










Smart Survey Implementation

Grant Agreement Number: 101119594 (2023-NL-SSI)

Work package 5

Design level Legal-Ethical

Deliverable 5.3: Elaboration and application of a modular strategy towards data protection impact assessments (Smart advanced stage)

Version 1.1, 2025-06-25 (final review)

Prepared by:

Natale Renato Fazio, (ISTAT, Italy) - Coordinator
Barry Schouten (CBS, The Netherlands)
Monica Attias (ISTAT, Italy),
Nicoletta Belvedere(ISTAT, Italy)

Work package Leader:

Natale Renato Fazio (ISTAT, Italy)
e-mail address : nafazio@istat.it

Index

Summary

The main objectives of this WP5 deliverable are the definition of the legal context, the elaboration, application, and possible expansion of the proposed modular data protection impact assessment strategy. We elaborate the modular data protection impact assessment strategy. We apply the strategy to the three SSI case studies, assuming they process personal data. We present the outline of the resulting Data protection impact assessment (henceforth DPIA) modules. We explain the consequences of including data processors and provide tips and tricks for doing so. We move to tactics to keep assessments up to date in time. Finally, we discuss what to do when new smart survey applications need to be assessed.

Furthermore, we provide guiding questions to develop DPIA modules, not actual modules per smart feature and smart application. We recommend that NSIs use the strategy and guiding questions in the design and evaluation phases of field tests and share the resulting modules, to allow a new working group to take up the process of preparing the actual modules.

This deliverable reads as follows: in the Introduction, we present the legal context in which the modular data protection impact assessment strategy is rooted; in Section 1, we expose some considerations for drafting DPIAs guidelines for the treatment of personal data in smart surveys; in Section 2, we present the expanded and revised strategy. We then move to a discussion of the three case studies in Section 3, including generalizations. Next, we present the DPIA smart module as it may be included in DPIA's in Section 4. In Section 5, we address the role of data processor and third-parties and give recommendations how to implement smart features in such a setting. We move to the PDCA-cycle in Section 6. We end with a discussion on how to extend the modular strategy to categories of features and surveys that have not yet been considered in Section 7.

Introduction – The legal context

From a legal perspective, the ultimate goal is to create clear and complete guidelines for the creation of a Data Protection Impact Assessment (henceforth DPIA) for ESS-surveys that process personal data using one or more smart features from a specified set of smart features. The emphasis on a set of smart features is made because new features may be developed and/or added gradually in time. It is the task of the SSI project to identify guidelines for features used in the three SSI case studies: receipt processing, geo-tracking and energy meter data donation. Given that in time more features and more applications will be added, the DPIA modular strategy will, by nature, be a dynamic document.

When defining and applying the DPIA guidelines on smart surveys, reference must first be made to the principles and rules of the Regulation (EU) 2016/679 (henceforth GDPR), especially those concerning processing for statistical purposes.

Article 5 (Principles relating to processing of personal data) provides that personal data shall be:

- processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
- accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation')[…];

Article 6 (Lawfulness of processing) provides that processing shall be lawful if the data subject has given consent [...], if it is necessary for the performance of a task carried out in the public interest or in exercise of official authority vested in the controller.

Article 9 (Processing of special categories of personal data) provides, as well, that processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation is authorized if it is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

Article 89 (Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes) provides that :

1. Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.
2. Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15 (Right of access by the data subject), 16 (Right to rectification), 18 (Right to restriction of processing) and 21 (Right to object) subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
3. [...]
4. Where processing referred to in paragraphs 2 and 3 serves at the same time another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs.

Finally, regarding procedural aspects, Article 35 (Data protection impact assessment) provides as follows:

1. Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

2. The controller shall seek the advice of the data protection officer, where designated, when carrying out a data protection impact assessment.

3.-11. [...]

Article 36 (Prior consultation) provides that the controller shall consult the supervisory authority prior to processing where a data protection impact assessment under Article 35 indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk.

When drafting the DPIA relating to the processing of personal data for official statistical purposes, account must be taken not only of the GDPR, but also of the general principles contained in the European Statistics Code of Practice, and of relevant national and European legislation.

1. Considerations for drafting DPIAs guidelines for the treatment of personal data in smart surveys

The Regulation (EU) 2016/679 does not formally define the way to carry out a DPIA, but outlines its minimum content (Article 35 (7)). Furthermore, no templates or guidelines on how to conduct a DPIA for the processing of personal data in the context of statistical surveys have been formally adopted under the ESS. However, for purpose the development of the DPIA for smart surveys guidelines can be taken as a reference the "Guidelines on DPIA" adopted by Article 29 Working Party (https://ec.europa.eu/newsroom/article29/items/611236) and the Guidelines 4/2019 on Article 25 Data Protection by Design and by Default (https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en) adopted by European Data Protection Board (EDPB). In the examination of the DPIAs shared within the SSI project, both of surveys with or without smart features,[1]despite the templates diversity, we can identify common elements referring to the aforementioned acts. This deliverable, D5.3, which originates from deliverables D5.1 e D5.2, is a stand-alone paper and can be read independently from the earlier ones. Nonetheless, we like to mention what is different and what is new. In D5.2, the modular strategy was mostly based on legal considerations without direct empirical support. Here, we have elaborated the strategy based on user tests, pilots and field tests, as far as they have been conducted and evaluated at the time of writing. Also a discussion was organized with the CBS ethical committee about criteria that demand for a full ethical review. Furthermore, the modular strategy has been revised following subtle changes in the categorization of smart features and smart surveys. The most important change is the availability and implementation of alternatives to smart features, especially when the so-called output gap is large. New in this deliverable is the discussion of third party involvement and the role of processors in smart surveys. We apply the proposed modular strategy to all three SSI case studies and attempt to generalize to smart features and smart surveys with similar categorization.

Under the legal-ethical point of view, the logic underpinning the construction of the DPIA guidelines is: a) analysing the process under the methodological point of view, and b) proposing a list of checks on accuracy and output gaps that form a decision tree to decide whether the smart process is really necessary and proportionate. Assessing the process' necessity is fundamental when weighing the restriction of fundamental data subject rights and freedoms.

As the necessity has to be justified on the basis of objective evidence, its assessment has to be carried out before evaluating the proportionality of the limitation of rights and freedoms.

Proportionality is also fundamental to strike a balance between the means used and the intended aim. When considering the right to data protection of the data subjects, proportionality is key for its limitation.

On the basis of the above mentioned considerations, in order to proceed with the impact assessment of the data subjects rights and freedoms in the light of GDPR, the legal-ethical design level has to start from a factual analysis of the aim and advantage of adopting a smart feature in the framework of a current statistic, and the possibility to achieve quality results by using non-smart methods.

[1] DPIAs by NL, IT, NO, DE; CNIL (FR) DPIA template

The data protection module proposes a set of questions to test the level of compliance with the GDPR principles (art. 5) as well as the conformity to the indicators listed in the above mentioned Data Protection by Design and by Default Guidelines.

The proposed modular strategy has three main elements. The first element is the distinction of types/categories of features and feature-application combinations. The motivations are timeliness and harmonization. Recognizing that two different settings are the same by nature will speed up assessments but also harmonize them. The second element is the distinction between application-independent and application-dependent modules in risk assessments. The motivations for separating the two are efficiency and again harmonization. Surveys using the same features are subject to the same risks and have the same set of mitigation measures. It is only through the application that risks may be evaluated against added value. The last element is the cyclical nature of risk assessments. Assessments require empirical support but also need to account for time change.

The legal-ethical design level needs constant input from the methodology and IT design levels and needs to be embedded in the logistics design level. The legal-ethical design level needs to know what are the magnitudes of accuracy and output gaps, the perceptions of the general population on the logic and utility of smart features in a given context, and the viability of non-smart alternatives in a dynamic survey climate. This deliverable 5.3, therefore, is leaning on results reported in D2.3 and D3.3. It uses the categorization/taxonomy of smart features and smart surveys as introduced in deliverables D4.1 to D4.3. The risk assessments underlying to DPIA's are GSBPM building blocks in D4.2 and D4.3.

2. The modular strategy for data protection impact assessments in smart surveys

This section forms the core of the deliverable. We elaborate the proposed modular strategy. As input we have used existing and on-going data protection impact assessments, discussions with the WP5 Working Group Legal, a consultation of the Statistics Netherlands formal ethical committee and findings from WP2 Methodology and WP3 IT.

In Deliverable 5.2, we introduced the overarching modular decision tree for creating a smart survey DPIA. We repeat it here as we will refer to the various steps in subsequent sections and we prefer this deliverable to be a stand-alone paper. However, we also make two changes to the strategy.

We, first, briefly discuss revisions to the classification of smart features and smart surveys. These affect the modular strategy. Next, we present the revised modular strategy and terminology. We then introduce the modules where we distinguish between application-independent and application-dependent modules. We end with describing how assessments may converge based on empirical evaluations as performed within SSI field studies.

We note that other SSI WP's refer to domain-specific design features and non-domain-specific design features. In our context, application-dependent modules are domain-specific and application-independent modules are non-domain-specific.

2.1 Revised smart feature and smart survey classification

Based on pilot results, the classification ('taxonomy') of smart features and smart surveys has been adapted. Here, we summarize the changes and how they affect the modular strategy.

Two changes have been made; one to the classification of a smart feature and one to the classification of the combination of feature and application.

The change to the smart feature classification concerns the type of measurement characteristic. It had three categories in deliverable 5.2: Q&A/no sensor measurement, internal mobile device sensors and external sensor system. The second category is split into to two categories: internal mobile device sensor on-demand and internal mobile device sensor continuous. Receipt scanning is an on-demand internal mobile device sensor. Location tracking is a continuous internal mobile devices sensor. The split is made because continuous measurement has implications on all design levels. For the legal design level, privacy-by-design choices need to be organized differently when the surplus of information is large. Given that the data time resolution is high, such an output gap will often occur for such features. However, when respondents delay data checking and validation, data minimization may be stalled. Another implication for the legal design level is that respondents have to opt out of a feature rather than opt in. Respondents are asked for permission to employ the feature, but it is obviously infeasible to ask permission on a continuous basis. For the modular strategy this change is relatively minor. However, since the classification is expanded it implies we further diversify application-independent modules in our strategy.

The change to the smart survey classification is the introduction of a new characteristics next to the output gap. The criterion is termed Presence of alternatives. It has four categories: no viable alternative, alternative available but burdensome, non-burdensome alternative available but low data quality, and non-burdensome low-error alternative available. The new characteristic comes from the observation in pilots/field studies that respondents ask for the logic of a feature within a specific

context. Location tracking may be logical in travel/passenger mobility surveys, but (much) less so in time use surveys or household budget surveys. The availability of alternatives is an important back-up in settings where there is a surplus of information. Respondents would have the choice to avoid an output gap. Alternatives to smart features are even more important when data handling is (partially) local, i.e. in-device. According to the principles of GDPR, NSIs must be very transparent and clear in explaining to the respondent the different alternatives, advantages and disadvantages. However, the NSI must assess the risk and the organizational and technical measures to be applied, already from the design stage, for the application to guarantee the freedom and rights of the data subject. The NSI must for example suggest the use of complex passwords, to log out each time new information is entered, and, if the data is temporarily stored on the smartphone, to check whether it can be kept encrypted. Doing so, the NSI shows that it is not after the surplus of information per se, but only that it is after respondent burden reduction and/or better data quality.

We make a final side remark to the classification: We like to stress that the output gap can be twofold: shortage of information and surplus of information. A shortage of information means that the respondent needs to supplement smart data. A surplus means that the NSI needs to omit part of the smart data. Both may occur simultaneously. For example, in a time use survey location data do not contain information on with whom an activity was done, which is part of the output need. A shortage of information does not have legal implications, unless it occurs in conjunction with a surplus. The shortage of information will make it harder to perform data minimization. The respondent may have to supplement data dependent on the smart data already collected before part of the smart data can be deleted.

2.2 Modular strategy, definitions and terminology revisited

We present the revised modular strategy. The addition of an extra application-dependent criterion made us expand the former steps 3 and 4. We also add one more step, step 6, namely keeping the assessment up to date in time, following the plan-do-check-act paradigm.

The revised modular decision tree is as follows:

1. Repeat for all smart features to be applied in a survey/application:
    a. Determine the type of smart feature from the taxonomy Φ;
    b. If the feature is not yet part of the taxonomy, initiate a new research and development project (preferably in ESS context);
2. If the feature:
    a. Is a documented feature of type $F \in \Phi_D \subseteq \Phi$, then insert the following modules:
        1. The description of the smart services for type F
        2. List of risks for type F
        3. List of mitigation measures for type F
    b. Is one of the undocumented types, $F \in \Phi_U \subseteq \Phi$, not yet evaluated (but in the taxonomy)
        1. Launch a full risk assessment for F
        2. Perform pilot studies to empirically evaluate the accuracy gap of F
        3. Aim to add the type to the available types Φ_D (preferably within the ESS context)
3. Classify the smart feature(s) – application/survey dimension, i.e. the output gap and the presence of alternatives;
4. For a smart survey of:
    a. Known and documented type $G \in \mathrm{X}_D \subseteq \mathrm{X}$, add the modules:

1. Ethics for G
2. Accepted data minimization considerations for G
3. Special categories of personal data deliberations for G, if any
4. Privacy-Enhancing-Techniques (PET) implemented for G, if any

b. Unknown type $G \in \mathrm{X}_U \subseteq \mathrm{X}$, prepare an analysis DPIA through
1. Consultation of an ethics assessment or ethical committee
2. A perceptions survey or other form of respondent consultation
3. Empirical evaluation of the output gap based on a field test or pilot
4. If applicable, a motivation of absence of alternatives because of respondent burden and/or low data quality
5. Aim to add the documentation as available module to X_D (preferably within the ESS context)

5. Check whether all DPIA modules are implemented as they have been evaluated and accepted.
6. Periodically check whether the added modules are up to date:
   a. Check whether new risks have arisen and/or identified risks have changed
   b. Check whether the accuracy gap has changed form/size
   c. Check whether the output gap has changed form/size
   d. If YES to any of the checks, re-iterate all corresponding steps

We note that the decision tree is generic within the European Statistical System (ESS), because of the explicit focus on the GDPR EU-legislation. Categorization of application-independent and application-dependent components should be done jointly within the ESS. However, individual NSI's may add additional constraints based on national legislation, in particular on (cyber)security. Furthermore, the acceptance of a smart application depends on public perception in a country. The SSI perception survey displayed clearly that perceptions vary between countries.

We advise, for efficiency and comparability reasons, to share DPIA modules and assessments among NSIs. This would favour the availability of ready-made analyses for similar cases, even only in terms of common characteristics examined (accuracy and output gap, type of measurement, data existence, location of processing), which may be equivalent to similar risk mitigation measures. It is clear that NSIs that have similar characteristics of maturity of the production process, in the experience on smart surveys and similar local policies, have more possibilities of sharing information. A proposal could be to create a network of DPOs of the different NSIs for which confidential information on DPIA and risk assessment could be exchanged. In any case, the decision tree can be applied within an NSI with efficiency gains, even if it is limited to not take advantage of the experiences of the community of NSIs.

The decision tree and the elaborations in the following subsections contain a number of terms that we briefly reiterate here as well:

- ***Smart feature***: A smart feature is a data collection action through a smart device such as:
  - In-device storage and/or computing
  - Employment of in device-sensors
  - Linkage to external sensor systems
  - Linkage to public online data
  - Data donation through the respondent
  - Data donation through the statistical institute, i.e. requiring identification keys to link data already in possession
- ***Smart data*****:** Smart data are data collected through one or more smart features;

- ***Smart task***: A smart task is a processing action applied to smart data;
- ***Smart service/solution***: A smart service is a combined and implemented series of smart tasks (i.e. with a single input and a single output);
- ***Device***: hardware unit, electronic device; in particular, high-tech and small devices;
- ***Sensor***: a device that interacts with the quantity to be measured and its environment and detects its variations;
- ***Passive data collection***: involves gathering data without active participant involvement and is well-suited for continuous, objective data;
- ***Active data collection:*** relies on participants actively providing data and is used for subjective information and specific insights.
- ***Application***: we mean one smart survey instance, such as HBS using receipt scanning and uploading or TUS using geo-tracking.
- ***Sources***: These are additional data about respondents or groups of respondents, used as features in methods for cleaning, editing, imputing, predicting, or transforming data. Sources come in two forms:
    - Data already possessed by the institute (e.g., administrative data).
    - Linkage of public online data, which may require preprocessing and editing.

Now, steps 1b, 2b, 4b and 6d imply that action is needed. Step 1b corresponds to an entirely new smart feature. Step 2b means that the statistical properties of the smart data are not yet (fully) known. Step 4b implies that the surplus of information has not yet been evaluated ethically and legally. Step 6d links the PDCA-cycle and enforces periodic re-assessments. In the following two Subsections 2.2 and 2.3, we elaborate steps 2b and 4b, respectively. Step 1b is discussed in Section 7. Step 6 is elaborated in Section 6.

2.3 Application-independent (non-domain-specific) modules

Here, we elaborate step 2b of the modular strategy. Hence, we are in the setting where we can categorize the smart feature but have not yet (fully) documented the tasks, risks, measures and how they are integrated into smart services. We see two dominant characteristics in this module from the legal viewpoint: yes/no local handling of data and the presence and size of the smart data accuracy gap.

Local handling is not specific to smart surveys, but the type of data, i.e. smart data, and the type of processing, i.e. smart tasks, often are different from regular non-smart data collection. As the NSI is the data controller, it is its responsibility to make sure that both unprocessed and processed smart data are secure. This holds even when the respondent may be careless in how he/she secures her mobile devices and/or stores smart data. The key lies in clear explanation in invitation letters and the user interface of applications how smart data should be stored. In case of local handling, the module must provide:

- A description of the potential options a respondent may store and process smart data beyond outside the security net of the application.
- If such options exist,
    - a description of these are communicated in invitation materials and application user interface,
    - a description if and how respondents can abort collecting smart data at any given time during the reporting period.

The other key ingredient is an understanding of the accuracy gap, both in terms of technology and in terms of methodology. When the accuracy of the smart data from one or more smart features added to a survey have not been (fully) analyzed, then action is needed. This amounts to research and empirical study both in the methodology design level and in the IT design level. We describe what is needed to add a module.

The accuracy gap concerns the following components:

I. Type of (quality) metadata needed in case of an accuracy gap, in particular whether if and how longitudinal smart data are needed
II. Process split into in-device, in-house and third party
III. Inclusion of special categories of personal data
IV. List of security risks. Per risk:
    - chance of occurrence
    - consequence
    - list of mitigation measures

The additional mention of longitudinal smart data under component I follows the revised categorization of the type of sensor. When internal sensors produce a continuous stream of data, then data preceding or succeeding a data error may be needed to check and, if needed, adjust.

The sub-steps as defined in Deliverable 5.2 are:

I. Decide whether there is an accuracy gap. If so, specify the type of metadata needed
II. Split the process into smart tasks and per task specify the type of location: in-device, in-house or third party
III. Determine whether special categories of personal data are included in the smart data
IV. Create the list of security risks, and per risk specify:
    - chance of occurrence
    - consequence including severity
    - list of mitigation measures including risk reductions
V. For risks where severity x chance of occurrence exceeds a specified level, demand choosing a mitigation measure with at least a sufficient risk reduction impact

So what makes up the accuracy gap? A loss of accuracy of smart data may come from random (sensor) noise, outliers, sensor drift, systematic (sensor) bias and gaps. Examples are elaborated in Section 3. Inaccuracy has one further dimension and that is a differential across different target population units. If population units have different smart data accuracy then incomparability/inequivalence may result. Causes for such differential can be technological, i.e. different devices, and methodological, i.e. different life styles and device usage.

2.4 Application-dependent (domain-specific) modules

Here, we elaborate step 4b. We are in the setting of a known smart feature but for which the output gap and availability of viable alternatives have not been (fully) documented. The output gap, essentially, amounts to a confrontation of output need and smart data content. When the smart data content have not yet been compared against output needs, then an evaluation of alternative options for respondents, potential privacy-by-design measures and general ethics need to be performed. This is mostly the responsibility of the methodology design level. However, also the IT design level comes in. We discuss the actions needed.

The proposed sub-steps in Deliverable 5.2 are further elaborated:

I. Identify possible alternatives to the smart feature(s) and evaluate their respondent burden and measurement data quality.
II. Determine whether there is an output gap in the smart data. If so, determine whether:
    a. Special categories of personal data are included in the surplus
    b. The average respondent may not (fully) understand the content and implications of the smart data collection
    c. No viable (non-smart) alternatives are available
    d. Not being able to employ the smart feature is stigmatizing
III. If at least one of the criteria II.a – II.d is scored as positive, then:
    a. Re-evaluate the consequences of a security breach
    b. Re-evaluate the strength of the selected mitigation measure(s)
    c. Re-assess the added value of the smart data
    d. Ask for an explicit ethical approval
IV. If there is an output gap but also a viable (non-smart) alternative(s), then implement the alternative(s):

An output gap may have different forms: The time resolution may be higher than needed. It may be hard or impossible to demarcate the exact time window needed. The granularity may be larger than needed, i.e. AI-ML classification routines imply a strong dimension reduction and/or feature importances can be low for some features. It may be hard or impossible to isolate eligible events, i.e. events may be included that are out-of-scope. In most cases, a surplus of information occurs because smart data collection cannot be (fully) controlled or any control would imply a high burden on respondents. As for accuracy gaps, also output gaps may be subject to a differential across target population units. Such differential may arise from different devices producing different smart data content, e.g. iOS against Android. Such differential may also result from behavior that is specific to certain types of respondents. In Section 3, we give examples.

2.5 From idea to production

In Subsections 2.3 and 2.4, we list ingredients to perform steps 2b and 4b. In both, there is a strong relation to the other SSI work packages, and, when extending beyond the SSI, to the various design levels. We discuss the dependencies and relations to all smart survey design levels and SSI work packages.

Let us start by zooming out to the overall business process. Here, the link to WP4 Logistics comes in. WP4 is responsible for the process building blocks and for the overarching view.
The modular decision tree must be embedded in the smart survey GSBPM phases. This is important in particular for the Specify needs, Design, and Evaluate phases. Up to now, all smart applications went through two or more cycles in development and testing before going into production. Through these cycles, the smart survey converges to sufficiently mature design. However, per cycle assessments are needed and legal-ethical choices need to be made explicit. This points at the relevance both at the early GSBPM phases and at the final phases. The implication is that explicit building blocks must be distinguished in which actors, roles, responsibilities, tools are specified. The building blocks need to guarantee a form of empirical support for accuracy gap – output gap trade-offs in privacy-by-design and respondent burden.

The realization of a DPIA is a building block in D4.2 of WP4. It is positioned within the Design phase, to follow the sub-Phase 2 processes. We advocate that the Write DPIA block appears at multiple phases and is detailed following the steps listed in the previous subsections.

The prevalence of an accuracy gap requires empirical support from WP2 on methodological solutions and from WP3 on IT solutions to reduce/adjust. We list the following general questions that may form the basis to such support:

- Methodology - WP2:
    - Will the smart data need adjustment for missing data? If so,
        - How much is deemed necessary post-survey?
        - What is the role of the respondent?
        - Can respondents be expected to understand the tasks needed?
    - Will smart data include measurement errors and outliers that must be adjusted? If so,
        - How much is deemed necessary post-survey?
        - What is the role of the respondent?
        - Can respondents be expected to understand the tasks needed??
    - In case AI-ML routines are included, is performance too weak to exclude manual review? If so,
        - To what extent must manual review be in-house by expert staff?
        - What is the role of respondents in reviewing classifications?
        - Can respondents be expected to understand how to evaluate and adjust classifications?
- Technology/IT - WP3:
    - What additional data protection measures have been implemented because of the collection of smart data?
    - Can technical errors and deficiencies lead to missing data? If so,
        - Do these depend on the device or system used?
        - Can respondents resolve the errors themselves?
    - Can technical errors and deficiencies lead to measurement error or outliers? If so,
        - Do these depend on the device or system used?
        - Can respondents resolve the errors themselves?

An output gap requires empirical support from WP2 on respondent perceptions and alternatives that can be offered, information from WP3 on potential privacy-by-design IT solutions, and evaluations from WP5 itself on ethical decisions. We see the following general questions that may inform choices in handling the output gap:

- Methodology - WP2:
    - Is there a surplus of information in smart data relative to output needs? If so,
        - Can this surplus of data be identified during the respondent reporting period?
        - Is this surplus imperative for accurate adjustments in the analyses/exploration phase of the smart survey?
        - Is this surplus imperative for accurate adjustments in the production phase of the smart survey?
    - Is their empirical evidence that respondents know to what smart data collection they consent to?
    - Is their empirical evidence on ow respondents perceive the collection of the surplus of smart data?
    - Are their sufficiently accurate alternatives for the smart features?

- Technology/IT - WP3:
    - What privacy-by-design measures are possible in case of a surplus of information?
    - What privacy-by-design measures have been implemented in case of a surplus of information?
    - What additional data protection measures have been implemented to protect against disclosure of this surplus of information?
    - Do devices or systems vary in the extent to which they collect a surplus of information? If so,
        - Do privacy-by-design measures vary across devices or systems?
        - Do data protection measures vary across devices or systems?

In Section 3, we illustrate how the answers to these questions are answered and employed for the SSI case studies.

3. Application of the modular strategy

The approach we take to gradually elaborate the modular strategy is to fix thoughts through the SSI case studies. Here, we apply the strategy from the assumption that we can categorize the smart features but have not yet documented application-(in)dependent modules. We, thus, perform steps 2b and 4b. In particular, we interviewed WP's 2 and 3 coordinators about the (anticipated) accuracy and output gaps. For some components of the gaps we can only give anticipated results. Pilots and field studies do not cover all questions. This holds especially for the energy data donation case study.

For each case study, we first run through the overarching strategy given in Subsection 2.2 and then move to the details as discussed in Subsections 2.3 and 2.4.

3.1 Receipt scanning and uploading in the Household Budget Survey

We start with the overarching decision tree:

1. Repeat for all smart features to be applied in a survey/application:
    a. Determine the type of smart feature from the taxonomy Φ:
    Receipt scanning = Data is non-existent, Internal mobile device sensor, Accuracy gap requires respondent assistance, Handling is partly in-device.
    Receipt uploading = Data is existent, Internal mobile device sensor, Accuracy gap is negligible, Handling is in-house.
    b. If the feature is not yet part of the taxonomy, initiate a new research and development project (possibly in ESS context):
    It concerns features that can be categorized.
2. If the feature:
    b. Is one of the undocumented types, $F \in \Phi_U \subseteq \Phi$, not yet evaluated (but in the taxonomy):
    Receipt scanning has been investigated for a number of years but we include it anyway.
    Receipt uploading is new.
        1. Launch a full risk assessment for F
        Risk assessments have been performed by CBS and SSB within SSI, and by Stat Finland outside SSI. See below for details.
        2. Perform pilot studies to empirically evaluate the accuracy gap of F
        The accuracy gap has been evaluated within SSI WP2.2 and WP3 and earlier within ESTAT-project ESSnet Smart Surveys. See below for details.
3. Classify the smart feature(s) – application/survey dimension, i.e. the output gap and presence of alternatives:
    Receipt scanning = Output gap can be handled through questions, Manual data entry alternative burdensome and risk of low data quality
    Receipt uploading = Output gap can be handled through questions, Manual data entry alternative burdensome and risk of low data quality.
4. For a smart survey of:
    b. Unknown type $G \in \mathrm{X}_U \subseteq \mathrm{X}$, prepare an analysis DPIA through
        1. Consultation of an ethics assessment or ethical committee
        No formal ethics assessment or request has been performed because of the negligible output gap.
        2. A perceptions survey or other form of respondent consultation

Within the SSI smart perceptions survey, respondents in IT, NL and SI were asked for their perceptions on receipt scanning and uploading. User tests have been performed for receipt scanning in WP2.3 and in ESTAT-projects @HBS and @HBS2. See below for details.

3. Empirical evaluation of the output gap based on a field test or pilot
   The output gap has been evaluated within ESTAT-project ESSnet Smart Surveys and within the field test of SSB. See below for details.
4. If applicable, a motivation of absence of alternatives because of respondent burden and/or low data quality
   Implemented alternatives are manual data entry. It is known to be very burdensome (Household Budget Surveys have relatively low response rates and high attrition). It is also known to be prone to recall error.

5. Check whether all DPIA modules are implemented as they have been evaluated and accepted:
   CBS and SSB have implemented modules in DPIA's.
6. Periodically check whether the added modules are up to date
   To be performed in 2026

To prepare the modules, we have consulted deliverables prepared in earlier ESTAT-projects and we have asked questions to WP2 and WP3 coordinators. The questions considered are:

Methodology:

- Accuracy gap:
  - Is adjustment for missing product descriptions and/or prices needed? Are respondents asked to do this? And are they able to?
    Yes, text extraction has missing items and respondents need to check and supplement.
  - Is adjustment for errors in product descriptions and/or prices needed? Are respondents asked to do this? And are they able to?
    Yes, text extraction has both spurious items and errors in product texts and product prices. Respondents are needed to check and correct.
  - Is adjustment for unknown product classification needed? Are respondents asked to do this? And are they able to?
    Yes, product classification accuracy is around 90%. Respondents are not asked and in-house manual review is included. Respondents are considered insufficiently knowledgeable to classify products. They may, however, give additional information.
  - Is adjustment for errors in product classification needed? Are respondents asked to do this? And are they able to?
    See above.
  - Will respondents need to be involved in updating/retraining models?
    At this stage, the updating/re-training strategy is in development. It is unclear whether respondents will be asked to assist. As it appears now, this will not be the case.
- Output gap:
  - Is there smart data extracted from receipts that are not needed?
    Potentially yes, but it is not kept and submitted.
  - Are there special categories of personal data?
    Yes, expenditures cover also health care, memberships of clubs/parties, etc. Most of the special categories of personal data are collected through separate questionnaires. Through the questionnaires they contain no surplus of information. However, scanned or uploaded receipts may contain more details.

- Can it be assumed that respondents know what are the implications of submitting receipts?
  Yes, user tests indicate that respondents understand what is being extracted. They see the scan that is made and the extracted texts.
- Do they have an alternative?
  Yes, they can enter data manually. They can also remove receipts during the reporting period.
- Is not being able to scan/upload stigmatizing?
  Only to a minor extent, because the diary and questionnaires can be completed on any device. However, not being able to use scanning implies older mobile devices that are (no longer) supported. Having an outdated mobile device only may be considered stigmatizing.

IT and technology:

- Accuracy gap:
  - Does performance of receipt text extraction vary across devices/platform? Can respondents do anything to improve?
    Yes, performance varies. Respondents are notified of lower performance for smartphones of five years and older. They can check and edit receipt scans. They can enter data manually.
  - Have there been any extra data protection measures for receipts?
    There are three risks: A breach of the app, an interception of data transfer and a breach of the backend. Relative to existing (platform) solutions the only new feature is the submission of receipt scan files and e-receipts. These files are treated as images and not as executables. Any malware in scans or e-receipts cannot have any consequence, therefore. No additional measures have been implemented. There is a difference between applications in how image files are stored. For the MOTUS application, images are stored in the backend and not locally in-device. The implication is more data transfer back and forth when respondents inspect or edit images. A study can be configured such that after completion of the study by the respondent are deleted or kept for model training. The CBS HBS application does keep local copies and images are transferred once. As for MOTUS, image files can be deleted or kept. The default for both applications is that file are deleted.
- Output gap:
  - What privacy-by-design choices have been made, if any?
    The most important choice is that non-relevant text extracted from receipts is not kept. Only relevant information also asked in manual data entry is kept. The applications can, however, be configured such that for model training non-relevant text can be kept. The default is that non-relevant text are deleted.
  - Have there been additional data protection measures for the surplus in receipt extracted data?
    No

Summary: The privacy-by-design measure to discard irrelevant text information and the option to enter data manually make the smart features relatively easy to handle. The submission of (scanned) text files does not give additional security risks. Data protection measures are comparable to non-smart counterparts.

Generalization to overarching smart survey categorization: The evaluation can be extended to scanning and uploading of text images as long as 1) irrelevant texts are discarded, 2) data can be entered manually as an alternative and 3) it is logical in terms of burden and data quality.

3.2 Location tracking in time use and passenger mobility surveys

The overarching decision tree:

1. Repeat for all smart features to be applied in a survey/application:
    a. Determine the type of smart feature from the taxonomy Φ:
    Location tracking = Data is non-existent, Internal mobile device sensor, Accuracy gap requires respondent assistance, Handling is in-device[2]
    b. If the feature is not yet part of the taxonomy, initiate a new research and development project (possibly in ESS context)
    It concerns a feature that can be categorized.
2. If the feature:
    b. Is one of the undocumented types, $F \in \Phi_U \subseteq \Phi$, not yet evaluated:
    Location tracking has been investigated for a number of years but we include it anyway.
        1. Launch a full risk assessment for F
        Risk assessments have been performed only by CBS but outside SSI. See below for details.
        2. Perform pilot studies to empirically evaluate the accuracy gap of F
        The accuracy gap has been evaluated within SSI WP2.2 and WP3. See below for details.
3. Classify the smart feature(s) – application/survey dimension, i.e. the output gap
    Location tracking has a large output gap that requires respondent assistance. Manual data entry is a viable alternative but burdensome and prone to specific types of measurement error (recall errors, underreporting errors).
4. For a smart survey of:
    b. Unknown type $G \in \mathrm{X}_U \subseteq \mathrm{X}$, prepare an analysis DPIA through
        1. Consultation of an ethics assessment or ethical committee
        A formal ethics assessment has been performed at CBS outside SSI. ISTAT has prepared an assessment within SSI but without location tracking. No formal assessment was made of location tracking within SSI.
        2. A perceptions survey or other form of respondent consultation
        Within the SSI smart perceptions survey, respondents in IT, NL and SI were asked for their perceptions on location tracking. User tests have been performed in WP2.3. See below for details. Willingness to be tracked was relatively low and varied strongly across countries.
        3. Empirical evaluation of the output gap based on a field test or pilot
        The output gap has been evaluated within two field tests of CBS outside SSI. See below for details.
        4. If applicable, a motivation of absence of alternatives because of respondent burden and/or low data quality

[2] The handling of location data is classified as in-device for two reason. The first is that part of the processing is local, namely the pre-processing of outliers and noise, in order to present the tentative data to respondents. The second is that as long as location data remain tentative, i.e. not accepted, rejected or altered by respondents, they are completely stored in-device.

The alternative is the manual entry of begin and end times of all travels including transport modes, and the elaboration of all activities for stops. Time use diaries are known to be subject to recall error and underreporting error, in particular for shorter trips and shorter stops.

5. Check whether all DPIA modules are implemented as they have been evaluated and accepted:
   CBS has implemented modules in a DPIA.
6. Periodically check whether the added modules are up to date
   Unclear when a re-evaluation will be performed given that implementation was postponed.

We consulted WP2 and WP3 deliverables and contacted coordinators of WP's 2 and 3. The location tracking smart feature has not as far advanced as planned, however, and some choices are open at the end of SSI. The following questions were evaluated:

Methodology:

- Accuracy gap:
  - Is adjustment for missing location data needed? Are respondents asked to do this? And are they able to?
    Yes, location data are subject to a fair amount of missing data arising from a range of causes. The amount and frequency of missing data depend on the type of device and on the platform. Missing location data may lead to missed travels or missed stops.
    In time use surveys, the requested classification of type of activity is very detailed. Often, multiple activities take place at the same location, so that supplementing is imperative anyway. Furthermore, the details of a stop (with whom, any side activities) will not be derived from location data regardless of being complete or not. Location data function as a tentative time frame for further supplementation. Respondents need to complete the diary and also fill in any gaps.
    It is not yet known to what extent respondents accurately fill in gaps. Given that the exact location itself is not needed, the conjecture is that respondents are able to complete diaries if they are sufficiently motivated. Field studies at CBS in the context of travel surveys show, however, that a non-negligible proportion of respondents leaves gaps in the diaries. In the context of time use, this may imply that respondents underreport certain types of activities.
  - Is adjustment for errors in location data needed? Are respondents asked to do this? And are they able to?
    Yes, errors occur frequently. Location data are subject to modest sensor noise at any given location. However, larger errors occur as well, especially in more urban areas. As a consequence, travels may be missed or vice versa spurious travels may occur.
    Like gaps in location data, respondents are asked to check and correct. Again it is unknown to what respondent can and will correctly detect artefacts.
  - Is adjustment for errors in transport mode classification needed? Are respondents asked to do this? And are they able to?
    Likely, yes. At the time of writing transport mode prediction is still under development. Early results in master thesis projects by CBS show that accuracy will likely converge to values around 70% for uni-modal travels. The correct prediction of multi-modal travels depends strongly on the performance of the stop-track segmentation. It must be expected that a quarter or more of all travels will need to be corrected or provided by respondents.

Since mode of transport is a part of the output need, respondents will be asked to assist. For multi-modal travels it must be assumed that they cannot or will not provide all details when predictions are false. However, empirical evidence is lacking.

- Is adjustment for unknown transport modes needed? Are respondents asked to do this? And are they able to?
  Likely, yes. ML models will be pre-trained on the most prevalent transport modes plus an 'other' category. This 'other' category must be further specified according to the time use guidelines.
  Since mode of transport is part of the output need, respondents will be asked to detail the prediction of an 'other' category. Empirical evidence is lacking in how willing they are and how well they can do this.
  User tests point at potential unease among respondents in including transport mode prediction. They may explain the service as a form of monitoring, even though predictions are fully automated.
- Is adjustment for errors in activity classification needed? Are respondents asked to do this? And are they able to?
  Likely, yes. At the time of writing activity type prediction is still under development. Early results in a master thesis project by CBS show that accuracy will likely converge to values around 80% for single purpose stops. The prediction performance of all activities in multi-purpose stops will likely be low. It must be expected that multi-purpose stops will need to be elaborated by respondents.
  Since type of activity is the core output need, respondents will be asked to assist. For multi-purpose stops it must be assumed that they cannot or will not provide all details. However, empirical evidence is lacking to date.
  See above for checking predictions.
- Is adjustment for unknown activity needed? Are respondents asked to do this? And are they able to?
  Likely, yes. Like transport mode prediction, there will be 'other' and/or multi-purpose categories. However, for time use, the range of activities contained in those categories will be broad. Respondents will have to supplement a fair amount of information. It must be assumed that performance is similar to non-smart time use surveys.
  User tests point at potential unease among respondents in including stop purpose prediction. They may explain the service as a form of monitoring, even though predictions are fully automated.
- Will respondents need to be involved in updating/retraining models?
  Yes, but assistance in updating/retraining will likely go through the changes and supplements that respondents have to make anyway.
  Again see the earlier remarks about unease.

- Output gap:
  - Is there smart data extracted from location tracking that are not needed?
    Yes, location data contain detail that is not needed for time use surveys. It will give exact physical locations. These locations are not linked to detailed contextual data such as address, street name, name of a shop, name of an employer, name of a school, etc. The smart data by themselves are, thus, meaningless without further contextualization. However, any malevolent user or intruder could easily derive such context through open source databases. In other words, there is a potential surplus.
  - Are there special categories of personal data?

Potentially, yes. Location data are not special categories of personal data, but they may indirectly point at special categories of personal data. Physical locations may correspond to hospitals, churches, political venues, etc that may be viewed as special categories of personal data.

- Can it be assumed that respondents know what are the implications of being tracked?
  Yes, before location tracking starts, the vast majority of respondents will know what it comprises of. Location tracking has strong similarities to commonly used apps that do the same. Location tracking data trajectories, including anticipated stops, are shown to respondents during the reporting period. Hence, if they would not know beforehand, they will see what data are collected. The SSI smart perception survey demonstrated that respondents are aware of the implications
  User tests seem to indicate that respondents are surprised by any form of enrichment of their travels to inform transport modes and stop purposes. Despite documentation and help options inform them, they may not notice this.
- Do they have an alternative?
  Yes, respondents do not have to enable location tracking in a time use survey. It is merely a service and not mandatory. They can enable and disable location tracking at any time. When a respondent decides not to use location tracking, then travels and stops need to be entered manually.
- Can not being able to deploy tracking be considered stigmatizing?
  Likely, no. Older mobile devices are not supported and location tracking is not possible without a mobile device. Respondents not able to use the feature either do not have a mobile device or only one that is relatively old. To some extent such respondents may feel that are treated differently. Given that a fair amount of requested data have to be entered manually anyway and given that the traditional approach is non-smart, the potentially stigmatizing nature is deemed acceptable.

IT and technology:

- Accuracy gap:
  - Does performance of location tracking vary across devices/platform? Can respondents do anything to improve?
    Yes, there is a strong dependence. Location tracking is handled differently for iOS and for Android. iOS is much more restrictive than Android in time resolution and battery management. Within both platforms there also is substantial variability in the frequency and accuracy of location data. Respondents are pointed at options to improve performance through the app UI. They can 'white-list' the app, i.e. exempt the app from battery saving interventions. They should also turn off overall battery saving mode. Beyond these options, there still is variability in performance that respondents have to accept.
  - Does performance of transport mode prediction vary across devices/platform? Can respondents do anything to improve?
    No. Transport mode predictions will only vary when the underlying location data differ. The latter will often be true, i.e. two different devices will give different data when making exactly the same travels. However, when data are the same, predictions will also be independent of the device.
  - Does performance of activity prediction vary across devices/platform? Can respondents do anything to improve?

No. Activity purpose predictions will only vary when the underlying location data differ. The latter will often be true, i.e. two different devices will give different data when making exactly the same travels. However, when data are the same, predictions will also be independent of the device.

- Have there been any extra data protection measures for location tracking?
  The handling of location tracking data is application-specific. The MOTUS application sends location data to the backend and no data are kept in-device. Location data are segmented to tentative tracks and stops. As long as tentative data have not been confirmed, revised or rejected by respondents, all data are kept. Each time a respondent decides to view tentative data for a certain day or part of a day, the corresponding data are displayed in the app. Hence, there is transfer of anticipated stop-track segments. After respondent evaluation, data are deleted by default. A study can be configured such that data are kept for model training. The CBS travel application does store location data locally and segments location data into stops and tracks locally. Once a respondents confirms or edits, segments are submitted to the backend.

- Output gap:
  - What privacy-by-design choices have been made, if any?
    All location data are segmented into tracks and stops, but kept until a respondent confirms or edits the segments. Once stops and tracks have been confirmed location tracking data are deleted. Points-of-interest data can be merged in the backend to assist segmentation and to predict travel modes and/or travel purposes. Results can be shown to the respondents. Again, once a decision is made by the respondent, the additional data are deleted by default.
  - Have there been additional data protection measures for the surplus in location tracking data?
    By default the surplus is deleted once a respondent confirms or edits. The study can be configured such that data are kept for model training purposes.

Summary: Location tracking is subject to both a large accuracy gap and a large output gap. This makes location tracking a complex smart feature. The trade-off in accuracy gap – output gap is very influential. The solution may be clear communication, clear presentation of what is collected, and explicit request to change location data from tentative to accepted, and a clear manual (but more burdensome) alternative. User tests seem to point at the need to inform respondents in-the-moment rather than prior through recruitment materials.

Generalization to overarching smart survey categorization: For generalization purposes, it is meaningful to compare time use to passenger mobility/travel. In time use surveys the added value for location tracking is smaller than it is in passenger mobility/travel surveys. In the latter, exact locations are a core output need as they are input to traffic infrastructure models. As a result, also the output gap is much smaller, and, hence, the trade-off to the accuracy gap easier to motivate and explain. The evaluation of location tracking in time use extrapolates to all non-existent sensor data with a large accuracy gap and a large output gap. Findings indicate that a continuously monitoring mobile device sensor requires clear logic, clear communication and respondent control.

3.3 Donation of energy data in Energy and Housing surveys

Donation of energy meter data is a new smart feature that was added deliberately to evaluate to what extent experiences can be extrapolated. The findings are based on a small pilot in one country only and must be interpreted with some care. The primary interest in energy data donation comes from

housing and living conditions surveys. However, energy usage may also point at specific types of (side) activities that can be included in time use surveys.

Let us first look at the overall decision tree:

1. Repeat for all smart features to be applied in a survey/application:
    a. Determine the type of smart feature from the taxonomy Φ:
    Energy meter data donation = Data is existent, External sensor system, Accuracy gap is unknown, Handling is in-device
    b. If the feature is not yet part of the taxonomy, initiate a new research and development project (possibly in ESS context)
    It concerns a feature that can be categorized except for the accuracy gap dimension.
2. If the feature:
    b. Is one of the undocumented types, $F \in \Phi_U \subseteq \Phi$, not yet evaluated:
    Energy meter readings have been implemented in dedicated applications, but not in the context of official statistics about households. It is, thus, mostly undocumented.
        1. Launch a full risk assessment for F
        CBS performed a small-scale qualitative study in which 12 colleagues used a commercial app linked to a dongle. The logistics were ad hoc. Given that in the future another solution may be used, no risk assessment was conducted.
        2. Perform pilot studies to empirically evaluate the accuracy gap of F
        Three data streams could be downloaded from the cloud maintained by the commercial party: electricity, gas (if applicable) and solar panel (if applicable). Afterwards participants were invited for an evaluation in which four relevant questions were asked: Did they experience missing data?, Did they experience implausible data?, Did they see unexpected large peaks?, and Did the app and/or dongle show malfunction. All participants replied 'no' to all questions. For now, there is not yet reason to believe, therefore, that data shows a large accuracy gap. However, it must be noted that participants are no experts in energy usage of devices. An expert look based on a larger sample will still be needed. Furthermore, it is unclear to what extent the provider performs any form of preprocessing of the collected data.
3. Classify the smart feature(s) – application/survey dimension, i.e. the output gap and presence of alternatives.
    Energy meter data have an output gap. No viable alternatives exist that have acceptable burden and data quality.
4. For a smart survey of:
    c. Unknown type $G \in \mathrm{X}_U \subseteq \mathrm{X}$, prepare an analysis DPIA through
        1. Consultation of an ethics assessment or ethical committee
        At this point an ethics assessment has not yet taken place.
        2. A perceptions survey or other form of respondent consultation
        Within the SSI smart perceptions survey, respondents in IT, NL and SI were asked for their perceptions on energy meter data donation. Willingness to donate, if possible, was relatively high but with some variation across countries.
        3. Empirical evaluation of the output gap based on a field test or pilot
        In the CBS pilot data streams were displayed in the app at a very high frequency. However, data could only be downloaded every 15 min. This resolution is too low for the prediction of devices. It is possible to donate data at a much higher frequency, but in the pilot it was, thus, impossible to evaluate the output gap with

respect to machine learning predictions. If submitted at high frequency there is an output gap. It must be noted, however, that energy statistics classify devices at a relatively high level. For example, all kitchen devices are clustered to one category.
The CBS pilot showed that the net electricity usage of solar panel and electronic devices is measured. Consequently, the electricity usage is zero when the solar panel production is larger than the electricity needed. Hence, the presence of a solar panel masks the usage of electricity. So here, there is actually also a data shortage.

4. If applicable, a motivation of absence of alternatives because of respondent burden and/or low data quality
   The only alternative is detailed diary keeping of the usage of electronic devices and high frequency reporting of energy meter status. In practice, the burden is too high and data quality too low. So there is no viable alternative then to adjust the detail in output need.

5. Check whether all DPIA modules are implemented as they have been evaluated and accepted:
   There is not yet a DPIA.
6. Periodically check whether the added modules are up to date
   Not applicable yet

Methodology:

- Accuracy gap:
  - Is adjustment for missing energy data needed? Are respondents asked to do this? And are they able to?
    Preliminary conclusion: No. But unclear what processing is done by a commercial provider.
  - Is adjustment for errors in donated energy data needed? Are respondents asked to do this? And are they able to?
    Preliminary conclusion: No. But unclear what processing is done by a commercial provider.
  - Is adjustment for unknown electronic device needed? Are respondents asked to do this? And are they able to?
    Cannot be answered from the CBS pilot.
  - Is adjustment for errors in electronic device classification needed? Are respondents asked to do this? And are they able to?
    Cannot be answered from the CBS pilot.
  - Will respondents need to be involved in updating/retraining models?
    Cannot be answered from the CBS pilot.
- Output gap:
  - Is there smart data extracted from donated energy data that are not needed?
    Time stamps during the day are provided but they are not relevant for energy statistics.
  - Are there special categories of personal data?
    No
  - Can it be assumed that respondents know what are the implications of donating energy data?
    In the app respondents can see all data that are collected. Data are only donated at the end of the reporting period. They can decide anytime to not donate. However, it must be assumed that prior to the data collection they are not fully aware of what is contained in

the data. The reason is that the average respondent will not be familiar with this type of data.

- Do they have an alternative?
  No. Detailed diary keeping leads to very low data quality. Even in the CBS pilot, the diaries showed big gaps.
- Can not being able to donate data be considered stigmatizing?
  The CBS pilot is too small to draw a conclusion.

IT and technology:

- Accuracy gap:
  - Does quality of donated energy data vary across smart energy meters? Can respondents do anything to improve?
    Unknown. Only one solution was used in the CBS pilot.
  - Have there been any extra data protection measures for donated data?
    The pilot was low-profile with ad hoc logistics. The commercial provider makes statements about encryption of data transfer and security of the backend cloud. However, a solution as is used in the pilot will likely not be employed when moving to implementation. No strong conclusions can be drawn.
- Output gap:
  - What privacy-by-design choices have been made, if any?
    None have been made in this small pilot. In future studies, attempts will be made to aggregate during the reporting period and only submit summaries. This data minimization seems feasible given the relatively small accuracy gap.
  - Have there been additional data protection measures for the surplus in energy data?
    No

Summary: The main preliminary conclusion is that energy data have a modest accuracy gap but do have an output gap. It is unclear whether real-time predictions of the type of devices can be made. It is, thus, also unknown whether energy data can be aggregated and only summaries can be submitted. The conjecture is that respondents will need to assist and that post-survey evaluation across the full energy time series will be needed to recognize patterns.

Generalization to overarching smart survey categorization: The energy data donation case study shows strong resemblance to data donation of physical activity trackers. It also shows some resemblance to data extracted from wearable research-grade physical activity trackers and indoor climate systems that are provided to respondents. Studies with respondent-owned trackers and with trackers and climate systems provided to respondents have been performed by CBS over the past five years and within ESSnet Smart Surveys. See Luiten et al (2022a and b), Kompier et al (2024) and De Wolf et al (2025). What they share is the dependence on third parties, the logistics of distributing devices and uploading data, the small accuracy but large output gap, the continuous stream of data, the non-centrality of what is being measured for average respondents, and the need to let respondents assist. For the physical activity studies, CBS did perform risk assessments and created a DPIA. It is anticipated that much can be borrowed for the energy data donation.

4. DPIA module

In the previous sections we discussed the three case studies at the hand of a list of questions. We also made a first extrapolation step to the larger taxonomy categories to which a feature plus application belong. Here, we go into more detail what modules look like.

4.1 Outline of DPIA 'smart' modules

As we argued in Section 2, for each smart feature category and for each smart survey application category, a module must be added to a DPIA. The outline of the modules is described here.

The considerations and responses to the preceding questions must be adequately documented in accordance with the principle of accountability established by Regulation (EU) 2016/679 (Article 5, paragraph 2, and Article 24). To this end, the preferred approach appears to be the DPIA (Data Protection Impact Assessment) of the investigation, within which each smart feature used should be the subject of a specific module. This module should describe the nature, scope, and context of the personal data processing associated with its use and demonstrate compliance with the principles set out in Article 5 of Regulation (EU) 2016/679.

The preparation of the module must take into account the indications that emerged from the Consultation of an ethics assessment or ethical committee, from the perceptions survey or other form of respondent consultation and from the field tests or pilot, if they have been performed, and to justify the measures that deviate from them. The module can be submitted to the DPO for an opinion, as of Art. 35, par. 2 of Regulation (EU) 2016/679.

Each module contains the following topics: Lawfulness and purpose limitation, Fairness, Minimization, Transparency, Accuracy, Storage limitation, and Integrity and confidentiality. Per topic, a list of subtopics/questions is presented in Section 4.2.

We are not providing actual modules. The main reason is that the case studies have not yet converged at the time of writing. Field tests by CBS and ISTAT take place in the near future, but outside the scope of SSI. We strongly advise that, as a spin-off of project SSI, a small working group is formed that takes up the task of integrating all field test results and completes modules for location tracking and for scanning and uploading of texts.

4.2 Content of DPIA 'smart' modules

For the completion of this module, taking as a reference the Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0 - EDPB, it is possible to start from the following questions.

Description of smart feature:

Feature (survey)	Data existent prior	Type of measurement	Accuracy gap	In-device handling	Output gap
Es. Location tracking (TUS, Passenger Mobility)	*YES/NO*	*Internal sensor ..*	*Respondent ..*	*YES/NO*	*Q&A smart ..*

PRINCIPLES OF ARTICLE 5 OF REGULATION (EU) 2016/679

LAWFULNESS AND PURPOSE LIMITATION

The Data Controller must identify a valid legal basis for the processing of personal data; the measures and guarantees should contribute to the obligation to ensure that the entire lifecycle of the processing is in line with the relevant legal basis. Moreover, personal data must be collected and processed for specific, explicit, and legitimate purposes and the processing must be necessary for the pursuit of such purposes.

Both principles must be considered with reference only to the collection or processing of personal data exceeding what is strictly necessary for the achievement of the statistical purpose, which would not occur if the smart feature was not used (surplus of information), as the processing of personal data for the survey is examined in another section of the DPIA.

1. **Is there a surplus of information in smart data relative to output needs?**
   1.1 No. [In this case, the legal basis for the processing is the one which has been already identified for the survey].
   1.2 Yes. [In this case, proceed to the following questions].
2. **What categories of personal data are involved?**
   ……. [Indicate the categories of data (e.g., type of device, IP address, time and/or place of data provision, etc.), specifying if there are special categories of data or information that allow tracing special categories of data (e.g., health data contained in receipts for specialist medical examinations or purchases made in pharmacies)].
3. **What is the legal basis for the processing of the surplus of information?**
   ….. [Indicate the specific legal basis].
   The legal basis could be represented by the data subject's consent or by a task carried out in the public interest: Article 6, paragraph 1, letter a) and e), of Regulation (EU) 2016/679, for personal data other than special categories of data, and Article 9, paragraph 1, letter a) and g) of Regulation (EU) 2016/679, for special categories of data.
4. **In case the legal basis is consent.**

   4.1 **When and through which methods is the data subject's consent obtained?**
   ….. [Describe].
   Specific issues occur in the case of minors. While it is necessary to preliminarily assess the existence of national regulatory requirements, it is preferable that, where the DPIA related to the statistical survey has documented the need to collect personal data referring to minors, the information should be collected from a third party. If the nature of the questions requires direct participation of the minor (e.g., survey of personal habits or opinions), the use of the smart feature should be specifically evaluated and justified, and, in case of use, it should be ensured that the consent for the use of the smart feature and the related processing of the surplus of information is provided by the parents or the person exercising parental authority. Specific measures should also be adopted to limit the scope and risks associated with the processing of the surplus of information related to minors, and the procedure under Article 36 of Regulation (EU) 2016/679 may be activated if necessary.
   In the case of subjects unable to provide consent, it will be necessary to resort to legal guardians.

4.2 **How and for how long is the documentation proving the acquisition of the data subject's consent kept?**

……. [Describe].

4.3 **How can the data subject withdraw the consent provided?**

….. [Describe].

The measures provided by the Data Controller for the revocation of consent must be similar or equally easy as those provided for its acquisition.

4.4 **In the case of withdrawal of consent, how is the deletion of the surplus of information guaranteed?**

….. [Describe the procedures put in place to comply with the withdrawal of consent and the techniques adopted for the deletion of data].

On this point, it is important to assess whether the data collected for the survey are also processed on the basis of the data subject's consent or not. In the first case, it will be necessary to evaluate whether it would be appropriate to set apart the two consents (one for the use of the smart feature and one for the survey data processing); this distinction will be possible (and recommended) if the participation of the data subject in the statistical survey can occur even without the smart feature. The distinction in the consent acquisition mechanism - and consequently in its withdrawal - will allow the Data Controller, in the case of revocation of the sole consent for the use of the smart feature, to continue processing the data collected for the survey.

5. **Is the data processing responsibility shared with other parties?**

5.1 No.

5.2 Yes. In this case, answer the following questions:

5.2.1 **Who are the joint controllers?**

…… [Indicate the name(s) of the joint controller(s)].

5.2.2 **Has a joint controllership agreement been drafted defining the responsibilities of each joint controller in the processing of personal data and, in particular, in the relationship with the data subjects, pursuant to Article 26 of Regulation (EU)?**

a) No. [Explain why the agreement has not been signed] ……

b) Yes. [Indicate the details of the agreement and attach a copy of the same to the DPIA].

6. **Which is the specific purpose for collecting and processing the surplus information?**

…… [Indicate the purpose pursued]

If multiple purposes are indicated, it is necessary to specify which data will be used for each purpose and whether it will be necessary to integrate the surplus of information with other data already in the possession of the data controller or specifically collected by them.

7. **Can the purposes indicated in the previous question be achieved using anonymous data?**

…… [Briefly explain the reasons that make the processing of personal data necessary]

8. **Which organizational and technical measures have been implemented to limit the possibility that the surplus of information is used for a purpose other than the one indicated in question 1?**

Type of measure	Adopted measure	Implementing rules (*description)*
Encryption	☐	
Pseudonimization	☐	

Anonimization	☐	
Other (*specify*) ……..	☐	

FAIRNESS

Personal data must not be processed in a discriminatory or misleading manner for the data subject, nor in a way that could cause harm to them. This principle refers to the ethical dimension, which must be observed in the processing of personal data.

The data subject must be aware of the purposes and means of the processing, including the risks associated with it, and the possibility to exercise the rights granted to them by Regulation (EU) 2016/679. These aspects also affect the validity of the consent given by the data subject to the processing.

1. **Is the data subject informed of the specific risks associated with the use of the smart feature?** (e.g., in the case of loss of the device and the data contained within it, and the measures to be adopted to mitigate these risks)
   1.1 Yes. …… [Describe briefly]
   1.2 No. …….. [Explain why]
2. **Is an explicit ethical approval asked before using the feature?**
   2.1 Yes. …… [Describe briefly]
   2.2 No. …….. [Explain why]
3. **Are viable (non-smart) alternatives available?**
   3.1 Yes. …… [Describe briefly]
   3.2 No. …….. [Explain why]
4. **Can the data subject choose to participate in the statistical survey without using the smart feature?**
   4.1 Yes. …… [Describe the alternative method of participation]
   4.2 No. …….. [Explain why]
5. **Are there support tools available to make the data subject more aware of the processing of their personal data?** (e.g., help line, video tutorials, dedicated sections on the website, FAQs)
   5.1 Yes. …… [Describe]
   5.2 No. …….. [Explain why]
6. **How is the data subject guaranteed the ability to exercise their rights concerning the processing of their personal data?** (e.g., procedures for submitting and managing requests)

   …….. [Briefly describe]
7. **Are there any exceptions or limitations to the exercise of the data subject's rights?**
   7.1 Yes. …… [Describe the exceptions or limitations]
   7.2 No.
8. **In case of processing based on consent, is the data subject informed of the procedure to withdraw consent**
   8.1 Yes. …… [Describe the procedure]
   8.2 No. ..... [Explain why]
9. **Was the choice of using a smart feature made in consultation with the data subjects or their representatives?** (e.g., through focus groups, consultations, testing phases)

9.1 Yes. …… [Indicate who was involved and how]
9.2 No.

10. **Is data processing carried out using algorithms?**
    10.1 Yes. …… [What measures are adopted to evaluate the proportionality of the algorithm to the purpose of the processing (principle of PURPOSE LIMITATION) and to eliminate any distortions, as well as to inform the data subjects of their use]
    10.2 No.

MINIMIZATION

The personal data processed must be adequate and relevant to the intended purpose.

In this context, the assessment of necessity concerns the surplus of information and the information stored or processed in the smart feature.

1. **Why can't the collection of surplus information be eliminated?**
    1.1 The smart feature is not customizable, or only partly customizable … [Explain why (e.g., inherent in the smart feature package , unacceptable decrease in data quality, or it would be too burdensome for the data subject)]
    1.2 The collection t is controllable but necessary… [Explain why it is not possible to identify alternative tools that could exclude or limit the processing]
2. **Are date handling: in-device, in-house, other person (processor)?**

    …… [Describe]

3. **Which are the measures taken to limit the amount of personal data processed ?**

Measure	**Adopted measure**	**Implementing rules (*description)***
Encryption	☐	
Pseudonimization	☐	
Pet (Privacy Enhancing Enhancing Techniques)	☐	
Aggregation (e.g., of some modes of the variables)	☐	
Partitioning (of the data in multiple environment)	☐	
Partial deletion (of unnecessary data)	☐	
Anonymization or Total secure data erasure	☐	

TRANSPARENCY

The data controller is obliged to inform the data subject of the purposes and methods of processing their personal data and to allow them, if necessary, to exercise their rights.

Information regarding the collection of surplus information and the operation of smart features must be provided before the processing begins, in clear, concise, and understandable language, and must be easily accessible to all data subjects.

1. **How is the data subject informed?**
   …… [Describe, specifying whether a layered approach (e.g., layered notices in a digital environment) and/or a multi-channel approach (e.g., in addition to written notices, the possibility for the data subject to access dedicated web pages or request clarifications via email or phone) is foreseen]
   The use of a layered notice and/or a multi-channel approach should be adopted when the nature, scope, and methods of the processing might be difficult for the data subject to understand.
2. **Which information is provided to the data subject regarding the surplus data collected and/or the processing carried out by or within the smart feature?**
   …… [Describe]
3. **Are specific measures adopted to ensure the information intelligibility to the data subject?** (e.g., translation into multiple languages, use of language that takes into account the specific characteristics of the respondents or their heterogeneity) or to facilitate accessibility (e.g., when information is provided using digital tools, measures are taken to ensure the accessibility of these tools, such as using machine-readable language)
   3.1 Yes. …… [Describe briefly]
   3.2 No. ……. [Explain why]
4. **Is the notice provided before the processing begins?**
   4.1Yes ……. [Describe when the notice is provided]
   4.2 No. ……. [Describe when the notice is provided and explain why it cannot be given beforehand]

ACCURACY

The personal data processed must be accurate and updated concerning the pursued purpose; inaccurate personal data should be corrected without delay or deleted.

The principle of accuracy is of course related to the accuracy gap and output gap connected to the use of the smart feature and the measures to be taken for its containment.

1. **In terms of the output gap, which measures are taken in order to avoid the storage of the irrelevant extracted information?**
   …….. [Briefly describe]
2. **Is there any additional data protection measures for the surplus data?**
   …….. [Briefly describe]
3. **What are the main risks related to data accuracy?**
   …….. [Briefly describe]
4. **What measures are in place to mitigate the risks of the accuracy gap?**
   …….. [Briefly describe]
5. **Are respondents asked to correct manually the data? Are they able to do this?**

In cases of data residing in the data subject's device, the controller may not be able to intervene directly to correct it. This aspect must be carefully evaluated and communicated to the data subject in the information. In this context, the possibility for the data subjects themselves to correct their own data must also be verified.

5.1 Yes. ..... [Briefly describe]

5.2 NO.

6. **Is the right of rectification of the data subject guaranteed?**

6.1 Yes. ..... [Describe the procedure and methods for intervening on the data (e.g., Until what stage of the processing process can the right be exercised?)]

6.2 No. ..... [Explain whether the lack of guarantee is due to technical or methodological problems (e.g., inability to re-identify the data subject with certainty) or to exemptions or limitations provided by European or national legislation.]

STORAGE LIMITATION

The controller must ensure that personal data is kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed. It is vital that the controller knows exactly what personal data the company processes and why. The purpose of the processing shall be the main criterion to decide in how long personal data shall be stored.

Considering that a surplus of data related to the use of the feature always could have as legal basis the consent, contrary to the survey questionnaire data which can be based on public interest (Article 6, paragraph 1, letter e) of Regulation (EU) 2016/679), it is considered appropriate that, in general, the data in surplus should be treated separately and possibly with additional protection.

1. **Is there a surplus of information in smart data relative to output needs?**

1.1 No. In this case, storage limitation is the same of storage time of the survey.

1.2 Yes. In this case, proceed to question 2.

2. **What are the expected retention choices for privacy-by-design?**

…..[Specify the retention policy of this surplus of data, the duration of storage and the justification for which such period is necessary, the procedures for deletion and/or anonymization of the data, the effectiveness of pseudonymisation/encryption techniques]

3. **Is it possible that surplus of data is also stored on the device of the respondent?**

3.1 No

3.2 Yes. In this case proceed to question 4

4. **What policies are used to inform the data subject of the risk of retaining such data on the device and to support the data subject in the deletion of the data?**

..... [Specify organizational measures planned and adopted, possible technical measures]

INTEGRITY AND CONFIDENTIALITY

With reference to the smart feature, describe how it is planned to protect personal data from unauthorized or unlawful processing and from accidental loss, destruction or damage, as well as what measures are taken to prevent and manage data breach incidents, ensure the proper execution of

data processing tasks and compliance with other principles, facilitating the effective exercise of individuals' rights.

Describe what additional data protection measures have been implemented because of the collection of smart data

1. **Have the risks to the security of personal data, the impact on individuals' rights been assessed? For use in risk assessment, a «threat modelling» has been developed and managed comprehensive, systematic and realistic analysis of the attack surface in relation to the specific software so as to reduce attack vectors and opportunities to exploit any weaknesses and vulnerabilities?** *(consider among others unauthorized and illicit treatments)*
   …….. [Describe]
2. **Have specific tests been planned from the design phase to concretely assess risks, particularly those arising from destruction, loss, modification, unauthorized disclosure, or accidental or unlawful access to personal data transmitted, stored, or otherwise processed?**
   2.1 Yes. ….. [Describe]
   2.2 No.
3. **Are there policies and procedures for periodic review and verification of software, hardware, systems, and specific services used to detect potential vulnerabilities in the data processing support systems?**
   (e.g., Change management and software lifecycle: static analysis, dynamic testing, penetration testing, interactive testing; Perimeter protection: runtime protection, web protection, etc.)
   3.1 Yes. …. [Describe the procedures or provide references to specific documents addressing these questions]
   3.2 No.
4. **Are personal data transfers protected against modifications and unauthorized or accidental access?**
   (e.g., channels used: HTTPS, SFTP, VPN, TLS, secure Wi-Fi)
   4.1 Yes. …. [Describe procedures and measures adopted]
   4.2 No
5. **Is data storage protected against modifications and unauthorized access? Is there a deletion mechanism to archive common data or permanently erase stored data at the end of its retention period?** (Consider all different data repositories throughout the statistical process.)
   5.1Yes. …. [Describe measures adopted]
   5.2 No
6. **Are necessary backups and event logs maintained for information security throughout the different phases of the statistical process?**
   6.1Yes. …. [Explain why]
   6.2 No
   6.3 Partially [Justify which parts of the process are not covered]
7. **Are security measures such as pseudonymization of personal data and backups/event logs used to minimize the risks of potential data breaches?**
   (e.g., pseudonymization and/or encryption of sensitive data variables using encryption techniques aligned with state-of-the-art standards)
   7.1Yes. …. [Describe measure adopted]
   7.2 No
8. **Are activity logs and event monitoring used as security controls on a routine basis, protected against unauthorized and accidental modifications and access, and periodically reviewed?**

8.1 Yes. …. [Describe the procedures or provide references to specific documents addressing these questions]

8.2 No

9. **Are there operational disaster recovery and business continuity procedures in place to restore the availability of personal data in the event of major incidents?** (Consider the requirement, under Directive (EU) 2022/2555 (NIS2), to have a disaster recovery plan and a business continuity plan.)

   9.1 Yes. …. [Describe the procedures or provide references to specific documents]

   9.2 No

10. **Are special categories of personal data adequately protected against security breach risks? Are high-risk data kept separate from other personal data?** (Always consider all phases of the statistical process.)

    10.1 Yes. …. [Describe the procedures or provide references to specific documents]

    10.2 No

11. **Are there methodologies, procedures, and resources in place to detect, mitigate, manage, and report data breaches while learning from them?**

    11.1 Yes. …. [Describe the procedures or provide references to specific documents addressing these questions]

    11.2 No

12. **Are there procedures for handling breaches and incidents, including notification procedures such as incident management and information handling?**

    12.1 Yes. …. [Describe the procedures or provide references to specific documents addressing these questions]

    12.2 No

13. **Which measures are put in place against the risk of data breach of the app, interception of data transfer and a breach of the backend?**

    ……. [Describe]

5. Involvement of a data processor and third parties

In this paragraph we introduce the involvement of a data processor and/or a third party in the smart survey.

A brief clarification:

- **processor** means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;
- **third party** means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under direct authority of the controller or processor, are authorised to process personal data.

Article 28 of the General Data Protection Regulation (GDPR) concerns the **data processor** and sets out the requirements that must be met when a data controller entrusts data processing to an external processor. Another reference in the GDPR are recitals 81 and 82, the first defines the specific requirements for the contract between the controller and the processor, stating it must be in writing and detail the responsibilities; the second emphasizes the importance of ensuring that processors do not subcontract data processing without the controller's consent.

Other reference points are:

- the guidelines 07/2020 on the concepts of controller and processor in the GDPR produced by the European Data Protection Board (EDPB), in which relationships between controller and processor are explained in detail ;
- the Commission implementing decision (EU) 2021/915 on standard contractual between controllers and processors

A data **processor** could for instance manage the cloud in which NSI applications run, manage all the software lifecycle for NSI, could treat only some personal data through a ML/AI service via API.

In the preparation of a DPIA, information about the presence of one or more data processors is part of the main section of the same DPIA, precisely because it will be important to apply all seven data protection principles also to the data processor. Considering that the success of a smart survey will depend not only on the effort of the NSI, but also on the correct involvement and expertise of the data processor, this aspect should not be underestimated. The following is a list of steps to be taken.

1. Assessment of the Data Processor
- Verify that the data processor is reliable and capable of ensuring compliance with the GDPR.
- Evaluate their experience, expertise, and technical and organizational security measures.

2. Definition of the Purpose of Processing
- Clearly identify the purposes for processing personal data.
- Specify the types of personal data that will be processed by the processor.

3. Drafting a Contract (DPA - Data Processing Agreement)
- Prepare a written agreement between the controller and the processor, as required by Article 28 of the GDPR.
- The contract must include:
    - The subject matter and duration of the processing.
    - The nature and purpose of the processing.

- The types of personal data and categories of data subjects.
- The obligations and rights of the controller.
- Technical and organizational security measures.
- Prohibition of the processor using the data for their own purposes.
- Confidentiality obligations.
- Procedures for managing sub-processing.
- Assistance to the controller in ensuring the rights of data subjects.
- Support in case of data breaches.
- Deletion or return of data at the end of the service.

4. Assessment of Sub-Processing

- If the processor intends to use sub-processors, the controller must be informed and must give a written authorization.
- Ensure that sub-processors adhere to the same contractual obligations as the main processor.

5. Evaluation of Security Measures

- Ensure that the processor has implemented adequate security measures to protect personal data.
- Request documentation on security policies, certifications, and audits.

6. Monitoring the Processor

- Establish a monitoring system to verify that the processor complies with contractual obligations and the GDPR.
- Request periodic reports or conduct audits if necessary.

7. Management of Data Breaches

- Agree with the processor on procedures for managing data breaches.
- Ensure that the processor promptly notifies the controller in case of a breach.

8. Support for Data Subject Rights

- Define how the processor will support the controller in responding to data subject requests (e.g., access, rectification, erasure of data).

9. Termination of the Relationship

- At the end of the relationship, ensure that the processor deletes or returns all personal data, as agreed in the contract.
- Verify that no copies of the data remain.

10. Documentation and Compliance

- Maintain complete documentation of the relationship with the processor, including the contract and assessments conducted.
- Ensure everything is aligned with GDPR requirements.

Within the SSI project and in any case in the wider container of the European statistical system (ESS), we are interested to consider as data processors companies that design, develop, provide and possibly manage data collection software using smart features. Here it becomes clear, already from the design phase of the smart survey, the actors that should come into play immediately.

- Department of Statistics: Provides the specifics of what is needed, in terms of required input and output, time, resources, respondents involved, to deploy the smart survey to the field;
- Department of data collection: provides the specifications on how to best implement the questionnaire, follows the usability and functional testing phases, manages the field data collection phase;

- Methodology department: provides methodological specifications on how to implement some smart features; analyses data from small test study to improve the methodology;
- IT department: Depending on the choices of implementation of the proposed solution, the commitment can vary greatly, because you can integrate and manage in your own systems implemented solutions (in-house) or, conversely, entrust all management to external (SaaS), with all possible intermediate variants;
- Legal department: You will have to deal with the legal aspects in the contract and later on regarding the privacy aspects;
- Procurement department: The contract or tender should be drawn up (among others, service level agreement and service level indicators need to be clarified) and all administrative aspects covered;
- DPO: must be informed of any new types of data processing that involve the drawing up of a DPIA, DPO should be consulted early enough to verify that data processing should or may require a DPIA.

Focusing on the ethical and legal aspect related to the processing of personal data in the involvement of a data processor the reference figures will be the ethics committee (if present), the legal office, DPO.

We want to highlight two aspects related to methodology and IT. Have the knowledge and competence of ML/AI algorithms used, parameters, models , tagging and training is a very sensitive issue, also from an ethical point of view, where it is required that in the decisions taken by algorithms the principle of human intervention and supervision (so-called "human in the loop") be guaranteed.

Regarding IT, even outsourcing the development and possibly the entire process of managing the software platform, it will be necessary to maintain a knowledge and competence on new technologies both in order to be able to examine with awareness the proposed solutions, the risks they may entail for the rights and freedom of data subjects, either to be able to integrate or to make these solutions interact with their own information systems, a subject that is not always trivial.

The issue of **Third Parties** is even more complex. These entities are autonomous, determine the purposes and means of processing data on their own. They do not act on behalf of the data controller but are responsible for their own processes, for example companies that use external marketing platforms to analyze and segment user data are considered autonomous controllers, therefore third party, as they determine the purpose of the processing (e.g., sending personalized advertising campaigns), also tools like Google Analytics process data autonomously to gather information about user behavior on a company's website and are considered third parties because they independently determine the data processing.

There are several disadvantages and risks of using Third Parties:

- **Limited Control:** The indirect involvement of third parties has a consequence that a NSI loses some direct control over data processing.
- **Data Security and GDPR compliance :** One of the main risks associated with third parties is data security. If the external provider does not adopt appropriate measures, the data could be vulnerable to hacking or unauthorized access.
- **Transparency and Consent:** With the involvement of multiple third parties, data subjects may not be informed about the involvement of the third party and therefore not be able to provide their consent or denial.

To mitigate risks and ensure GDPR compliance, NSIs should identify the third parties and, where possible, exclude (e.g. deactivate) them. In case it is not possible to exclude the third party, the data controller should provide the data subjects with clear, easily accessible, and comprehensive information, so they can exercise their rights knowingly.

6. Plan-do-check-act and legal-ethical evaluations

To date, smart survey applications still are a new and relatively unknown area in official statistics. Several statistical institutes have experimented with smart features and in some cases applications already made it to production. Nonetheless, all design levels are converging and so do risk assessments. However, also when designs have converged, the dynamics in technology, methodology and population uptake are such that frequent smaller and larger redesigns must be anticipated. Smaller redesigns may be seen as regular maintenance. Examples are changes in IT and technology such as app store policies and requirements, new versions of operating systems and new types and models of mobile devices. But also changes in the methodology/data science such as improved performance of AI/ML methods that are applied to smart data or different forms of active learning. Larger redesigns may come from advances in IT, technology and methodology/data science or from major shifts in public perceptions on the use of smart features. Examples are the inclusion of e-receipts in a household budget survey or the shifting perception on the use of location tracking by authorities.

All in all, we see three reasons to embed risk assessment in the GSBPM Evaluation phase and to view them explicitly as part of the PDCA-cycle of smart surveys:

1. Design levels of a smart survey application are changing maturity level
2. Regular dynamics and maintenance
3. Advances in technology, methodology and public perception

We refer to D4.3 for details on maturity and GSBPM building blocks in the Evaluation phase. Here, we focus on steps in the modular strategy that may be especially sensitive to the three mentioned changes.

Table 6.1 summarizes the proposed evaluations and actions. We distinguish three types of consequences: changes in risks, changes in the accuracy gap, changes in the output gap and change in the availability of alternatives to the smart feature(s). We explain the evaluations and actions per type of change.

Table 6.1: Evaluation phase for risk assessments

		Risks	Accuracy gap	Output gap	Alternatives
Maturity	Evaluation	More effective mitigation measures?	Accuracy gap smaller? Improvement ML?	Output gap smaller?	NA
	Action	Update descriptions	Update application-independent module	Update application-dependent module	NA
Maintenance	Evaluation	Is frequent default maintenance performed and aligned?	Are there changes in missing smart data? Are there changes in errors/outliers in smart data?	Has the surplus of information contained in smart data changed?	NA
	Action	If yes, update documentation If no, escalate as the DPIA may no longer be valid	If yes, perform an empirical assessment of the order of magnitude and the need to	If yes, re-assess ethics, the potential implementation of alternatives and the need for consulting public perception	NA

			increase the output gap		
Innovation	Evaluation	Have new risks emerged?	Has the smart feature changed categorization?	Has the output need changed?	Have new alternatives emerged? Have alternatives devaluated?
	Action	Update the identified risks and measure of the smart feature	Restart the modular strategy and if needed collect empirical support	Restart the modular strategy	Conduct the modular strategy and add findings

Maturity level: We make the assumption that a change in maturity level is an improvement. We see threats to the maturity level of a smart survey, i.e. a deteriorating maturity, as part of maintenance and/or innovation. However, in case maintenance is stalled or delayed or larger innovations are avoided, then a drop in maturity may occur and the DPIA may no longer be valid. Such events are addressed as well under the other two types of change. A growing maturity may come from many design decisions. In the legal-ethical context, the relevant improvements are more effective risk mitigation measures, smaller accuracy gaps through advanced use of technology or sophisticated methodology, and smaller output gaps through stronger data minimization. Output gaps and accuracy gaps go hand in hand. More accurate smart data may relax the need for rich quality metadata. Faster and more effective methods to adjust smart data errors may enable real-time processing and aggregation to data summaries. In all cases, the only action needed is to update documentation.

Maintenance: Maintenance consists of relatively modest year-to-year actions due to changes in IT, technology and methodology. We list possible changes to fix thoughts: (small) modernisations in the frontend UI, updating of libraries, updates in browsers, new OS versions, changes in mobile device sensor queries, updates to mobile app frameworks, app store policy changes, small code refactoring, bug fixes, updates to backend servers and data pipelines, certification and domain registration, security updates, updates to libraries used in machine learning software embedded in microservices, and active/online learning for trained machine learning models.
Many of these maintenance tasks are not smart survey specific and may be viewed as standard tasks any NSI has to conduct at a minimal frequency. Nonetheless, it is imperative that a DPIA remains fully aligned with these default tasks.
Some changes may, however, affect accuracy of smart data and/or change the surplus of information contained in smart data. A few examples: Policy of the European Union forbids the use of native routines in Chinese smartphones, affecting the frequency of location tracking data. New iOS models are more restrictive in location tracking battery management, causing more gaps in location tracking data. Stores may change the format of and information contained in their shopping receipts, making some information obsolete or changing the performance of product recognition and classification. This means that questions posed in Subsection 2.5 must be re-answered at a certain frequency in time, say once a year. A larger accuracy gap in smart data may demand for more quality metadata and less restrictive data minimization protocols to maintain accuracy of statistics. In other words a larger accuracy gap may imply a desire to increase the output gap. If so, also the application-dependent questions must looked at.

Innovation and major redesigns: Larger changes in time may demand for major redesigns and innovation projects in order not to lose maturity. These changes can be very diverse in nature. They can offer new opportunities but also make current design choices hard or infeasible. We identify a few realistic larger changes:

- Output specification revision: The specifications of the output behind the smart survey may be altered. In ESS-context, it may mean that the regulation is being modified. It could, however, also be that the smart data provide better and more detailed proxies of the concepts of interest. This is the case, for example, in energy statistics where energy meter data provide information that before could not be collected. A revised output need may obviously imply a different surplus of information collected in the smart survey and, consequently, a renewed look at the application-dependent parts of the smart feature module(s).
- New smart feature: A new smart feature may emerge that can be added to smart surveys. An example are e-receipts in online purchases or through apps deployed by stores. The feature may be added or even replace an existing feature. The modular strategy should be launched as explained in Section 2.2.
- Smart feature devaluation: An existing smart feature may loose potential. For example, printed receipts may become optional rather than default. Respondents have to ask explicitly to get paper copies. The accuracy gap of the smart feature may get larger to a point where it is hardly sustainable and alternatives have to be found. The implication is that the smart feature categorization may change and a renewed application of the modular strategy is imperative.
- Third part changes: Third parties may be involved in the execution of a smart feature or in the processing of smart data. Examples are external sensor systems such as activity trackers, energy meters and indoor climate systems. These parties may disappear, or introduce major accuracy or cost changes to the sensor systems. Other examples are in the processing such as the availability and richness of points-of-interest data in machine learning models. As a consequence a revisit of all empirical support is necessary.

7. Towards new smart surveys

An important ambition of SSI is to generalize to smart features and smart survey applications yet to come. The energy data donation case study, having a very different feature and aiming at a very different application, was included for this reason. Ultimately, only two SSI NSI's supported the case study. Also the other two case studies had to make stronger claims on time and budget than foreseen. Consequently, conclusions on some legal-ethical aspects cannot be derived from the energy data donation case study. However, there are still important and useful lessons-learned.

The energy data donation case study does show strong resemblance to case studies in physical activity tracking and indoor climate tracking. These studies have been performed outside SSI, but, in part, within Eurostat-project ESSnet Smart Surveys. In generalizing towards new smart surveys, we recommend to apply the modular strategy to these case studies. It likely gives important additional lessons in how to deal with new features.

The most important lesson comes from the application of the modular strategy to the energy data donation case study. We conclude that the questions that need to be answered for an 'undocumented' new feature and application are complete and valuable. The whole modular strategy and all follow-up steps were easily transferable to the new case study. We did not miss steps or questions. We could not answer all questions, but we feel we will have a complete view once empirical results are broader and more quantitative. Perhaps the only new element is the dependence on commercial operators, vendors of devices/sensor systems. They may 'mask' the true accuracy gap but also complicate privacy-by-design choices. This dependence may be relaxed by using devices oriented at the research-grade market in combination with tailored dedicated applications.

Obviously, in time, smart features and applications will emerge that are beyond our current classification and taxonomy. The dynamic and quickly progressing AI/ML likely will lead to new options that to date are hard to predict. We do believe, however, that the core of the modular strategy, accuracy gap versus output gap and the collaborative role of the respondent, will remain key.

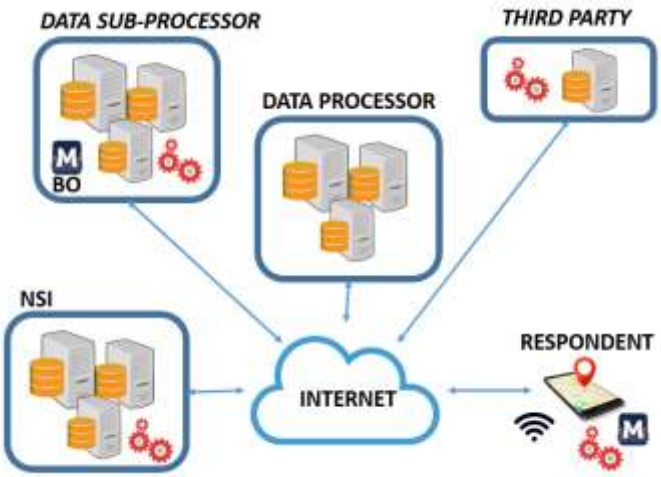
8. References

- Guidelines 01/2025 on pseudonymization – EDPB
- Guidelines 2/2023 on Technical Scope of Art. 5(3) of ePrivacy Directive – EDPB
- Guidelines 07/2020 on the concepts of controller and processor in the GDPR – EDPB
- Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0 – EDPB
- Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 (WP 248 rev.01 - 2017) - Working Party article 29
- EDPB opinions regarding the processing operations exempt from/subject to the requirement of a data protection impact assessment: https://www.edpb.europa.eu/our-work-tools/consistency-findings/opinions_en?f%5B0%5D=opinions_topics%3A138
- Privacy impact assessment (PIA) – Application to IOT devices – CNIL feb.2018 edition
- Practice guide GDPR – Security of personal data Version 2024 – CNIL
- Luiten, A., Toepoel, V., Schouten, B., Cierpiał-Wolan, M., Kapica, K., Szlachta, P., Lusyne, P., van der Beken, H. (2022a), Deliverable 2.1: health - measuring physical activity, ESSnet Smart Surveys, Statistics Netherlands, Heerlen, The Netherlands
- Luiten, A., Schouten, B., Blanke, K., Knapp, D., Volk, J., Lusyne, P. (2022b), Deliverable 2.1: Living conditions - Measuring indoor environment, ESSnet Smart Surveys, Statistics Netherlands, Heerlen, The Netherlands
- Kompier, M., Elevelt, A., Luiten, A., Mulder, J., Toepoel, V. (2024), Data donation of personal physical activity trackers, Survey Practice 17, available al https://www.surveypractice.org/article/124520-data-donation-of-physical-activity-trackers
- De Wolf, I., et al. (2025), How to integrate physical activity tracking data into self-reported physical activity?, CBS Discussion paper, forthcoming at www.cbs.nl
- Guidelines for securing the Internet of Things – Enisa
- Mobile Application Security Verification Standard – OWASP
- Web security testing guide – OWASP
- The PET guide – The United Nations guide on privacy-enhancing technologies for official statistics
- Emerging privacy enhancing technologies – Current regulatory and policy approaches – OECD Digital Economy Paper March 2023

Annex 1. Insights for risk analysis on smart surveys

In this section we will give some ideas on issues that need to be considered when developing a risk analysis of certain aspects of smart surveys, including the use of a data processor or third party. We do not want to go into detail on the risk analysis that requires a much more in-depth examination, but rather to point out some points of attention on new technologies and the related legal ethical aspects

An initial clarifying aspect may be a diagram describing the system infrastructure as a whole and the flow of personal or identification data between the actors involved



The diagram will show where data storage is planned, where business logic is planned, the presence of sub-processors, third parties, the first information outlining, among other things, the maps of the attack surface useful for risk analysis, to be included in the DPIA, highlighting the different responsibilities among the different actors involved (data Controller, data Processor, sub-processor, third-party, NSI).

The following are some points to be considered:

Software distribution policies

It is important to know the distribution type of software to be used, from which depend the different processing responsibilities.

- SaaS (Software as a Service) is a software distribution model where applications are hosted on remote servers and accessed via the Internet. Users do not need to install or manage the software locally but use it directly through a web browser;
- IaaS (Infrastructure as a Service) provides virtualized computing resources over the Internet, such as virtual machines, storage, and networking. Users have full control over the operating systems and applications but do not manage the physical hardware.
- PaaS (platform as a Service) provides a cloud platform for developing, testing, and deploying applications. It includes development tools, databases, middleware, and infrastructure, without requiring developers to manage the underlying hardware or operating system.
- CaaS (Container as a Service) provides container-based virtualization, allowing users to deploy and manage applications using container technology

- NmaaS (Namespace as a Service) provides an isolated namespace for organizing and managing resources in cloud or distributed environments. It is often used in containerization or microservices contexts. It allows users to create, manage, and utilize namespaces within a Kubernetes environment easily. In Kubernetes, a namespace provides a form of isolation within a cluster, dividing resources among multiple users or teams. This isolation can apply to applications, user access, storage volumes, and network traffic.

Service	Main Focus	Main Advantages	Main Disadvantages
SaaS	Application software	Accessibility, automatic updates	Dependence on Internet, limited customization
PaaS	Development platform	Faster development, infrastructure management	Limited customization, potential costs
IaaS	Virtualized infrastructure	Full control, scalability, no hardware investment	Requires technical expertise, higher management overhead
CaaS	Container management	Portability, automatic scaling	Learning curve, high costs
Nmaas	Resource isolation	Organization, multi-tenancy	Complexity, performance overhead

Depending on the solution chosen, the responsibilities for data processing can vary considerably, and therefore this should also be taken into account in the risk analysis and DPIA. As mentioned in the previous paragraph, several of these solutions may involve close collaboration between the service provider and the IT department at NSI. The on-premise choice, in which the software solution is deployed on your own systems, will require attention and collaboration with the software provider if the solution has also been developed partly externally.

Mobile application taxonomy

As regards the first two smart features that are based on sensors present on the respondent's smartphone, one should consider the type of software program used.

- Native Apps: They are built specifically for a mobile operating system (iOS or Android) using platform-specific programming languages. These apps are downloaded from an app store and run directly on the device.
- Web Apps: They are browser-based applications that run on mobile devices but don't need to be downloaded. They are accessed through a mobile web browser.
- Hybrid apps: They combine elements of both native and web apps. They are built using web technologies (HTML, CSS, JavaScript) but wrapped inside a native container to run on various platforms.
- Progressive Web Apps: PWAs are web apps designed to work offline and function like native apps, offering a near-native app experience without the need to install the app from an app store. They

are typically accessed via a web browser of your choice but can be installed and accessed on a device.

App Type	Platform Compatibility	Development Cost	Performance	User Experience	Access to Device Features
Native App	iOS, Android	High	Very High	Excellent	Full
Web App	Any (via web browser)	Low	Low	Fair	Limited
Hybrid App	iOS, Android	Medium	Medium	Good	Moderate
Progressive Web App	Any (via web browser)	Low	Medium	Good	Moderate

We must then proceed to the modeling of possible attacks and for this we start from the analysis of the attack surface. For our purpose of dealing only with the additional part related to the use of smart features in the preparation of the DPIA, also the risk analysis will be focused on how far it differs from the normal survey. So below we consider as attack surface only the device in the hands of the respondent

From the perspective of the rights and freedoms of the data subject, the fundamental point to consider is the use (for receipts and location tracking) of their own mobile device, over which, understandably, we have no control. As we well know from years of experience in CAPI (Computer-Assisted Personal Interviewing) surveys with interviewers, one of the key measures when using tablets for interviews was the pre-installation of mobile device management (MDM) software. This software allows for complete isolation of the application used and also enables administrators to remotely deploy OS updates or security patches, enforce password policies, and blacklist apps or device functionality. It can even remotely lock or erase data. However, this is not the case here, so we must consider the data subject's mobile device as an unsecure device.

Indeed, if we consider the common types of threats to a mobile device, we can easily notice how few countermeasures we can give implemented by the interested party

TIPOLOGY	DESCRIPTION	EXAMPLES	RISKS	COUNTERMEASURES
Malware	Malicious software designed to harm or exploit a mobile device	Trojan, spyware, ramsomware	Datatheft, information loss, device lockout	Install antivirus software, download apps only from trusted sources, keep the operating system updated
Phishing	Attempt to obtain sensitive information by pretending to be a trustworthy entity	Fraudolent emails or SMS	Credential theft, financial fraud	User education, use of anti-phishing filters, verifying sources

Man-in-the-Middle (MitM)	Interception of communications between the device and another system	Attacks on public Wi-Fi networks	Data theft, interception of sensitive information	Use VPNs, avoid insecure Wi-Fi networks, use HTTPS connections
Jailbreaking/ Rooting	Removal of restrictions imposed by the operating system	-	Exposure to malware, loss of warranties.	Avoid jailbreaking/rooting, keep the device in its original state
Fraudulent Apps	Apps that appear legitimate but contain malicious code		Data theft, unauthorized monitoring	Download apps only from official stores, check reviews and requested permissions
Exploitation of Vulnerabilities	Exploitation of bugs or vulnerabilities in the operating system or apps		Unauthorized access, device control	Regularly update the operating system and apps, apply security patches
Smishing and Vishing	Phishing via SMS (smishing) or voice calls (vishing).		Theft of sensitive information, fraud	Do not respond to suspicious messages or calls, verify the sender's identity
Physical Theft	Theft of the mobile device		Unauthorized access to data, information loss	Use strong passwords, enable remote lock, encrypt data
Interception of Calls and Messages	Interception of voice or text communications.		Privacy violation, information theft	Use encrypted messaging apps, avoid insecure networks.
Social Engineering Attacks	Psychological manipulation to obtain sensitive information		Data theft, unauthorized access	User education, verifying information requests

It will therefore be important that on the one hand, the data subject is made well aware of the risks and the ways to mitigate them (transparency), on the other hand, all possible security measures are implemented (integrity and confidentiality). We refer for example to the continuous updating of the software and related libraries, to the local database encryption where present (for example it can be present in cases of native, hybrid and pwa apps), a configuration that avoids the use of unsecured WIFI. For completeness it will be appropriate to verify or have certified that the provisions and suggestions for risk mitigation of ENISA and OWASP in this matter have been taken into account.

Microservices

A microservices architecture breaks down the traditional monolithic software deployment model into independent, distributed microservices that developers can deploy and scale separately. This software development approach builds a single application as a collection of small services. Each service runs within its own process and communicates with lightweight mechanisms, such as APIs and HTTP resources.

Developers build microservices around business capabilities, using automation to deploy them independently. A key advantage of the microservices architecture is that it enables writing services in various programming languages and using different data storage technologies. A microservices architecture is highly distributed and dynamic, introducing unique security risks. To address these risks, DevOps teams require a new approach to security. Ideally, teams should implement security into the design and architecture patterns, integrating security measures across the entire software development lifecycle (SDLC).

In general, on one side the use of microservices architecture specific to smart features is very interesting because it allows a decoupling between platform used and delivered services that can be called by APIs or http calls. On the other side microservices architecture is not free from attacks, but usually the attack surface widens. Even the technology on which microservices are implemented (containers and Kubernetes) is not free of attacks, so whether you choose to use a cloud or opt for on-premises, attention should be paid to continuous security updates of software and related libraries.

MITRE ATT&CK framework is an interesting way to simplify, categorize, and provide security recommendations for the various ways that Kubernetes and its components can be exploited. Below are the tactics and techniques representing the MITRE ATT&CK® Containers platform. The techniques below are known to target containers and container orchestration systems such as Kubernetes. The Matrix contains information for the Containers platform.

All this highlights the complexity of new technologies and architectures from the point of view of integrity and confidentiality and the need to have specific expertise on these issues.

Anti-tracking software

The "Guidelines 2/2023 on Technical Scope of Art. 5(3) of ePrivacy Directive" adopted by the EDPB on October 7, 2024, provide clarifications on the scope of Article 5(3) of the ePrivacy Directive, which concerns the storage or access to information on the user's terminal devices. This article applies not only to cookies but also to similar technologies used to track users online.

The guidelines identify three key elements for the applicability of Article 5(3):

1. Information: This refers to any data stored or accessed on the user's terminal device, whether personal or not;
2. Terminal device: Includes any equipment used by the user to access an electronic communications network, such as computers, smartphones, tablets, and IoT devices;

3. Access or storage: Refers to both the act of storing information on the user's terminal device and accessing information already stored.

The guidelines also provide a detailed analysis of various use cases, including:

- **Tracking via URLs and pixels**: Techniques that use unique URLs or invisible pixels to monitor users' online activities. A tracking pixel is a hyperlink to an asset, usually an image file, embedded in content such as a website or email. In the case of an email, for example, the sender may include a tracking pixel to detect when the recipient reads the email. Tracking pixels on websites are able to track user behavior;
- **Local processing**: Processes that occur directly on the user's device. "*If at any point and for example in the client-side code, the processed information is made available to a third-party, for example sent back over the network to a server, such an operation (instructed by the entity producing the client-side code distributed on the user terminal equipment) would constitute a 'gaining of access to information already stored without transferring data to external servers'*".
- **Tracking based solely on IP addresses**: Use of IP addresses to identify or track users.
- **Intermittent and mediated IoT communications**: Data exchanges between IoT devices and servers, which may involve access to or storage of information on terminal devices. For example IoT devices could be instructed by the manufacturer to stream the collected information, through the use of WIFI or a cellular SIM card, while storing it first locally until a connection is available. Other IoT devices, which do not have a direct connection, may be instructed to transmit information to another device (usually a smartphone) via a Bluetooth connection. In both situations, Article 5 (3) as through the instruction of the IoT device to send dynamically stored data to the remote server, an "access" occurs.
- **Unique identifiers**: Use of unique identifiers to recognize or track users across different sessions or devices.

The main goal of these guidelines is to ensure that tracking practices respect user privacy and comply with the ePrivacy Directive, especially in an ever-evolving technological context. It is important report that the applicability of this article does not systematically mean that consent needs to be collected. The EDPB reminds that in each case it would have to be assessed if a consent is needed or whether an exemption under Article 5(3) ePD could apply.

There are several open source tools that help to analyse these aspects, the same EDPB and EDPS have released open source software tools to analyse legal compliance.

Software Bill of Material (SBOM)

Software transparency involves providing clear and accurate information about the components used in an application, including their name, version, supplier, and any dependencies required by the component. This information helps identify and manage the risks associated with the software whilst also enabling compliance with relevant regulations and standards. With the growing importance of software in our daily lives, transparency is critical to building trust in software and ensuring that it is safe, secure, and reliable.

Software Bill of Materials (SBOMs) are the vehicle through which software transparency can be achieved. With SBOMs, parties throughout the software supply chain can leverage the information within to enable various use cases that would not otherwise be easily achievable. SBOMs play a vital role in promoting software transparency, allowing users to make informed decisions about the software they use.

The SBOM is extremely useful for software development teams of an external party, NSIs and end users. Its use can help ensure that open source and third-party components are up to date, and provides visibility into which project dependencies have known vulnerabilities that could be exploited in the software. Software buyers, on the other hand, can use SBOMs to analyse the risk inherent in a product through vulnerability assessments.

NSIs would benefit from working with their suppliers to ensure that they have access to correct and up-to-date information on the project components implemented in systems and/or products. They should also regularly evaluate their SBOMs to minimize the risks from using open source and third-party components.

Privacy enhancing techniques (PETs)

The Privacy enhancing techniques (PETs) comprise a set of tools and methodologies designed to strengthen the security of personal data. These technologies play a key role in reducing the visibility of sensitive data during processing and transfer, mitigating exposure and abuse risks. The effective adoption of PETs not only helps NSIs to comply with regulatory requirements, but also could provide greater confidence in data subjects.

Among the PETs, differential privacy, synthetic data and homomorphic encryption emerge as preferred techniques for their potential alignment with the principles of "data minimization" and "privacy by design", with practical results similar to anonymisation, masking the data at all times.

Within the SSI project, homomorphic encryption could be a solution that allows personal data to be processed in encrypted form, without ever exposing the data in plain text. This technique guarantees the confidentiality of data during the entire life cycle of the treatment, conforming to the principle of "privacy by design" and ensuring that the operations carried out on personal data do not compromise the confidentiality of the same.

Several pilot and POC projects are under way in the Eurostat community and throughout the statistical and non-statistical world. Although these techniques are not fully mature and readily available, since they do not include all possible operations on the data and require considerable computing power, there remains a field to be explored and possibly used for specific concrete cases.