The Conceptual Framework for the European Platform for Trusted Smart Surveys

**Keywords:** Architectural framework, Smart surveys, Trusted Smart Surveys.

# Introduction

The development of Smart Surveys - surveys in which respondents are asked to employ smart devices (e.g. smartphones, tablets, activity trackers) to collect survey data through active and passive data collection - and Trusted Smart Surveys (TSS) - surveys in which respondents are also asked to share existing data collected by trusted third parties, like government authorities and larger, stable enterprises willing to establish data delivery agreements - offers new challenges to improve the quality of social surveys in the NSIs. Different smart survey solutions were developed within the Eurostat Grants for “Innovative Tools and Sources” for two specific surveys: the Harmonized European Time Use Survey (HTUS) and the Household Budget Survey (HBS). 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.

The ESSNet on Smart Surveys, which started its activities at the beginning of 2020, constitutes a contribution towards many important achievements foreseen within ESS: (i) testing and developing (trusted) smart surveys within the ESSNet, (ii) the conceptualization, development and implementation of a new reference architecture for trusted smart statistics as well as the evolvement of new skills within the ESS. The ESSNet will deliver preparatory work to create a European wide platform to share und re-use smart survey solutions and components. The platform will be flexible and implementing a set of common functions and configurable services that can be used to build particular instances of trusted smart surveys for specific application domains and/or target areas. Such a platform should be modular, evolvable, extensible and agnostic to particular application domains. The platform should include support for secure private computing to avoid data concentration (e.g., secure multiparty computation), full transparency and public auditability.

In this context, the work-package 3 (WP3) is working on the idea of TSS platform through two main tasks: (i) conceptualization and development of a general platform for trusted smart surveys, following a top-down design approach; (ii) development of proofs-of-concept in the form of modular prototype elements for essential aspects of the architecture.

In the following, the features of the six tasks composing the work-package are described, in order to provide an overview of the different aspects of the framework.

# The sub-projects of WP3: the areas of the conceptual framework

## Smart Survey Methodology

Smart surveys could be designed for any device that has access to passive data collection features. Usually these devices have one or multiple sensors that collect data for their own functionality or provide access to their sensor data to other applications. Due to the widespread use of mobile devices, the growing number of available sensors in the device itself and via connected devices, they are an obvious choice for smart surveys [1,2]. In addition, wearable devices and IoT devices for home automation became commonly available for consumers, which allows for even wider range of data access. Consequently, new challenges for privacy and informed consent arose [3].

A smart survey methodology does not only cover smart data collection, processing, and analysis, but is also concerned with making the survey smart itself. Thus, survey automation is part of the design process. Design choices, which are explored within this task, are based on informed choices of the integration of big data, reference architectures for big data for official statistics, and preceding methodological research within the ESS for the use of mobile devices and sensors. Methodological discussions of error treatment – as new forms of data also pose new forms of problems – and applied reviews of machine learning techniques are the analytical backbone of this task. As a result, it delivers guidelines, infrastructure and staff profile requirements for the implementation of smart data in smart surveys and ways to deal with hybrid forms of data already in the design process.

## Technical infrastructure

One of the phases in smart surveys development is to define logical components of the technical architecture. In this case, the BREAL reference architecture is used, developed in ESSnet Big Data II grant by work package F. It divides business processes of the architecture into different phases, i.e. acquisition and recording, data wrangling, data representation, modelling and interpretation with shape output as the last phase.

The general assumption of the architecture for smart surveys is to gather the data without or with minimum user interaction. Thus, there is a variety of different sensors that can be used to provide information from mobile devices but not all the sensors can be used in all smart surveys. Then, the data from selected sensors can be linked to other, more specific information, for instance from administrative data. The general framework for acquiring the data is presented in Figure 1, where a key component is a mobile phone which collects the data from sensors. The data is loaded into structured and unstructured databases via RestAPI. Data from mobile phone are also linked to the data obtained from administrative data sources (via API) and text data, e.g. from public databases.

Diagram

Description automatically generated

**Figure 1. The general framework of BREAL data acquisition and recording phase for smart surveys**

## Integration in existing architectural framework

The main objective of this task is to design a trusted smart surveys framework compliant with existing frameworks such as GSBPM, GSIM, CSPA, SPRA and EARF. The target architecture will model *smart* statistical processes resulting from the combination of traditional and new data sources (e.g. sensor data, geolocation). Further, the architecture should fulfil the following general requirements: security, user authentication, dynamic and passive data collection, machine learning, privacy preservation, data transformation and statistical production.

In addition to official statistical standards, several ESSNet research projects and relevant initiatives will affect the design of the TSS architectural framework: Innovative tools and sources for living conditions surveys; ESSNet Mimod; ESSNet Big Data II; Motus platform. The first outcome of the task is a set of business principles resulting from the alignment with existing frameworks. These principles applied to the TSS domain will allow to model the business layer of the target architecture.

## Privacy preserving

To maintain full respondent’s trust and be fully compliant with the EU and national privacy regulation ESS Smart surveys need to apply established principles, develop guidelines and introduce new technologies and techniques in area of privacy and transparency. This task explores best practice, identifies concrete challenges related to Smart surveys, and will proposes guidelines as well as methodological and technical solutions.

Guidelines will be based on established privacy-by-design principles such as seven principles listed below [4]: 1. Proactive not reactive; preventive not remedial; 2. Privacy as the default setting; 3. Privacy embedded into design; 4.Full functionality – positive-sum, not zero-sum; 5. End-to-end security – full lifecycle protection; 6.Visibility and transparency – keep it open; 7. Respect for user privacy – keep it user-centric.

Processing the Smart surveys input personal data will also require new architectural, methodological and technological techniques and approaches, for example: Architecture that doesn’t need to centralise sensitive data at a single entity to reduce concentration risk (for example Edge computing); Advanced privacy-preserving methods for particularly sensitive data such as anonymization and advanced cryptographic techniques (for example Secure Multi-Party Computation); Solutions for full auditability and complete transparency of the processing methods applied to the data.

Combined together, these principles, guidelines and techniques will assure high level of trust in private, decentralized and automated collection as well as processing of (sensor)data, for usage in ESSnet smart surveys and will be particularly important for new types of data such as smartphone sensors (for example GPS location, images taken by phone camera etc.).

## Incentive Schemes

Surveys are dull, offer no particular benefit for the respondent and if there is no penalty involved easy to quit. A powerful incentive to keep respondents engaged in a survey or how to engage people into activities is to gamify them. To gamify a survey means to apply game mechanics to motivate and engage respondents in the activity. Research on gamification of online surveys suggests that they offer resulting benefits in regard to the experience, motivation, participation, as well as amount of data and quality of data [6,7]. However, implementing gamification comes with more methodological and time-consuming tasks while designing a survey.

Trusted Smart Surveys offer the potential to include gamified parts or full gamification. Especially mobile devices as a gateway for smart data and questionnaires are prone to that implementation. Thus, the context for the implementation of gamifications fits the every-day life of respondents. Yet, not only context, also users are important, because not all mechanics appeal in the same way to people, as user preferences and personality does matter [7]. In addition, Trusted Smart Surveys should be engaging and not addictive in nature, which could violate the trust respondents put into it.

## Auditability and Metadata

As part of the definition of the TSS platform, a specific line of work is dedicated to the specification of the metadata that are useful in the context of smart surveys. All the metadata traditionally associated to statistical operations stay relevant, including definitions of concepts, code lists, descriptive metadata, information and process models, etc., but for the TSS environment, special attention will be given to data structures and flows, process execution and traceability, and privacy. Also, the use of sensors represents a relatively new field for statistics, and thus their representation and the specification of related data and processes will receive a specific focus.

# Future work: Proofs-of-concepts

The preliminary framework will be described in a deliverable for the ESSNet. After, some aspects of the architecture will be experimentally developed in Proofs-of-Concept. These will focus on machine learning, metadata, incentive schemes, privacy preservation and IT components of the platform, through the identification and implementation of use cases.

# References

[1] T. D. Buskirk and C. Andres, “Smart Surveys for Smart Phones: Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones,” *Survey Practice*, vol. 5, no. 1, 2012.

[2] F. Keusch, B. Struminskaya, C. Antoun et al., “Willingness to Participate in Passive Mobile Data Collection,” *Public Opinion Quarterly*, vol. 83, pp. 210–235, 2019.

[3] F. Kreuter, G.-C. Haas, F. Keusch et al., “Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent,” *Social Science Computer Review*, vol. 38, no. 5, pp. 533–549, 2020.

[4] A. Cavoukian, "Privacy by Design, The 7 Foundational Principles Implementation and Mapping of Fair Information", 2011, https://iapp.org/media/pdf/resource\_center/pbd\_implement\_7found\_principles.pdf

[5] J. Harms, S. Biegler, C. Wimmer et al., “Gamification of Online Surveys: Design Process, Case Study, and Evaluation,” in *Human-Computer Interaction - INTERACT 2015,* J. Abascal, S. Barbosa, M. Fetter et al., Eds., pp. 219–236, Springer International Publishing, Cham, 2015.

[6] K. W. R. Oliveira and M. M. V. Paula, “Gamification of Online Surveys: A Systematic Mapping,” *IEEE Transactions on Games*, p. 1, 2020.

[7] T. Triantoro, R. Gopal, R. Benbunan-Fich et al., “Personality and games: Enhancing online surveys through gamification,” *Information Technology and Management*, vol. 121, no. 2, 2020.