, 6(12), December 2018. https://eprint.iacr.org/2018/450.pdf

[6] E. Ruppert, F. Gromme, F. Ustek-Spilda, and B. Cakici. Citizen data and trust in official statistics. Economie et Statistique / Economics and Statistics, 505-506, 2018. Published on 11/04/2019 https://doi.org/10.24187/ecostat.2018.505d.1971

[7] T. Hardjovo, D. L. Shrier, A. Petland (2020). Trusted data. A new framework for identity and data sharing. https://ieeexplore.ieee.org/servlet/opac?bknumber=8900792

[8] D. J. Leith, S. Farrell, Contact Tracing App Privacy: What Data Is Shared By Europe's GAEN Contact Tracing Apps, School of Computer Science & Statistics, Trinity College Dublin, July 2020. https://www.scss.tcd.ie/Doug.Leith/pubs/contact_tracing_app_traffic.pdf