CYBERNETICA

ESTAT 2019.0232 Data Protection Impact Assessment – Evaluation Report

Triin Siil, Riivo Talviste, Baldur Kubo, Ville Sokk, Toivo Vajakas

Version 1.10

19.11.2021

95 pages

Y-1440-3

Disclaimer

This document was prepared by Cybernetica AS as part of a procured project under Service Contract No ESTAT 2019.0232 (Ref. Ares(2020)2309804 - 30/04/2020). The opinions expressed in this document are those of the authors. They do not purport to reflect the opinions, views or official positions of the European Commission or its members.

Table of Contents

1. Introduction

1.1. Acronyms

CNIL	French Data Protection Authority (*Commission Nationale de l'Informatique et des Libertés*)
DPA	Data Protection Authority
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
ESS	European Statistical System
EU	European Union
MNO	Mobile Network Operator
NSI	National Statistics Office
PIA	Privacy Impact Assessment, as used in the CNIL Guides[1]
PSC	Pseudonymisation component of the Solution (deployed at MNO-ND)
SDC	Statistical disclosure controls
WP29	Article 29 Data Protection Working Party

1.2. Legal acts

Data Protection Directive (DPD)	Consolidated text: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data - amended by Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003 adapting to Council Decision 1999/468/EC the provisions relating to committees which assist the Commission in the exercise of its

[1] CNIL. Privacy Impact Assessment (PIA). Methodology. – In the Internet: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf (09.11.2020), see „Foreword“ p 2.

implementing powers laid down in instruments subject to the procedure referred to in Article 251 of the EC Treaty.

ePrivacy Directive (ePD)

Consolidated text: Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector

- amended by Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC (No longer in force, Date of end of validity: 03/05/2006)
- amended by Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws (Text with EEA relevance)

European Statistics Regulation (ESR)

Consolidated text: Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities

- amended by Regulation (EU) 2015/759 of the European Parliament and of the Council of 29 April 2015 amending Regulation (EC) No 223/2009 on European statistics (OJ L 123, 19.5.2015, p. 90–97)

European Data Protection Regulation (EDPR)

Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (Text with EEA relevance.), PE/31/2018/REV/1, *OJ L 295, 21.11.2018, p. 39–98 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)* – Internet: http://data.europa.eu/eli/reg/2018/1725/oj (04.04.2021).

General Data Protection Regulation (GDPR)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

1.3. Abstract

Cybernetica AS developed a proof-of-concept technical solution for a privacy-preserving statistical analysis of mobile location data for Eurostat. The goal was to improve the quality of official statistics by including more accurate data sources and updating the current statistics production methodologies accordingly. Synthetic mobile location data was used to test the solution. Relevant software development, risk assessment and user guide documents were created.

1.4. Scope

This Evaluation Report is meant to document and publish the results of the sample data protection impact assessment carried out concerning the further processing of pseudonymous mobile location data for producing official statistics, as further described below.

1.5. Target audience

This document is directed mainly to the EDPS and DPAs as well as the DPOs and legal personnel of interested NSIs and MNOs.

1.6. Dependencies

This document should be read and understood in conjunction with the following related deliverables under the Agreement:

1) ESTAT 2019.0232 Solution analysis
2) ESTAT 2019.0232 Solution architecture
3) ESTAT 2019.0232 Data Protection Impact Assessment – Scoping Report

2. Context

2.1. Nature and purposes of the processing

MNOs record and store certain types of data for purposes related to delivering telecommunications services (e.g., Call Detail Records collected for billing) and/or in support of network operation (signalling data), which embed information about the (approximate) position and movements of the mobile devices using the MNO network ("**mobile location data**"). Such position and movements of a mobile device can be attributed to a specific customer and/or individual who is using the MNO services ("**Subscriber**"[2]).

In addition to the primary purposes for which mobile location data is initially collected by MNOs, the data can also be useful for secondary purposes. For example, MNOs can provide certain value-added services to Subscribers based on analysis of mobile location data. Some MNOs apply short-term pseudonymisation to mobile device identifiers in order to prevent long-term tracking of the Subscribers during secondary use of mobile location data and decrease the risk and the potential impact of personal re-identification for mobile location data. Against this background, European Union Statistical Authority („**Eurostat**") has defined a reference scenario whereby it is assumed that this short-term pseudonymisation is based on a secret key that is changed periodically so that the generated short-term pseudonyms are different in each pseudonymization period ("**Reference Scenario**"). According to the Reference Scenario, as soon as a new key for the next pseudonymisation period is generated, the previous pseudonymisation key is deleted ("**change-and-forget method**").

The secondary use of mobile location data is of growing interest also to members of the ESS, comprising Eurostat and the NSIs. Traditionally, with a few exceptions, NSIs collected data directly from the data subjects via surveys or censuses, where respondents had to provide information on themselves. During the last decade, NSIs have started to extend their scope towards reusing secondary data sources for statistical purposes. First, statistical offices requested access to administrative data (data collected by public authorities) and the EU statistical legislations were amended accordingly.[3] Next, the current trend is to expand the secondary data sources to so-called "big data" that are often collected by private entities – similarly to administrative data sources, it may require explicit reference in statistical legislation to be recognised by potential data holders.[4]

In light of these circumstances, NSIs within the ESS are considering mobile location data as a potential new data source for the production of future official statistics. In principle, mobile location data can be used to extract information serving multiple statistical applications and use-cases,[5] e.g. enabling statistical reports about spatial density of present population

[2] For ease of reference, we assume the „Subscriber" includes both contractual clients of MNOs and any third parties who use the MNO's network by means of a Subscriber's mobile device.
[3] G. Somers. TASK 3: Legal review Deliverables: D.3.2 Report on legal review covering basic statistical laws and framework legislations D.3.3 Report on legal review covering other legislations. Services concerning ethical, communicational, skills issues and methodological cooperation related to the use of Big Data in European statistics (Contract number 11104.2015.005-2015.799). time.lex, 10 August 2017, p 28. – in the Internet: https://ec.europa.eu/eurostat/cros/system/files/deliverables_3.2_and_3.3_legal_review_final.pdf (04.12.2020).
[4] *Ibid.*
[5] F. Ricciato et al. Towards a methodological framework for estimating present population density from mobile network operator data, p 3. – Pervasive and Mobile Computing. Volume 68, October 2020, in the Internet: https://doi.org/10.1016/j.pmcj.2020.101263 (31.12.2020).

density and patterns of human mobility. In order to reuse the mobile location data for such purposes in the context of the Reference Scenario, an NSI would need to be able to draw insights from multiple records of the same mobile device over a period of time, which is longer than a single pseudonymisation period.

The processing operations subject to this DPIA process enable the extraction of statistics based on long-term analysis of mobile location data that are pseudonymised with short-term pseudonyms, while not increasing the level of risk and potential impact of re-identification of Subscribers beyond the time interval covered by a given short-term pseudonym. This is achieved thanks to privacy-enhancing technologies, which enable a new model for privacy-conscientious use of mobile location data of producing official statistics ("**Solution**"). It opens up the possibility to create new types of statistics production processes as well as provide new insights in established statistics production processes based on mobile location data. Ultimately, the aim is to apply the Solution in real-world scenarios of producing official statistics in the future, while keeping the current business processes at the MNO intact.

2.2. Scope of the DPIA

The first instance of the DPIA ("**Sample DPIA**") serves as a basis and guiding model for a potential partnership between an NSI and an MNO in the process of seeking authorisation by the relevant DPA at the national level and/or by the EDPS for the actual implementation of the Solution. It marks the start of a continual and iterative process of building and demonstrating compliance of the Solution with data protection laws. The Sample DPIA was carried out between November 2020 and May 2021, resulting in two reports ("**Sample DPIA Reports**"):

1) **"Scoping Report"** – documenting the questions raised, decisions made and the justifications relied on.

2) **"Evaluation Report"** – documenting the legal and technical risk analysis.

The Solution is envisioned to be adopted in incremental stages, starting with proof-of-concept stage, followed by pilot project stage and finally, production stage. The focus of the Sample DPIA is the proof-of-concept stage. In this stage, the basic functionality of the Solution is developed, tested and demonstrated using synthetically generated mobile location data. In the next stages, the Sample DPIA Reports will need to be revised, responding to updates in the development process as well as changes in the risks resulting from the use of real-world mobile location data. However, it is expected that the revisions will mainly concern the specific statistical analysis use case implemented by means of the Solution and the internal organisational setting of the NSI and MNO, whereas the Sample DPIA was conducted as if the Solution were already processing real-world mobile location data.

The scope of the Sample DPIA is defined at a conceptual level rather than in a concrete use case level. The reason for opting for the conceptual level is to reduce the overall complexity deriving from the need to integrate the NSI's statistical analysis process with the MNO's business process and bring it to a minimum acceptable level. The identification of potential statistical analysis use cases suitable for implementing in real-world scenarios by means of the Solution, as well as the design of appropriate statistical methodologies, is a work in progress – the exact use cases and statistical methods remain to be specified as a result of ongoing development efforts by the ESS. Once the choice has been made, the selected use cases along with the accompanying statistical methodologies will need to be

subjected to a “real” DPIA, which may be produced taking the Sample DPIA as a starting point.

Nevertheless, there is some initial information available which can be used to sketch out an example of how the Solution may be applied to a close-to-real-world statistical analysis use case in the future. This initial information was used to create a simplified version of a potential use case and the accompanying statistical methodology under consideration (see Section 2.5 below). Although far from complete, the initial information is helpful in understanding the feasibility and level of risks associated with a statistical analysis use case involving mobile location data. It is a vital starting point for involving specialists from different domains in a discourse concerning the viability and legitimacy of using mobile location data for producing official statistics by means of privacy-enhancing technologies.

For the reasons provided above, there will be some gaps in the documents produced as a result of the Sample DPIA, which will need to be filled gradually as new knowledge becomes available. It is a first step in a way towards demonstrating the technical feasibility and legal compliance of implementing privacy-enhancing technologies in facilitating secondary use of big data from sources outside the ESS, while maintaining the level of data protection and statistical confidentiality required under applicable laws and regulations. The relevant NSIs deploying the Solution in the pilot project and production stages in the future will need to carry out their own respective DPIA with regard to the real-world scenarios and specific statistical analysis use cases at hand. Furthermore, a separate legislative procedure may be needed in order to create a suitable legal framework to support the deployment of the Solution in a real-world situation, unless an appropriate legal basis is already available in the relevant national law. In both cases, the Sample DPIA should speed up the process by offering a basic understanding of the functioning of the Solution and the related risk management considerations with regard to protecting the fundamental rights and freedoms of individuals.

The Sample DPIA does not address:

1) **the initial conditions which determine whether or not a DPIA needs to be carried out**[6] – this question is usually addressed in the preparatory stage of a DPIA and thus analysed in a separate document. For the purposes of the Sample DPIA, it is presumed that a DPIA needs to be carried out in order to adopt the Solution as a means of further processing pseudonymous mobile location data for producing official statistics.
2) **the subsequent conditions which determine whether or not the supervisory authority needs to be consulted with**[7] – this question is usually addressed in the evaluation stage of a DPIA process and thus analysed in a separate document. For the purposes of the Sample DPIA, it is presumed that a supervisory authority needs to be consulted and this document can be presented to the relevant authority in order to assist such consultation.
3) **the specific legal basis for further processing pseudonymous mobile location data for the purposes of producing official statistics (including experimental statistics)** – the legal basis for conducting the processing activities envisioned in this document may derive from the national law of a Member State or from EU law. The Sample DPIA covers the legal grounds in EU law, which provide the conditions for choosing the most suitable legal basis under national law of a Member State or EU law. Because a specific Member State cannot be defined in this document due to the proof-of-concept nature of the use case at hand, the concrete legal basis will

[6] GDPR art 35(1).
[7] GDPR art 36(1).

need to be specified in the next stage, i.e. a pilot project or production stage in a real-world setting.

4) **the risk analysis of any components existing prior to deploying the Solution.** The scope of the Sample DPIA takes existing infrastructure and analysis processes (if any) at MNO as baseline and concentrates on processes and components added by the Solution itself. This is due to the fact that this Sample DPIA is meant as a template for different MNOs and NSIs with their own infrastructure and internal processes and the authors cannot make any assumptions on these.

2.3. Methodology

This document is primarily based on the PIA methodology developed by the CNIL, which is outlined in the CNIL guides[8]:

1) PIA methodology (setting out the approach),
2) PIA templates (facts that could be used for formalising the analysis),
3) PIA knowledge bases (a catalogue of controls aimed at complying with the legal requirements and treating the risks, and examples).

The data protection risk analysis in Section 3 should be read and understood in conjunction with the PIA methodology outlined in the CNIL guides – where there is no additional explanation added in the text of the document, the relevant CNIL guide acts as a support instrument for the reader.

Privacy risks analysis in Section 4 is based on technical risk analysis included as an appendix for this document. The technical risk analysis uses the STRIDE methodology to enumerate possible threats to processes and assets used in the Solution. The STRIDE methodology was used as it provides a systematic approach to identify possible threats so that no threat is overlooked. Moreover, it works on a data flow diagram that is a simplified version of the business process diagrams shown in the Solution analysis document[9].

For each identified threat, the technical risk analysis lists applied mitigations and any residual risks. The latter are then analysed further in the context of privacy in Section 4.2.

The final stage of “Formal Validation” was not completed, due to the proof-of-concept nature of the Sample DPIA. However, relevant further action points were provided, in order to facilitate the next steps.

2.4. Background concerning the territorial dimension of European statistics

The European Commission has recently made available data for several different territorial typologies across the EU, which has stimulated policymakers to carry out new kinds of policy analyses using a territorial dimension. The main territorial typologies can be divided into three different groups:[10]

1) **grid typologies** – Eurostat collects population statistics based on 1 km² grid cells. These are very detailed statistics, which are used to establish various cluster types — namely, urban centres, urban clusters and rural grid cells.

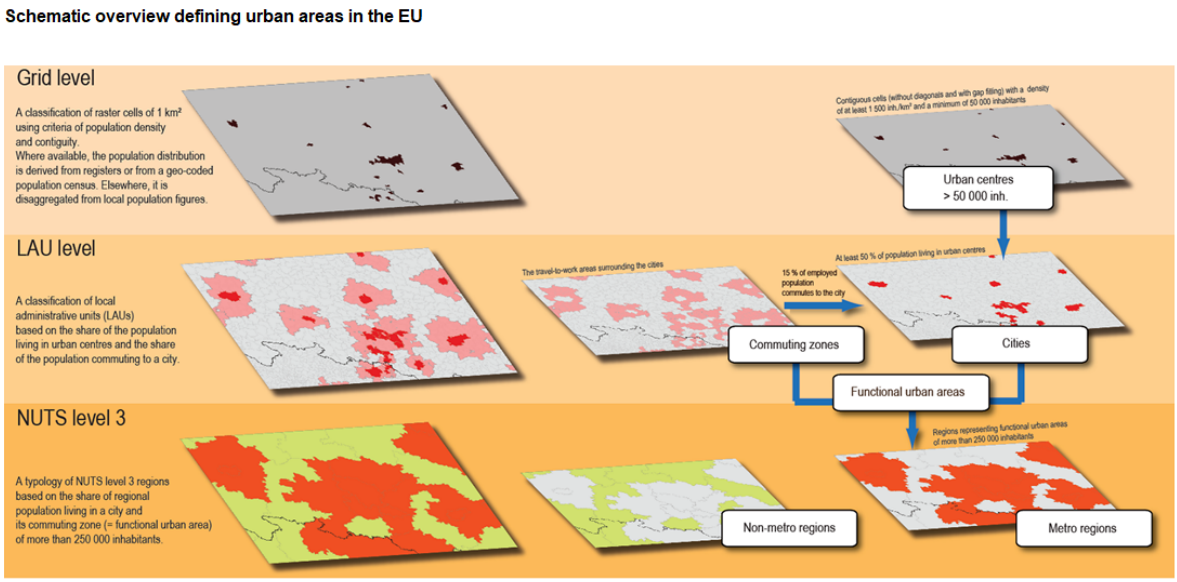
[8] CNIL. Privacy Impact Assessment (PIA). – In the Internet: https://www.cnil.fr/en/PIA-privacy-impact-assessment-en (09.11.2020).
[9] See: ESTAT 2019.0232 Solution analysis.
[10] Eurostat. The Methodological manual on territorial typologies. 2018 edition, p 7. – Internet: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-18-008 (23.04.2021).

2) **local typologies** – based on statistics for local administrative units (LAU), such as municipalities or communes across the EU. These statistics may be used to establish local typologies including the degree of urbanisation (cities; towns and suburbs; rural areas); functional urban areas (cities and their surrounding commuting zones); coastal areas (coastal and non-coastal areas).
3) **regional typologies** – statistics that are grouped according to the classification of territorial units for statistics (NUTS). They provide information at a relatively aggregated level of detail.[11]

The three different types of territorial typologies are all based on the same basic building blocks – classifying population grid cells to different cluster types and then aggregating this information either by LAU or by region to produce statistics for a wide variety of different typologies.[12]

Figure 1 below presents an example for how urban areas in the EU are defined at three different — but coherent — levels:

Figure 1 – Schematic overview defining urban areas in the EU[13]



Note: for more information, http://ec.europa.eu/regional_policy/sources/docgener/focus/2012_01_city.pdf
Source: European Commission, Directorate-General Regional and Urban Policy, based on data from Eurostat, JRC, national statistical authorities

eurostat

A population grid is one of the three basic building blocks that underpin the various territorial typologies. It is composed of a set of equally-sized cells containing population counts for each cell. Eurostat prefers the use of a 1km^2 square grid that is overlaid across the EU territory.[14] In practice, the population distribution data underlying the grid level is currently derived from registers or from a geo-coded population census or disaggregated from local population figures (see Figure 1 above).

Functional urban area ("**FUA**") is one of the local territorial typologies described above. It consists of a "city" (densely inhabited) and its "commuting zone" (less densely populated)

[11] *Ibid.*
[12] *Ibid.*
[13] *Ibid.*, pp 8-9.
[14] *Ibid.*, p 13.

whose labour market is highly integrated with the city.[15] "City" is a LAU where at least 50 % of the population lives in one or more urban centres.[16] "Commuting zone" contains the surrounding travel-to-work areas of a city where at least 15 % of employed residents are working in a city.[17] FUAs do not cover the whole territory of a country but rather the more densely populated areas. FUA as a type of classification is linked to other classification types, such as the degree of urbanisation and typology for metropolitan regions. It is used as a basis for the city statistics data collection.[18] Currently, there is no EU legislation on the collection of city statistics and they are provided on a voluntary basis only.[19]

Based on the current approach to including a territorial dimension in official statistics in the EU, it is possible to envision alternative population grid statistics and territorial typologies when using mobile location data as a source for evaluating population distribution. Inspired by the concept of FUA, the Eurostat staff has developed, solely for the purpose of the project described in the Sample DPIA, an analogous concept based on mobile location data, which indicates an approximation of cities and their commuting zones within the scope of what is realistically achievable by mobile location data – it is hereinafter referred to as Functional Urban Fingerprint[20] ("**FUF**").

2.5. Use case description

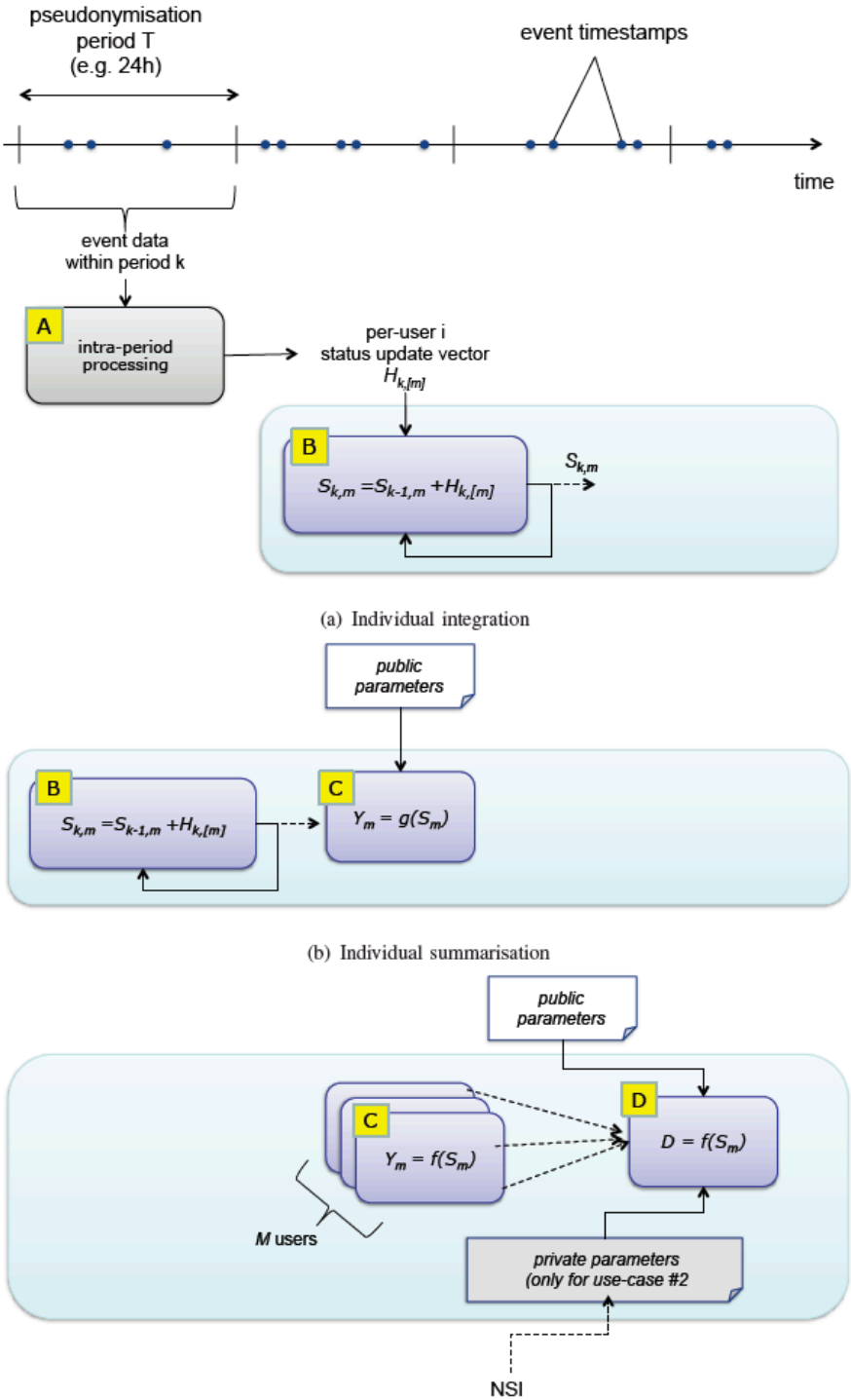
The target of evaluation of the Sample DPIA is the further processing of pseudonymous mobile location data by means of the Solution for the purposes of official statistics. The evaluation is conducted at a conceptual level, focusing on the change in the level of risks to the fundamental rights and freedoms of Subscribers, which may occur due to introducing mobile location data as a new source for computing the territorial dimension in official statistics by means of the Solution.

In order to better illustrate how the Solution may be applied in a close-to-real-world setting in the future, a simplified version of a potential statistical analysis use case along with the accompanying statistical methodology in the context of the Reference Scenario was created ("**Sample Use Case**"). The Sample Use Case has been designed to test how the current statistics production processes using the FUA concept can be emulated by means of the Solution, if the underlying population grid were derived from mobile location data. The Sample DPIA is not meant to assure the compliance of the Sample Use Case with applicable laws nor to evaluate the specific security and privacy risks related to the Sample Use Case – this shall be an object of future assessment. The Sample Use Case should be treated merely as a tool to measure how efficiently the privacy-preserving functionalities of the Solution work and what could be improved in terms of scalability, so as to facilitate identification and development of potential statistical analysis use cases suitable for implementing in real-world scenarios by means of the Solution in the future.

The schematic flow of the Sample Use Case is depicted in Figure 2 below.

[15] Eurostat. Statistics explained. Glossary. Functional urban area. – Internet: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Functional_urban_area (23.04.2021).
[16] Eurostat. Statistics explained. Glossary. City. – https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:City (23.04.2021).
[17] Eurostat. Statistics explained. Glossary. Commuting zone. – Internet: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Commuting_zone (23.04.2021).
[18] *Ibid.*, pp 59-60.
[19] Eurosat. Cities. Background. – Internet: https://ec.europa.eu/eurostat/web/cities/background (26.04.2021).
[20] Note the difference between the term "footprint", referring to data structures associated with individual mobile devices, and "fingerprint", for aggregate data structures associated with grid cells or Reference Areas.

Figure 2 – Schematic flow: the private computation blocks are embedded in the cyan box.[21]



[21] See: Specification of test use-cases for project ESTAT 2019.0232, p 9.

The Sample Use Case follows a step-by-step approach, whereby the pseudonymous mobile location data are gradually de-identified, each next step reducing the link between the mobile location data and the relevant Subscriber until its elimination. These activities can be observed in two levels:

1) "**Meta Level**" – the structural elements of the Sample Use Case (Module A, Module B, Module C and Module D), which will most probably be preserved in the official methodology used in the statistical analysis use cases implemented for real-world scenarios in the future;
2) "**Use Case Level**" – the dynamic elements of the Sample Use Case, which can differ from one statistical analysis use case to another, when implemented for real-world scenarios in the future (number of countries, amount of Subscribers, content of public and private parameters, specific computation algorithms, choice of SDC techniques etc).

The step-by-step approach of the Sample Use Case has three stages:

i. Individual integration

This stage is divided into two parts:

a) Pre-processing stage (Module A)

The data processing is carried out at the level of an individual Subscriber. Each Subscriber's mobile device has been previously assigned a pseudonym, which changes periodically after each 24h interval. The MNO provides mobile location data concerning cells in a 1km x 1km population grid that pseudonymous mobile devices have visited during each 24h interval.

Module A integrates all visits per each pseudonymous mobile device during each 24h period into a data structure ("**Intra-Period Footprint**"). The Intra-Period Footprint provides a score for each grid cell depending on how often and how long a pseudonymous mobile device spent time in it ("**Footprint Score**"). The Footprint Score is computed from a sequence of visits using a predefined algorithm provided by the NSI (public information).

The pre-processing stage carried out in the Module A is not part of the Solution.

b) Accumulation of individual footprint (Module B)

The Intra-Period Footprints and pseudonyms of mobile devices are provided as input to the Solution.

The first part of the Solution – Module B – links the different pseudonyms associated with the same mobile device across all 24h intervals and adds the Intra-Period Footprints of all associated pseudonyms together. As a result, all visits per each (non-pseudonymous) mobile device over a long period can be summarised in a data structure ("**Longitudinal Footprint**").

ii. Individual summarisation

The second part of the Solution – Module C – takes as input the Longitudinal Footprint for each single (non-pseudonymous) mobile device and selects the grid cells that have been visited more frequently and more regularly by a given mobile device. The output of Module C is a data structure consolidating the Longitudinal Footprints of all single mobile devices ("**Consolidated Footprint**") and, thus, still refers to such individual devices. The Consolidated Footprint has only a handful of non-zero entries (grid cells) and aims at representing the "usual environment" of a generic individual device.

The thresholds for determining most visited grid cells per mobile device will be provided as input to Module C by Eurostat (public information). Based on that, the top ranked grid cells ("**Top Tiles**") will be calculated per each mobile device in Module C and will therefore represent the estimated usual environment zones for each mobile device.

iii. Collective aggregation

The last part of the Solution – Module D – takes as input the Consolidated Footprint and computes aggregated statistics over the whole population of mobile devices.

An approximation of cities is given based on a pre-defined list of territories, which roughly correspond to administrative urban areas ("**Reference Areas**"). Each Reference Area is a list of contiguous grid cells. The Reference Areas are provided by the NSI as an input to the Solution (public information).

An approximation of commuting zones is calculated within the scope of what is realistically achievable by mobile location data. In order to estimate the load of movement between a city and its commuting areas, a value is calculated which indicates how strongly a grid cell outside the Reference Area is connected to the Reference Area. This value is related to the share of devices having a Consolidated Footprint intersecting both the grid cell and the Reference Area.

The NSI has the option to provide a secret input (can not be visible neither to the MNO nor to any other third party), which gives the resident count of each grid cell ("**Resident Count**"). The Resident Count has been estimated using a population census and can be used for calibrating the results of the analysis.

Statistical Disclosure Control ("**SDC**") is applied to the results of the analysis as a last step before releasing the output. SDC omits grid cell values that do not exceed a pre-defined threshold, in order to comply with the statistical principles laid out in the Treaty on the Functioning of the European Union[22] ("**TFEU**") Art 338 and Article 2 of the ESR[23]. This threshold is fixed beforehand and cannot be changed after the Solution has been deployed. In order to change the SDC threshold, it has to be changed in the source code, the new source code has to be compiled, the new version has to be deployed and enforcers have to approve it again. The state of the previous version will not be accessible in the newly deployed enclave. The threshold values are made public, along with a detailed description of the SDC method and the associated code.

In the last step, the following reports are made available as a result of the analysis:

1) **Fingerprint Report** – indicates for each grid cell how many mobile devices are typically found in this grid cell at a given time of day.
2) **Population Density Report** – indicates for each grid cell how many mobile devices had this grid cell as their No 1 Top Tile, corresponding to the most likely main place of living.
3) **FUF Report** – indicates an approximation of cities and their commuting zones within the scope of what is realistically achievable by mobile location data.

In addition to the reports, the following aggregate results are reported:

1) **Highly Nomadic Users** – the number of Subscribers who did not have Top Tiles or where their computation did not succeed.
2) **Observed Total Users** – the total number of observed individual Subscribers.

[22] Treaty on the Functioning of the European Union. Consolidated text: Consolidated version of the Treaty on the Functioning of the European Union – Internet: http://data.europa.eu/eli/treaty/tfeu_2016/2020-03-01 (04.04.2021).
[23] The thresholds are provided and calculated by the NSI in accordance with relevant SDC methodologies.

3) **Adjusted Total Users** – the total number of observed individual Subscribers after the optional calibration.

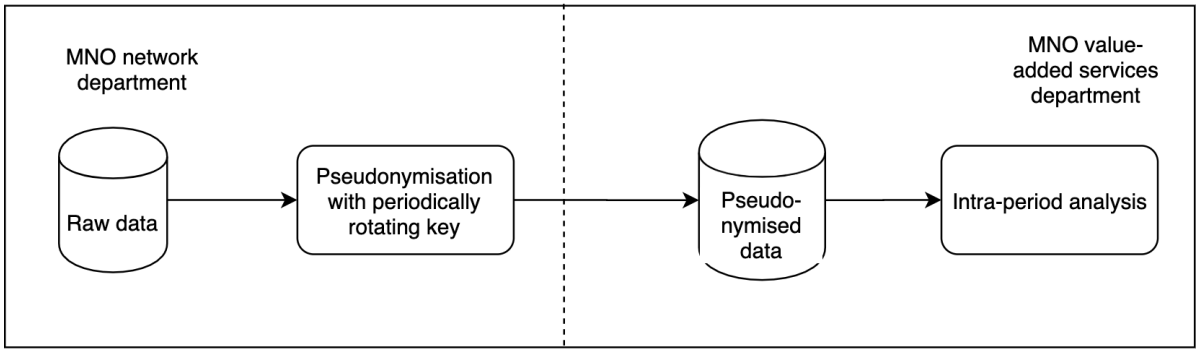
The Sample Use Case is further quantified by the following parameters:

1) The mobile location data used for statistical analysis by means of the Solution is provided by one hypothetical MNO. This is the only source of mobile location data, i.e., data from other sources (e.g. customer relationship management systems) is not used.
2) The mobile network of the MNO covers the full national territory of a hypothetical country, which is an EU Member State.
3) The MNO has approximately 100 000 000 Subscribers in the EU Member State.
4) The Subscribers' pseudonymous mobile location data is stored for the purposes of statistical analysis by means of the Solution for up to one year.
5) The statistical analysis by means of the Solution is ordered by one hypothetical NSI, located in the same country as the mobile network of the MNO.
6) The domain of official statistics and the broader production process embedding the Sample Use Case is unspecified.
7) In order provide a realistic assessment of the potential change in risk level and impact on Subscribers, it is presumed that the further processing of pseudonymous mobile location data is based on actual real-world personal data. Otherwise, the relevant data protection laws will not apply.

2.6. Solution design

One of the key requirements of the Solution was to add minimum overhead and changes to the pre-existing pseudonymisation and value-added service provision processes at the MNO, which is sketched in Figure 3 below:

Figure 3 – The current process of sharing and analysing mobile location data[24]

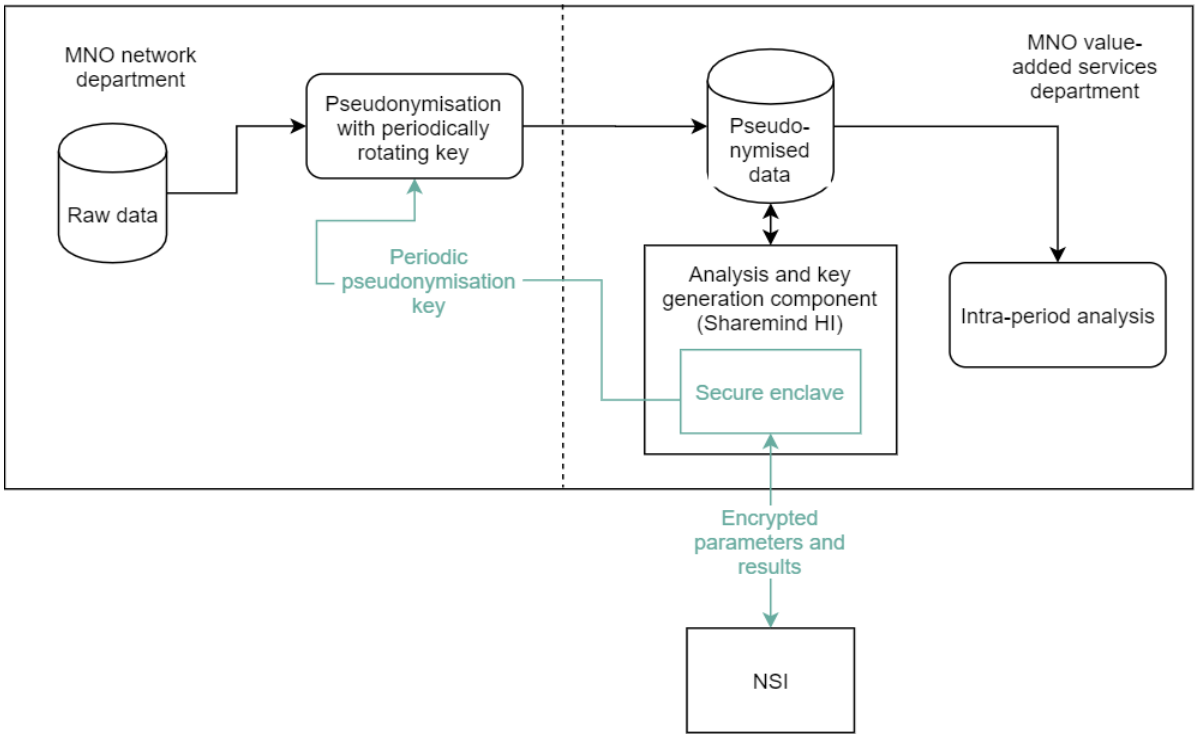


In order to meet the requirement to maintain the current processes at the MNO as well as other requirements deriving from the change-and-forget method and the Sample Use

[24] See: Solution Analysis. Figure 1: The current process of sharing and analysing mobile location data.

Case,[25] the Solution was implemented on the Sharemind HI platform[26] using the architecture depicted in Figure 4 below. Sharemind HI development platform relies on a Trusted Execution Environment ("**TEE**") technology. A TEE isolates security sensitive parts of an application from the rest of the system with the help of trusted hardware. The TEE technology used in Sharemind HI is Intel® Software Guard Extensions ("**SGX**"), which is available in modern Intel® processors.[27] In the Solution architecture, the TEE is located within the Analysis and key generation component of Sharemind HI – it is comprised of "enclaves" in Intel® SGX terms ("**enclave**" or "**enclaves**").

Figure 4 – Proposed secure architecture of sharing and analysing mobile location data[28]



According to the proposed architecture, the Solution has two main functions:

1) **to generate periodic keys for pseudonymising mobile location data** – when the Solution is active, raw mobile location data is pseudonymised by the MNO-ND at its premises using the periodic pseudonymisation key generated in the enclave. While the enclave is physically on a server processor located at the MNO-VAD, MNO-VAD does not have access to or control over it, other than allowing the set-up and (de-)activation of the Solution. This means that MNO-VAD, or any third party for that matter, is not able to access or otherwise make use of the pseudonymisation keys stored in the enclave.
2) **to perform data analysis tasks on pseudonymised mobile location data** – as the periodic pseudonymisation keys are stored in the Trusted Execution Environment (TEE), they can be used to reverse the pseudonymisation of mobile location data inside the Trusted Execution Environment (TEE) in order to carry out

[25] The specific technical requirements are provided in the Solution Architecture Document.
[26] The Sharemind HI platform is described in more detail in the other deliverables of the Project. – See: Solution Analysis. Chapter 3.
[27] See: Solution Analysis Document. Section 3.3.
[28] See: Solution Analysis Document. Figure 3: Proposed secure architecture of sharing and analysing mobile location data.

data analysis. Both the reversal of pseudonymisation and the analysis are carried out in Solution enclaves within the TEE, ensuring that neither MNO-ND nor MNO-VAD has access to or is otherwise able to make use of the mobile location data. Only once the analysis is complete, will the TEE release pre-agreed reports in encrypted form, which can be decrypted by the NSI and MNO-VAD. The reports contain aggregated data, which have already been subjected to SDCs inside the TEE before the release.

The only modification in these pre-existing processes (see Figure 3 above) concerns the periodic key generation process[29] – the currently used change-and-forget method will be replaced by an analogous mechanism offered by means of the Solution. This does not result in a change in the business process, *per se*, however, it introduces the possibility to add new functionalities on top of existing ones.

2.7. Applicable rules

Type of rules	Name of rules	Considerations
Legal norms	European Statistics Regulation	Provides the legal framework for the development, production and dissemination of European statistics. Applies to the NSI and Eurostat.
	ePrivacy Directive	Provides specific obligations concerning the protection of fundamental rights and freedoms *vis-à-vis* the processing of personal data regarding electronic communications. Applies to the MNO.
	General Data Protection Regulation	Provides the general rules to protect natural persons with regard to the processing of personal data and to ensure the free movement of personal data within the Union. Applies to all stakeholders.
	European Data Protection Regulation	Provides the rules relating to the protection of natural persons with regard to the processing of personal data by the Union institutions and bodies and rules relating to the free movement of personal data between them or to other recipients established in the Union. Applies to Eurostat.
Codes of conduct	European Statistics Code of Practice	Provides ethical guidance on how to perform official statistics – sets the standard for developing, producing and disseminating European statistics, along the lines of the institutional environment, statistical processes and statistical output. Applies to the NSI and Eurostat.

[29] See: Solution Analysis, Section 4.1.

	Quality Assurance Framework of the European Statistical System	Identifies possible methods, tools and good practices that can provide guidance and evidence for the implementation of the European Statistics Code of Practice. Applies to the NSI and Eurostat.
	Quality Declaration of the European Statistical System	Establishes a self-commitment to continuously develop, produce and disseminate high-quality European statistics and services in order to sustainably provide value to its users. Applies to the NSI and Eurostat.
	Other	Other, general quality management principles may apply. Applies to all stakeholders.
Standards	ISO/IEC 27001 Information security management	Provides requirements for keeping information assets secure. Applies to all stakeholders. It can be used as a benchmark for choosing the appropriate protection measures to complement the ones provided in the Solution.
	ISO/EIC 29101 ISO/IEC 29101:2018 Information technology — Security techniques — Privacy architecture framework	Incorporates the state-of-the-art of creating a privacy architecture for information and communication technology systems that process personally identifiable information. Applies to all stakeholders. The Solution has been developed based on the principles of this standard. It can be used as a benchmark for adjusting the privacy architecture of the Solution in case of further developments.
	Other	Other, sector-specific standards may apply, depending on the relevant statistical analysis use case.

2.8. Stakeholders and controllership

As a result of the controllership assessment in the Sample DPIA, it was concluded that either the MNO or the NSI can be designated as the controller in case of making the mobile location data anonymous by means of the Solution for further processing for the purpose of producing official statistics. This depends mainly on whether there is a legal obligation for the MNO to carry out the processing in question.

a) If there is such a legal obligation for the MNO, it presumably determines the MNO as a controller, possibly jointly with the NSI.

b) If there is no such legal obligation for the MNO, the processing must rely on a contractual arrangement between the MNO and NSI and thus presumes a consent

from Subscribers as a legal basis for processing. In such case, the NSI and the MNO can agree in the contract that the first acts as the controller and the latter as the processor.

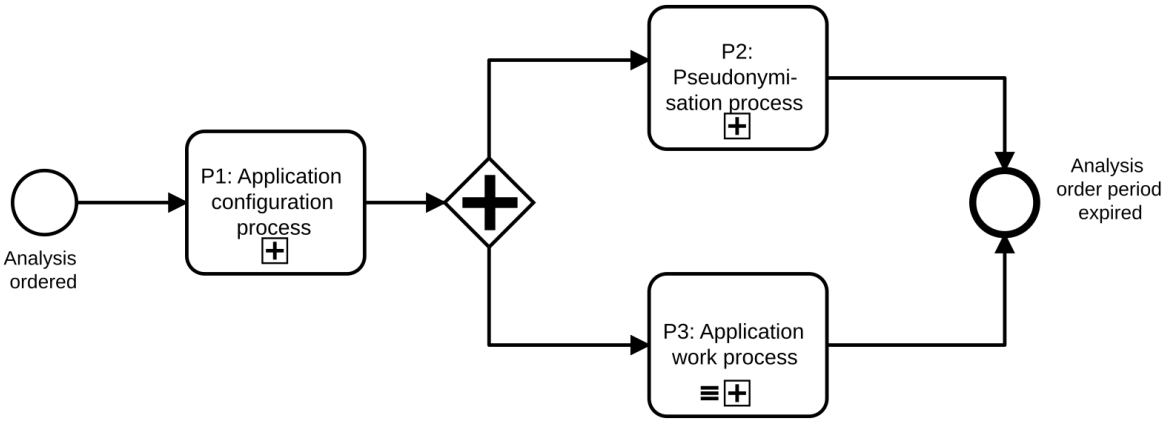
Stakeholder	Controller	Processor
NSI	The NSI acts in the general role of the initiator and direct beneficiary of further processing pseudonymous mobile location data by means of the Solution for producing official statistics. It decides both the means and the purposes of the processing, unless this decision-making task is divided between the NSI and the MNO according to a legal obligation in the applicable law.	-
MNO	The MNO has a power to decide (some of) the means and the purposes of the processing (jointly with the NSI), if there is a respective legal obligation in the applicable law.	MNO acts in the general role of a data source and service provider for the NSI. It acts under the authorisation of the NSI, unless a legal obligation in the applicable law grants the MNO with a power to decide (some of) the means and the purposes of the processing (jointly with the NSI). There is an internal separation of functions between two different departments within the MNO: - the Network Department of the MNO ("**MNO-ND**") – responsible for the pseudonymisation of mobile location data; - the Value-Added Services Department of the MNO ("**MNO-VAD**") – responsible for all other activities of the processing.

2.9. Process description

The Solution is envisioned to perform the following processes (see Figure 5 below), each of which will be summarised in the table below:

a) Application Configuration Process (P1),
b) Pseudonymisation Process (P2),
c) Application Work Process (P3).

Figure 5 – Solution general process



Process	Description	Data supporting assets
Application Configuration (P1)	The analysis of mobile location data will be carried out using an application, which is unique to each statistical analysis use case. In this process, the application specific to the Sample Use Case ("**Sample Use Case Application**") is developed, assessed for privacy risks, approved for implementation, set up, and attested.	Supporting assets in terms of mobile location data are not applicable because no mobile location data is being processed at this stage.
Pseudonymisation (P2)	In this process, the underlying mobile location data is being pseudonymised for the purposes of further processing with a key that changes every 24h. The pseudonymisation mechanism enables reversing the 24h pseudonyms inside the enclave of the Solution exclusively for the purposes of the Sample Use Case, so that the 24h pseudonyms can be linked to a specific mobile device denoted with a long-term pseudonym exclusively inside the enclave.	The processing of pseudonymous mobile location data is carried out for purposes other than the Sample Use Case. Therefore, the processing is subject to pre-existing pseudonymisation and value-added service provision processes at the MNO, subject to pre-existing data supporting assets. The pre-existing supporting assets are not within the scope of the Sample DPIA. However, the Solution introduces a new supporting asset for conducting the pre-existing pseudonymisation and value-added service provision processes – one of the functionalities of the Solution is used for generating the pseudonymisation keys (see P2 in Figure 6). For this reason, the

		relevant part of the Solution is further described on Figure 6.
Application Work (P3)	In this process, the NSI submits the statistical analysis request along with relevant inputs to the MNO and the MNO processes the pseudonymised mobile location data in two steps: 1) temporal summarisation of the pseudonymised mobile location data (in Module A) to prepare it for further processing by means of the Solution, 2) further temporal summarisation, aggregation and statistical analysis of the temporally summarised pseudonymised mobile location data by means of running the Sample Use Case Application on the Trusted Execution Environment (Modules B, C and D). None of the stakeholders will have access or visibility neither to the pseudonymisation keys nor to the long-term pseudonyms in the Trusted Execution Environment (except MNO-ND who has full access to raw mobile location data via other channels due to its central role in delivering telecommunications services to Subscribers).	See Figure 7 and Figure 8 below.

All the three processes described above include a remote attestation service process, but this is excluded from this analysis as it does not involve processing mobile location data.

Figure 6 – Pseudonymisation process (P2)

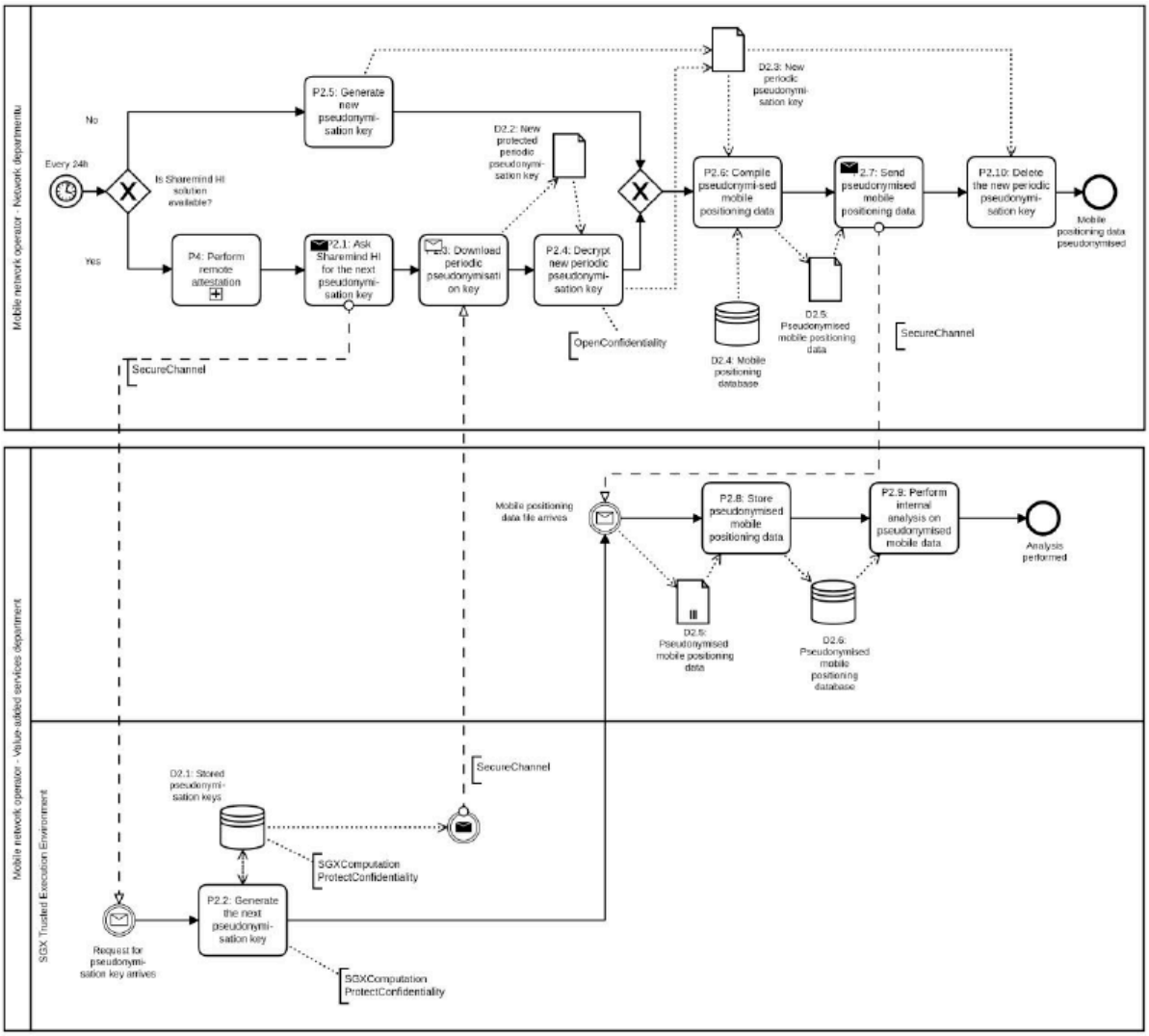


Figure 7 – Application work process (P3)

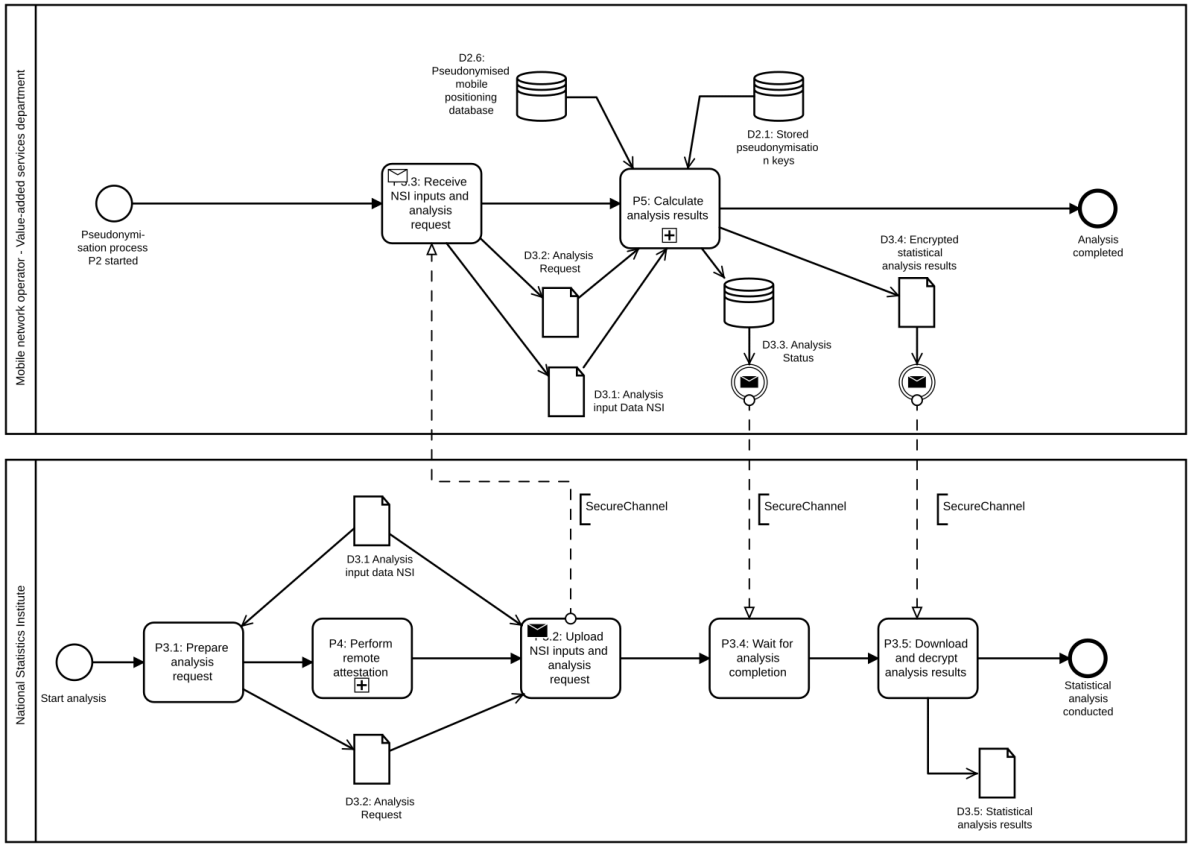
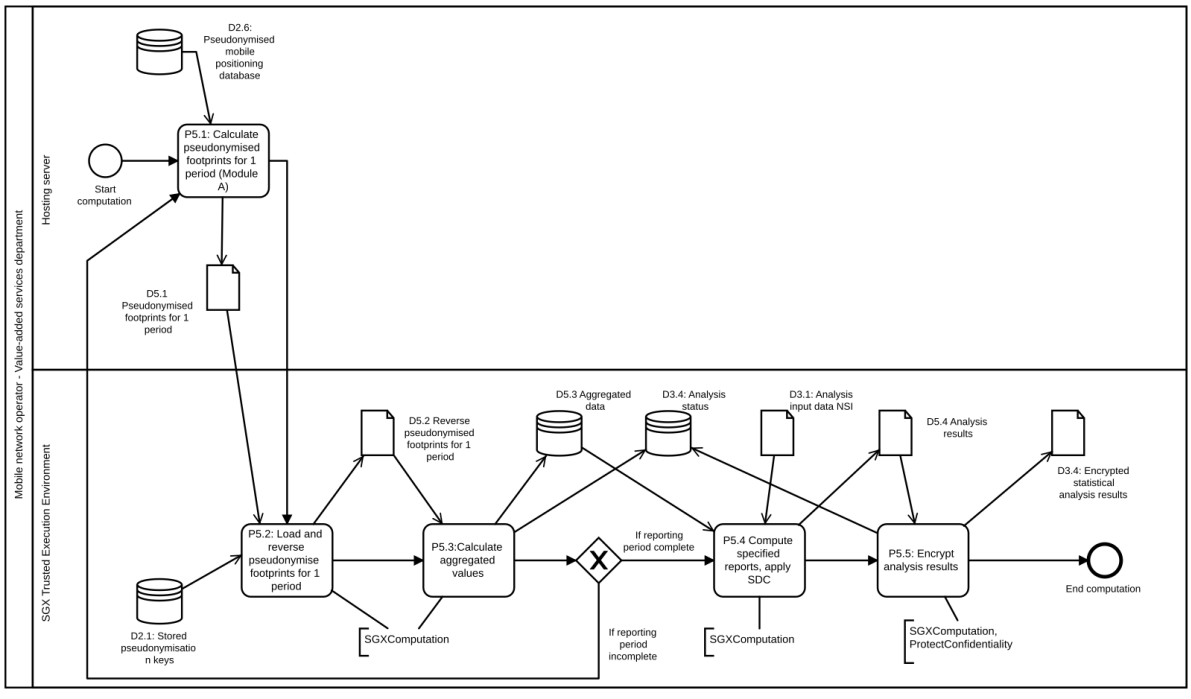


Figure 8 – Use case process (P5)



2.10. Data description

The Sample DPIA is focused on evaluating the impact of introducing mobile location data as a new source for computing the territorial dimension in official statistics by means of the Solution (see Section 2.5 above). Therefore, this document is dedicated to analysing the privacy-preserving processing of pseudonymous mobile location data by means of the Solution, as opposed to other types of personal data which may be provided as input or produced as output in the supporting processes of the Solution (e.g. user credentials and activity logs of the authorised users of the Solution). Such other types of personal data will be the subject of future legal analysis in the next iterations of the DPIA.

According to the CNIL guide on PIA methodology,[30] the study of the context of the processing has to define and describe in detail the personal data concerned. For this reason, this document does not cover non-personal data processed in the Sample Use Case by means of the Solution. Therefore, all non-personal data elements that were identified in other deliverables produced under the Agreement[31] have been excluded from the analysis contained herein, including, *inter alia*, the results of the collective aggregation step in Module D in the Trusted Execution Environment (see Figure 2 above), whereby the link between an individual mobile device and its location data is eliminated by deleting the underlying reverse pseudonymised mobile location data and the respective 24h pseudonymisation keys.

Although not part of the Solution, Module A also conducts personal data processing as part of the pre-processing step of the individual integration stage (see Figure 2 above). Since the results of the pre-processing in Module A are used as input for the Solution, it was considered necessary to include this step in the following analysis, in order to cover different processing operations at the macro level and better distinguish between processing with traditional technologies (Module A) and processing with privacy-enhancing technologies (the Solution).

The following tables in this section summarise the different types of mobile location data that are received by the relevant stakeholders (NSI and MNO) for further processing in the context of the Sample Use Case. Some of these personal data are initially collected by the MNO for purposes related to delivering telecommunications services and/or in support of network operation ("**Primary Processing**") and later re-used for the purposes of producing official statistics ("**Secondary Processing**"). The Sample DPIA is focused on the Secondary Processing, although Primary Processing is also covered to the extent it affects the Secondary Processing.

For the sake of better readability, the types of personal data processed have been divided into two tables:

1) the first table in Subsection 2.10.1 represents the types of personal data relevant in the Primary Processing phase,

2) the second table in Subsection 2.10.2 represents the types of personal data relevant in the Secondary Processing phase.

None of the tables include personal data which is generated in the Primary Processing phase, but not later re-used in the Secondary Processing phase. If the re-use of data

[30] CNIL. Privacy Impact Assessment (PIA). Methodology. February 2018 edition. – In the Internet: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf (24.08.2021).

[31] The excluded data elements are Analysis input data NSI (D3.1), Analysis request (D3.2), Aggregated data (D5.3), Encrypted statistical analysis results (D3.4) and Analysis results (D3.5, D5.4), as described in the document „ESTAT 2019.0232 Solution analysis".

elements in the Secondary Processing phase includes extracts of data elements generated in the Primary Processing, then only the entire data element as a whole is referred to, in order to avoid confusion regarding potential double processing.[32]

All data elements described in the tables refer to a numeric ID assigned to such data element in the "ESTAT 2019.0232 Solution analysis document" (in the format "Dn.n"). The same ID can be used to trace the relevant data element in the figures illustrating the process description in Section 2.9 above (see Figure 6, Figure 7 and Figure 8 above), as well as throughout the rest of the deliverables produced under the Agreement.

The non-identifiable state of the pseudonymous mobile location data within the Solution is marked with the colour green in the following tables. The importance of highlighting such data is explained in Section 2.10.3 below, which provides the legal reasoning for treating this data as anonymous.

[32] This concerns data elements D2.2 and D.2.3, which are extracts of data element D2.1 (encrypted and decrypted version, respectively).

2.10.1 Primary Processing

Data types	Recipients	Description of relevant data elements and processing activities
Raw mobile location data (D2.4)	MNO-ND	D2.4 – a database containing the following information: IMSI (International Mobile Subscriber Identity), timestamp and position. The MNO-ND creates the data element D2.4 as part of its usual business operations and maintains control over it.
24h pseudonymisation keys (D2.1, D2.2, D2.3)	Trusted Execution Environment (TEE)	D2.1 – a database composed of pseudonymisation keys, where each key is an independent random value valid for one period only (24h in the Sample Use Case). D2.2 – encrypted version of a pseudonymisation key extracted from the data element D2.1 for one period only (24h in the Sample Use Case). The TEE generates a new pseudonymisation key upon request from the MNO-ND as per the change-and-forget method and stores it in data element D2.1. The TEE encrypts the pseudonymisation key as data element D2.2 and sends it to the MNO-ND. The TEE protects the generated keys.
	MNO-ND	D2.2 – encrypted version of a pseudonymisation key extracted from the data element D2.1 for one period only (24h in the Sample Use Case). D2.3 – decrypted version of data element D2.2. The MNO-ND receives the data element D2.2 from the TEE, obtains data element D2.3 as a result of decrypting data element D2.2 and uses data element D2.3 in accordance with the change-and-forget method in its usual business operations.
Pseudonymised mobile location data (D2.5, 2.6)	MNO-ND	D2.5 – the pseudonymised version of an extract from the data element D2.4. The MNO-ND pseudonymises extracts of the data element D2.4, using the data element D2.3, and sends them to MNO-VAD as data element D2.5.
	MNO-VAD	D2.5 – the pseudonymised version of an extract from the data element D2.4.

		D2.6 – a database composed of data elements D2.5 over several periods. The MNO-VAD receives data elements D2.5 from the MNO-ND and stores them for further processing in data element D2.6 as part of its usual business operations.

2.10.2 Secondary Processing

Data types	Recipients	Storage duration
Pseudonymised mobile location data (D2.6)	MNO-VAD	D2.6 – a database composed of data elements D2.5 over several periods. The MNO-VAD sends data element D2.6 to the Module A.
Temporally summarised pseudonymised mobile location data (D5.1)	MNO-VAD (Module A)	D5.1 – a temporally summarised version of the data element 2.6. The Module A temporally summarises the data element D2.6, receives data element D5.1 as the output and sends it to the TEE.
	Trusted Execution Environment (TEE)	D5.1 – a temporally summarised version of the data element D2.6. The TEE receives the data element D5.1 from the Module A, stores and protects it for further processing by means of the TEE.
24h pseudonymisation keys (D2.1)	Trusted Execution Environment (TEE)	D2.1 – a database composed of pseudonymisation keys, where each key is an independent random value valid for one period only (24h in the Sample Use Case). The TEE uses the data element D2.1 to reverse pseudonymise data element D5.1. The TEE stores and protects the data element D2.1 for further processing by means of the TEE.
Temporally summarised reverse pseudonymised mobile location data (D5.2)	Trusted Execution Environment (TEE)	D5.2 – a reverse pseudonymised version of data element D5.1. The TEE reverse pseudonymises data element D5.1, receives data element D5.2 as the output and calculates aggregated values from it (the latter of the two can no longer be linked to individual Subscribers). The TEE stores and protects the data element D5.2 for further processing by means of the TEE.

2.10.3 Anonymous data

The Reference Scenario poses three cumulative requirements to the Solution ("**Core Requirements**"):

1) an MNO applies the change-and-forget method for every single pseudonymisation period (24h),
2) an NSI can produce official statistics from multiple records of the same mobile device over a period of time, which is longer than a single pseudonymisation period (1 year),
3) in order to produce official statistics, the NSI can add confidential calibration data as input to the statistical analysis process.

In order to fulfil the Core Requirements, the following principles were applied when designing the Solution ("**Core Design Principles**"):

a) nobody should be able to see, access or obtain pseudonymous mobile location data through the Solution. This includes, *inter alia*, no extracting of pseudonymisation keys which could be used for reverse pseudonymising the pseudonymous mobile location data at the MNO.

b) the Solution should be able to compute meaningful longitudinal statistical analysis based on the pseudonymous mobile location data,

c) no individual Subscribers should be identifiable from the output results of the Solution.

Essentially, the Core Requirements necessitate that the input data is pseudonymous (pre-processed mobile location data) and the output data is anonymous (official statistics). In legal terms, this means that the input data is personal data and output data is non-personal data. However, there is a strict legal regime under the EU law to protect mobile location data as personal data. In principle, only MNOs are allowed to process mobile location data for providing electronic communications services, as outlined in ePD Art 9. More specifically, ePD Art 9(1) first alternative prohibits MNO from sharing mobile location data with any third parties or otherwise further processing it without the data being "made anonymous". This means that re-use of mobile location data is generally prohibited. Any relevant national laws at Member State level must respect the limitation of ePD Art 9(1) because it is *lex specialis* in relation to GDPR. The exceptions from ePD Art 9(1) first alternative (ePD Art 9(1) second alternative, ePD Art 10(2), ePD Art 15) do not apply in case of the Solution.

According to the Sample DPIA, the most feasible legal route to processing pseudonymous mobile location data by means of the Solution for producing official statistics is to carry out the processing in a single step comprising two concurrent activities ("**2-in-1 approach**"):

1. **making the data anonymous** – the mobile location data is gradually made anonymous.
2. **statistical analysis** – the further processing of mobile location data for producing official statistics is carried out.

The data protection implications of the 2-in-1 approach depend on whether the further processing of pseudonymous mobile location data by means of the Solution qualifies as "made anonymous" under ePD Art 9(1):

a) **if yes**, then the "made anonymous" requirement of ePD Art 9(1) is fulfilled as soon as the pseudonymous mobile location data is encrypted for the Trusted Execution

Environment (TEE) within the Solution, considering that, with a sound design and implementation, it is not technically possible for any stakeholder or third party to access any intermediate data other than the final processing results, which are anonymised.

b) **if no**, then the Sample DPIA may need to be adjusted, so as to take better account of the (future) guidelines from the relevant data protection authorities and judgments of the relevant courts.

For the reasons above, the main focus of the Sample DPIA was on the "making the data anonymous" side of the 2-in-1 approach. The Sample DPIA concluded that the further processing of pseudonymous mobile location data by means of the trusted hardware component within the Solution (the enclaves), along with the complementing technical, legal and organisational protection measures applied in the Solution, qualifies as "made anonymous" under ePD Art 9(1), if the Core Design Principles are maintained. This is achieved thanks to a new state-of-the-art introduced by means of the secure computation model used in the Solution, which involves a combination of measures assuring input privacy, output privacy, as well as privacy during processing.[33]

"Made anonymous" is often understood as anonymisation. There are two well-known techniques to anonymisation – noise addition at the input level (anonymised database) and at the output level (anonymised query result). It is less known that "making anonymous" can also be achieved by means of anonymous processing. In that case, there is no suppression or noise needed – the underlying data remains intact and the "making anonymous" is conducted at the processing level, not only on the data.[34] The Solution implements the latter technique, i.e. anonymous processing by means of privacy-enhancing technologies, in combination with traditional anonymisation techniques and complemented by other technical, legal and organisational protection measures, as will be detailed below.

The EDPB has recently referred to an article by Y.-A. de Montjoye, et al.[35], which introduces four models for the privacy-conscientious use of mobile phone data. The authors of the paper demonstrate how the four models overcome the limits of traditional data anonymisation methods. According to the article, all four models have been "designed to fall under the legal umbrella of anonymous use of the data" [36]. Although none of them is considered to be a silver bullet according to the authors, each is believed to provide a reasonable balance between utility and privacy.[37]

The Sample DPIA concluded that the Solution adds to the existing state-of-the-art for the privacy-conscientious use of mobile phone data. It involves a combination of the following technical, legal and organisational protection measures[38]:

---

[33] For a detailed analysis, please refer to Sections 8.2.4.i) and ii).

[34] D. Bogdanov, T. Siil. Anonymisation 2.0: Sharemind as a Tool for De-Identifying Personal Data - Part 2: Sharemind and anonymization. – Internet: https://sharemind.cyber.ee/anonymisation-2_0-part-2-sharemind/ (25.08.2021).

[35] European Data Protection Board. Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak. Adopted on 21 April 2020, p 6. – Internet: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-042020-use-location-data-and-contact-tracing_en (02.05.2021). – Referring to: Y.-A. de Montjoye, S. Gambs, V. Blondel, G. Canright, N. de Cordes, S. Deletaille, K. Engø-Monsen, M. Garcia-Herranz, J. Kendall, C. Kerry, G. Krings, E. Letouzé, M. Luengo-Oroz, N. Oliver, L. Rocher, A. Rutherford, Z. Smoreda, S. Steele, E. Wetter, Alex "Sandy" Pentland L. Bengtsson. Comment: On the privacyconscientious use of mobile phone data. Scientific Data 5 (December 2018): 180286. – Internet: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6289108/pdf/sdata2018286.pdf (02.05.2021).

[36] *Ibid.*, Y.-A. de Montjoye, et al. Comment: On the privacyconscientious use of mobile phone data. Scientific Data, 2018, p 2.

[37] *Ibid*.

[38] For a detailed analysis, please refer to Sections 8.2.4.i) and ii) of the Scoping Report.

1) **TEE (the Solution enclaves)** – the Solution prevents visibility of data during computations even to privileged users, thanks to the Trusted Execution Environment (TEE) component (Solution enclaves) involved in its architecture, which provides a protected memory area with confidentiality and integrity guarantees. These guarantees hold even if privileged malware is present in the system, meaning that each enclave is protected even from the operating system that is running the enclave. This way, the pseudonymised mobile location data is added an additional layer of protection which effectively secures it from being manipulated during computations, even by relevant stakeholders directly involved in the Sample Use Case. No mobile location data is disclosed to the NSI or any third parties.
2) **no transfer of personal data** – the Solution is designed in a way that the mobile location data never leaves the MNO. The NSI can process the pseudonymised mobile location data in a secure environment at the MNO, separately from the MNO's regular business operations related to the provision of telecommunications services and value added services. The MNO keeps separately the raw (at the MNO-ND) and pseudonymised mobile location data (at the MNO-VAD), as well as the pseudonymisation keys which enable identification of individual mobile devices of Subscribers (at the MNO-ND). Moreover, the identities corresponding to the pseudonyms are kept separately from the Solution – they are stored at the MNO-ND, whereas the Solution along with the protected pseudonymization keys is hosted by the MNO-VAD. The NSI does not have access to the raw or pseudonymised mobile location data nor to the pseudonymisation keys – it can only prepare the Solution for deployment, activate it (enable the pseudonymisation process), initiate the statistical analysis process and receive the final reports.
3) **pre-approved computations** – the analysis code, including any automated controls (e.g. SDCs) are pre-approved by several independent stakeholders (NSI, MNO, potentially the local DPA or third-party auditor) and validated by an independent attestation service provider. The approval and validation procedure are technically enforced by the Solution set-up and remote attestation activities.
4) **change-and-forget pseudonymization –** raw mobile location data is pseudonymised by the MNO-ND at its premises using the periodic pseudonymisation key generated in the Solution enclave and shared with MNO-ND in encrypted form. The pseudonymisation key is changed after each 24h period. Only the MNO-ND has the means to decrypt it as per the change-and-forget method and is required to delete the decrypted key as soon as the pseudonymisation process is complete.
5) **temporal summarisation** – adding technical difficulties to attempts at re-identifying individuals and limiting the amount of information that could be uncovered if the data were to be re-identified (in Modules A, B and C).
6) **aggregation** – adding technical difficulties to attempts at re-identifying individuals and limiting the amount of information that could be uncovered if the data were to be re-identified (in Module D).
7) **governance of rights and roles** – the applied rights and roles concept limits the necessary rights and accesses in accordance with the least privilege principle.
8) **limited storage periods** – all data elements, whether in encrypted or decrypted form, are stored only until they are necessary to finalise the relevant computations. Only the 24h pseudonymisation keys (D2.1) are kept for the whole period of analysis (in the Sample Use Case deleted after 1 year) and have to be deleted manually thereafter.[39]

[39] For more details on storage limitations, see Section 3.1.1.5 below.

9) **encryption** – inputs and outputs are stored in the Trusted Execution Environment (TEE) in encrypted form, so that the data can be decrypted only inside a Solution enclave or by authorised clients.

10) **pre-approved outputs** – the output results are pre-approved by several independent stakeholders (NSI, MNO, potentially the local DPA or third-party auditor).

11) **SDC** – additional output privacy controls to assure confidentiality in accordance with applicable statistics laws.

12) **auditability** – auditor(s) (e.g. data protection authorities or internal audit divisions of MNO and NSI) can be involved in the Solution development and setup to verify the correctness of the Sample Use Case Application both *ex-ante* and *ex-post*. An auditor has access to the Sample Use Case Application source code and verifies *ex-ante* the fulfilment of privacy requirements (including e.g. the non-personal nature of the final output). The auditor has also access to the system audit logs to verify *ex-post* that the data processing with the Solution was in conformity with applicable law and agreements between relevant stakeholders and that the Solution had not been tampered with (identification of potential attacks against the Solution).

These cumulative techniques enforce the data minimization principle at its maximum level, resulting in the pseudonymized mobile location data being practically anonymous throughout the processing in the Solution. As a result, the content of the mobile location data is effectively hidden from all stakeholders and information about a particular person may not be learned by any stakeholder or third party in the Solution. The specific combination of the applied protection measures differs from one stage of processing to another, as illustrated in the Table 1 below.

Table 1. Protection measures applied in the Sample Use Case (Meta Level)

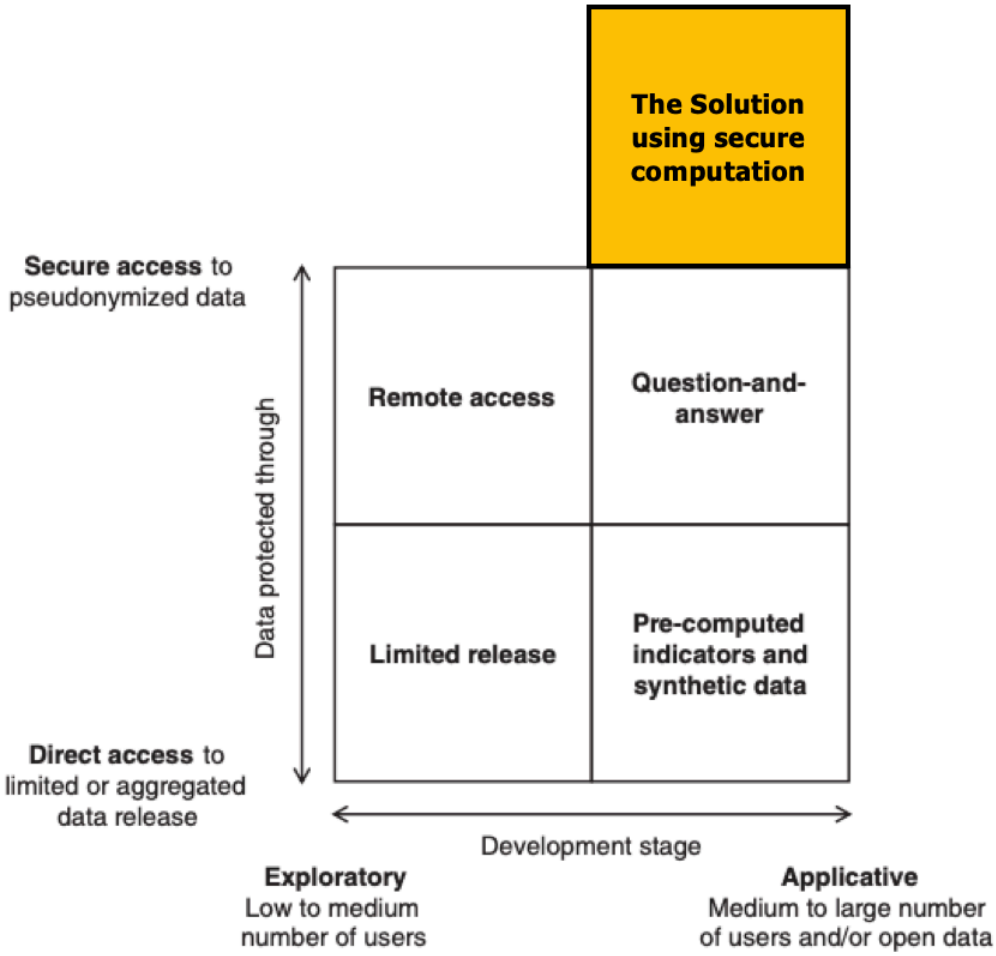
Applied protection measures	Module A*	Module B	Module C	Module D
1) TEE (Solution enclaves)	N/A	+	+	+
2) no transfer of personal data	TBS	+	+	+
3) pre-approved computations	N/A	+	+	+
4) change-and-forget pseudonymization	+	+	N/A	N/A
5) temporal summarisation	+	+	+	N/A
6) aggregation	N/A	N/A	N/A	+
7) governance of rights and roles	TBS	+	+	+
8) limited storage periods	TBS	+	+	+
9) encryption	TBS	+	+	+
10) pre-approved outputs	TBS	N/A	N/A	+
11) SDC	N/A	N/A	N/A	+
12) auditability	N/A	+	+	+

* Module A is not part of the Solution and is thus highlighted.

Legend:
"+" – measure is applied
"N/A" – non-applicable
"TBS" – to be specified (not covered within the scope of the Sample DPIA)

Relying on the above, the Solution is proposed in the Sample DPIA as a fifth model in addition to the four models introduced by Y.-A. de Montjoye, et al.[40] earlier (see Figure 9 below).

Figure 9 – The Solution using the secure computation model



Based on the above, it remains to be seen if the relevant data protection authorities and courts accept the novel interpretations of the concept "made anonymous" as proposed in the Sample DPIA. A typical statistical analysis process may require collection and analysis of individualised personal information to produce the relevant statistics. Usually, traditional anonymisation techniques or other protection measures are then applied to assure statistical confidentiality at the output level. By implementing the Solution, is is possible to introduce appropriate protection measures both at the input and output level as well as during processing, in order to assure privacy at its maximum:

1) input privacy – concerns Modules A (not part of the Solution), B and C;
2) privacy during processing – concerns Modules B, C and D;
3) output privacy – concerns Module D.

[40] *Op. cit.*, Y.-A. de Montjoye, et al. Comment: On the privacyconscientious use of mobile phone data. Scientific Data, 2018.

In this context, it is important to understand that the notion of anonymity has a different scope in data protection law and in statistics law. From a data protection view, the notion of anonymity would cover data that are no longer identifiable (previously DPD Rec 26, now GDPR Rec 26).[41] From a statistical point of view, anonymous data are data for which no direct identification is possible, implying that indirect identification of data would still qualify these data as anonymous.[42] Therefore, the threshold for treating personal data as anonymous is different in each case – it is lower under statistics law (includes indirectly identifiable data) and higher under data protection law (does not include identifiable data). The EDPS has emphasized that "in order to avoid possible misunderstandings when using these notions, the context and legal framework in which these notions are being used should be always clearly and precisely defined."[43] It was concluded in the Sample DPIA that the further processing of pseudonymous mobile location data by means of the Solution fulfils not only the requirements of the lower threshold for the notion of anonymity under statistics law but also the higher threshold for the same in data protection law under applicable law and regulations as they stand at the moment of completing the Sample DPIA.[44]

If the relevant data protection authorities and courts agree that the further processing of pseudonymous mobile location data by means of the Solution for producing official statistics is considered as "making anonymous" under ePD Art 9(1), then the Solution functions as a condition for the further processing[45]. This means that the condition of "making anonymous" under ePD Art 9(1) holds if and only if all Core Design Principles are maintained. This can be further ensured by applying additional organisational and legal and technical protection measures beyond the Solution.

"Making anonymous" in terms of ePD Art 9(1) first alternative is a type of personal data processing. Therefore, it needs to fulfil all the requirements of data protection regulations just as any other type of personal data processing, including compatibility with the purposes for which the data was collected (see the conclusions of the compatibility assessment below), a defined controller and processor(s) (see the conclusions of the controllership assessment below) and a legal basis for processing (see the conclusions of the lawfulness assessment below).

Consequently, the following analysis and the accompanying tables clarify the personal data processing activities which are carried out in order to transform the pseudonymised mobile location data into anonymous data by means of the Solution. As soon as the pseudonymous mobile location data enters the Solution, it is effectively hidden from both the NSI and MNO as well as any third parties and thus becomes non-identifiable.

[41] European Data Protection Supervisor. Opinion of the European Data Protection Supervisor on the proposal for a Regulation of the European Parliament and of the Council on Community statistics on public health and health and safety at work (COM(2007) 46 final), 2007/C 295/01, Brussels, 5 September 2007, Sec 19 – Internet: https://edps.europa.eu/data-protection/our-work/publications/opinions/community-statistics-health-data_en (25.08.2021).
[42] *Ibid*.
[43] *Ibid.*
[44] See Section 8.2. of the Scoping Report.
[45] See Section 8.2.5. of the Scoping Report.

3. Fundamental Principles

3.1. Data protection principles

3.1.1 Explanation

3.1.1.1 Purpose limitation

Purposes	Legitimacy
Producing official statistics (developing alternative means for computing the territorial dimension in official statistics, e.g. related to population distribution)	As a result of the compatibility assessment in the Sample DPIA, it was concluded that making the mobile location data anonymous by means of the Solution for the purpose of producing official statistics is compatible further use in terms of GDPR Art 6(4) and Art 5(1)(b), because the Solution qualifies as appropriate safeguards in terms of GDPR Art 89(1).

3.1.1.2 Lawfulness

As a result of the lawfulness assessment in the Sample DPIA, it was concluded that making the mobile location data anonymous by means of the Solution for further processing for the purpose of producing official statistics can, in principle, be based on consent (GDPR Art 6(1)(a)), legal obligation (GDPR Art 6(1)(c)), public interest/official authority (GDPR Art 6(1)(e)) and legitimate interest of the MNO (GDPR Art 6(1)(f)).

A different legal basis may be applied, depending on whether further processing pseudonymous mobile location data by means of the Solution for producing official statistics is carried out in the proof-of-concept, pilot project or production stage.

Note that each Member State currently has the possibility to choose a suitable legal ground under GDPR art 6 and thus the type of legal basis for “making anonymous” can change from country to country (e.g. consent, legal obligation, public interest/official authority, legitimate interest).

Lawfulness criteria	Applicable	Justification
Consent	+	As long as the provision of mobile location data is voluntary, it can rely on consent, similarly to voluntary responses to survey questions.
Contract	–	NSI will not be providing any direct value or services to the Subscribers, therefore there will not be a contract between them.

Legal obligation	?	No specific legal obligation under EU law or national law of the relevant Member State was established for the purposes of the Sample Use Case, as the Sample Use Case was not the object of the Sample DPIA (the Member State, the domain of official statistics and the broader production process embedding the Sample Use Case is unspecified). This does not exclude, however, that a suitable legal obligation may already exist in some EU Member States or may be in the process of being established in the future. It will be a matter of legal analysis for the next iterations of the DPIA to determine the relevant legal basis, if applicable.
Vital interests	–	Further processing pseudonymous mobile location data by means of the Solution for producing official statistics is not in the vital interests of the Subscribers.
Public interest task / official authority	?	The mere public interest task/official authority of the NSI alone may not be sufficient to oblige the MNO to provide mobile location data to NSI for analysis. It remains a matter of legal analysis for the next iterations of the DPIA to determine the relevant legal basis, if applicable.
Legitimate interest	?	It is questionable if an MNO can rely on the legitimate interest ground for the purposes of the Sample Use Case, because the MNO's legitimate interest to process mobile location data is not connected to the NSIs purpose of producing official statistics. NSI as a public authority cannot rely on the legitimate interest ground in the performance of its tasks (GDPR Art 6(1)(f), GDPR Art 6(1) last sentence). It remains a matter of legal analysis for the next iterations of the DPIA to determine the relevant legal basis, if applicable.

3.1.1.3 Data minimisation

a) Primary Processing

Data types	Data categories	Justification of the need and relevance of the data	Minimization controls
Raw mobile location data (D2.4)	D2.4 at the MNO-ND: Common personal data (perceived as sensitive) until deleted at the MNO-ND's discretion.	D2.4 at the MNO-ND: The justification is provided in: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).	D2.4 at the MNO-ND: Defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).
24h pseudonymisation keys (D2.1, D2.2, D2.3)	D2.1 at the Trusted Execution Environment (TEE): becomes anonymous data as soon as the 24h pseudonymisation keys are deleted by the MNO-ND in accordance with the change-and-forget method.	D2.1 at the Trusted Execution Environment (TEE): the 24 pseudonymisation keys are important for the MNO to fulfil its obligations to protect the mobile location data.	D2.1 At the Trusted Execution Environment (TEE): 1) processing within the TEE (the keys are generated within a Solution enclave), 2) no transfer of personal data (data stays within the MNO), 3) pre-approved computations (the key generation process (P2) is technically enforced by means of the Solution), 4) encryption (keys are encrypted before release to the MNO-ND), 5) storage limitations (see Section 3.1.1.5a) below), 6) governance of rights and roles

			(released only to the MNO-ND), Other controls defined by: 1) contractual arrangements between the NSI and the MNO (to forbid identification of Subscribers), 2) the applicable electronic communications, statistics and data protection laws and regulations.
	D2.2 and D2.3 at the MNO-ND: Common personal data (perceived as sensitive) until deleted by the MNO-ND in accordance with the change-and-forget method.	D2.2 and D2.3 at the MNO-ND: the justification is provided in: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).	D2.2 and D2.3 at the MNO-ND: 1) Storage limitations (see Section 3.1.1.5a) below). Other controls defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).
Pseudonymised mobile location data (D2.5, D2.6)	D2.5 at the MNO-ND: Common personal data (perceived as sensitive) until deleted at the MNO-ND's discretion.	D2.5 at the MNO-ND: Defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).	D2.5 at the MNO-ND: 1) Change-and-forget pseudonymisation, Other controls defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).
	D2.6 at the MNO-VAD: Common personal data (perceived as	D2.6 at the MNO-VAD: Defined by:	D2.6 at the MNO-VAD: 1) Change-and-forget pseudonymisation,

	sensitive) until deleted at the MNO-VAD's discretion.	1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-VAD).	Other controls defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-VAD).

b) Secondary Processing

Details about the data processed	Data categories	Justification of the need and relevance of the data	Minimization controls
Pseudonymised mobile location data (D2.6)	D2.6 at the MNO-VAD: Common personal data (perceived as sensitive) until deleted at the MNO-VAD's discretion.	D2.6 at the MNO-VAD: The pseudonymised mobile location data is essential for developing alternative means for computing the territorial dimension in official statistics, e.g. related to population distribution. The specific justifications for each statistical analysis use case implemented for real-world scenarios in the future may be further defined by applicable electronic communications, statistics and data protection laws and regulations.	D2.6 at the MNO-VAD: 1) Change-and-forget pseudonymisation, Other controls defined by: 1) contractual arrangements between the NSI and the MNO (to forbid personal data processing beyond the agreed statistical analysis use cases), 2) the applicable electronic communications, statistics and data protection laws and regulations.
Temporally summarised pseudonymous mobile location data (D.5.1)	D5.1 at the MNO-VAD: Common personal data (perceived as sensitive), until the MNO-VAD has deleted the data.	D5.1 at the MNO-VAD: This data has been pre-processed to the level where it can still be useful for producing adequate statistics without directly identifying any of the Subscribers.	D5.1 at the MNO-VAD: 1) Change-and-forget pseudonymisation, 2) temporal summarisation (in Module A), 3) storage limitations (see Section 3.1.1.5b) below).

			Other controls defined by: 1) contractual arrangements between the NSI and the MNO (to forbid personal data processing beyond the agreed statistical analysis use cases), 2) the applicable electronic communications, statistics and data protection laws and regulations.
24h pseudonymisation keys (D2.1)	D2.1 at the Trusted Execution Environment (TEE): anonymous data, if the MNO-ND has deleted the copies of the same keys in its premises outside the Solution in accordance with the change-and-forget method.	D2.1 at the Trusted Execution Environment (TEE): the 24 pseudonymisation keys are essential for enabling meaningful longitudinal statistical analysis.	D2.1 at the Trusted Execution Environment (TEE): 1) processing within the TEE, 2) no transfer of personal data (data stays within the MNO), 3) pre-approved computations (the reverse pseudonymisation process (P5.2 on Figure 8) is technically enforced by means of the Solution), 4) encryption (keys are decrypted only within a Solution enclave performing the reverse pseudonymisation process (P5.2 on Figure 8)), 5) storage limitations (see Section 3.1.1.5b) below), 6) governance of rights and roles (the keys are released only to another Solution enclave within the TEE to conduct reverse

			pseudonymisation in P5.2 on Figure 8), Other controls defined by: 1) contractual arrangements between the NSI and the MNO (to forbid identification of Subscribers), 2) the applicable electronic communications, statistics and data protection laws and regulations.
Temporally summarised reverse pseudonymised mobile location data (D5.2)	D5.2 at the Trusted Execution Environment (TEE): anonymous data	D5.2 at the Trusted Execution Environment (TEE): essential for conducting meaningful longitudinal statistical analysis.	D5.2 at the Trusted Execution Environment (TEE): 1) processing within the TEE, 2) temporal summarisation (in Modules B and C), 3) aggregation (in Module D), 4) SDCs (applied in Module D), 5) no transfer of personal data (data stays within the MNO), 6) pre-approved computations (the statistical analysis use case process (P5.3-P5.5 on Figure 8) is technically enforced by means of the Solution), 7) encryption (the analysis results are encrypted before release to the NSI), 8) storage limitations (see Section 3.1.1.5b) below), 9) governance of rights and roles (processing only in a Solution enclave, no

			access rights to third parties). Other controls defined by: 1) contractual arrangements between the NSI and the MNO to forbid identification of Subscribers and personal data processing beyond the agreed statistical analysis use cases), 2) the applicable electronic communications, statistics and data protection laws and regulations.

3.1.1.4 Accuracy (data quality)

Data quality controls	Justification
MNO internal operations	Since the processing concerns secondary use of mobile location data, the NSI relies on the MNO regarding good practices of ensuring the accuracy of the input data and keeping it up to date, i.e. the Subscribers may rectify and update their mobile location data, where feasible by means of the MNO systems.

3.1.1.5 Storage limitation

a) Primary Processing

Data types	Storage duration	Justification of the storage duration	Erasure mechanism at the end of the storage duration
Raw mobile location data (D2.4)	D2.4 at the MNO-ND: deleted by the MNO-ND at their discretion. Defined by: 1) applicable electronic	D2.4 at the MNO-ND: The justification is provided in: 1) applicable electronic communications and data	D2.4 at the MNO-ND: Defined by: 1) applicable electronic communications and data

	communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).	protection laws and regulations. 2) MNO internal business processes (MNO-ND).	protection laws and regulations. 2) MNO internal business processes (MNO-ND).
24h pseudonymisation keys (D2.1, D2.2, D2.3)	D2.1 at the Trusted Execution Environment (TEE): stored until the period of analysis expires and all data of the Trusted Execution Environment is deleted manually (in the Sample Use Case after 1 year). D2.2 is an encrypted copy of one element from D2.1. D2.2 is sent to the MNO-ND, who decrypts it and obtains D2.3. Defined by: 1) applicable statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc), 4) updates to the Solution.	D2.1 at the Trusted Execution Environment (TEE): The keys are needed as long as the period for which the statistical analysis is conducted (1 year) has come to an end, otherwise it would not be possible to reverse the 24h pseudonyms for the purposes of linking them to obtain the long-term pseudonyms, which are the basis for carrying out longitudinal statistical analysis.	D2.1 at the Trusted Execution Environment (TEE): Automatically (defined by the pre-approved computations in the Solution). Defined by: 1) applicable statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc), 4) updates to the Solution.
	D2.2 and D2.3 at the MNO-ND: deleted by the MNO-ND right after use in accordance with the change-and-forget method. Defined by:	D2.2 and D2.3 at the MNO-ND: The justification is provided in: 1) applicable electronic communications and data	D2.2 and D2.3 at the MNO-ND: Defined by: 1) applicable electronic communications and data

	1) applicable electronic communications and data protection laws and regulations, 2) MNO internal business processes (MNO-ND).	protection laws and regulations, 2) MNO internal business processes (MNO-ND).	protection laws and regulations, 2) MNO internal business processes (MNO-ND).
Pseudonymised mobile location data (D2.5, D2.6)	D2.5 at the MNO-ND: deleted by the MNO-ND at their discretion. Defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-ND).	D2.5 at the MNO-ND: The justification is provided in: 1) applicable electronic communications and data protection laws and regulations, 2) MNO internal business processes (MNO-ND).	D2.5 at the MNO-ND: Defined by: 1) applicable electronic communications and data protection laws and regulations, 2) MNO internal business processes (MNO-ND).
	D2.6 at the MNO-VAD: deleted by the MNO-VAD at their discretion. Defined by: 1) applicable electronic communications and data protection laws and regulations. 2) MNO internal business processes (MNO-VAD).	D2.6 at the MNO-VAD: The justification is provided in: 1) applicable electronic communications and data protection laws and regulations, 2) MNO internal business processes (MNO-VAD).	D2.6 at the MNO-VAD: Defined by: 1) applicable electronic communications and data protection laws and regulations, 2) MNO internal business processes (MNO-VAD).

b) Secondary Processing

Data types	Storage duration	Justification of the storage duration	Erasure mechanism at the end of the storage duration
Pseudonymised mobile location data (D2.6)	D2.6 at the MNO-VAD: deleted by the	D2.6 at the MNO-VAD: Storing the	D2.6 at the MNO-VAD:

	MNO-VAD right after relevant computations on it (P5.1 on Figure 8) have been completed in the Module A. Defined by: 1) applicable electronic communications, statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).	pseudonymous mobile location data outside the Solution is justified only until there are ongoing computations in the Module A. The specific justifications for storing data in each statistical analysis use case implemented for real-world scenarios in the future may be further defined by applicable electronic communications, statistics and data protection laws and regulations.	Defined by: 1) applicable electronic communications, statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).
Temporally summarised pseudonymous mobile location data (D5.1)	D5.1 at the MNO-VAD: Deleted by the MNO-VAD right after importing it from the Module A to the Solution (P5.2 on Figure 8). Defined by: 1) applicable statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).	D5.1 at the MNO-VAD: Storing the temporally summarised pseudonymous mobile location data outside the Solution is justified only until they are imported to the Solution. The specific justifications for storing data in each statistical analysis use case implemented for real-world scenarios in the future may be further defined by applicable electronic communications, statistics and data protection laws and regulations.	D5.1 at the MNO-VAD: Defined by: 1) applicable electronic communications, statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).
	D5.1 at the Trusted Execution Environment (TEE): deleted right after it has been reverse pseudonymised in the Solution's	D5.1 at the Trusted Execution Environment (TEE): The least period required in order to carry out the	D5.1 at the Trusted Execution Environment (TEE): Automatically (defined by the pre-approved computations in the Solution).

	enclave and relevant computations on it (P5.3 on Figure 8) have been completed. Defined by: 1) applicable statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc), 4) updates to the Solution.	longitudinal statistical analysis.	Defined by: 1) applicable electronic communications, statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).
24h pseudonymisation keys (D2.1)	D2.1 at the Trusted Execution Environment (TEE): deleted manually after the period of analysis expires and all data of the Solution enclave is deleted (in the Sample Use Case after 1 year). Defined by: 1) applicable statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc), 4) updates to the Solution.	D2.1 at the Trusted Execution Environment (TEE): The least period required in order to carry out the longitudinal statistical analysis.	D2.1 at the Trusted Execution Environment (TEE): Automatically (defined by the pre-approved computations in the Solution). Defined by: 1) applicable electronic communications, statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).
Reverse pseudonymised temporally summarised	D5.2 at the Trusted Execution Environment (TEE):	D5.2 the Trusted Execution Environment (TEE):	D5.2 the Trusted Execution Environment (TEE):

mobile location data (D5.2)	deleted right after the relevant computations on it (P5.3 on Figure 8) have been completed. Defined by: 1) applicable statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc), 4) updates to the Solution.	The least period required in order to carry out the longitudinal statistical analysis.	Automatically (defined by the pre-approved computations in the Solution). Defined by: 1) applicable electronic communications, statistics and data protection laws and regulations, 2) contractual arrangements between the NSI and the MNO, 3) settings of the MNO systems (e.g. back-up frequency etc).

3.1.2 Assessment

The above choices, if deployed in full, are considered sufficient to lower identified technical and privacy risks to a level where residual risks are acceptable.

The only exception concerns the processing of pseudonymised mobile location data in the Module A (not part of the Solution). As thoroughly analysed in the Sample DPIA, mobile location data has to be "made anonymous" before any further processing according to ePD Art 9(1) first alternative. Any pre-processing of the pseudonymous mobile location data before importing it to the Solution's enclave is processing personal data, not anonymous data. Therefore, according to current ePD norms, the pre-processing of pseudoymised mobile location data in the Module A precludes the possibility to rely on any legal basis established under GDPR Art 6 for making the mobile location data anonymous by means of the Solution. In order to overcome this risk under the current ePD norms, there are two options:

1) include the Module A in the Solution's enclave, thus ensuring that the pseudoymised mobile location data is effectively "made anonymous" as per ePD Art 9(1) first alternative, as proposed in the Sample DPIA. In this case, it needs to be further analysed if there are any existing legal bases under EU law and/or national law of the relevant Member State for making the pseudonymised mobile location data anonymous in accordance with the currently applicable ePD Art 9(1) first alternative. Due to the nature of the ePD as a directive, the ePD Art 9(1) first alternative needs to be implemented in Member States' laws. However, it is possible that it can also have direct effect in Member States' laws – it needs to be further analysed, if Member States can rely directly on ePD Art 9(1) first alternative or whether it needs to be harmonised first.

2) create a new legal basis under EU law and/or national law of the relevant Member State, which would allow pre-processing of (pseudonymous) mobile location data in pseudonymous form (e.g. by means of the Module A) before it is further processed for producing official statistics (e.g. by means of the Solution). Considering the high level of discretion Member States have under GDPR for introducing special rules for processing personal data for statistical purposes, it is worth considering whether the new legal basis may be derived from directly applicable EU law. For consistency purposes, it may be advisable to set up the new legal basis under EU law, so as to facilitate uniform interpretations and practices across the EU. This is a matter for further legal analysis.

Note that a revision of ePD towards a new ePrivacy Regulation ("**ePR**") has been underway for years.[46] If the ePD is amended or replaced with the ePR in the future, then the "make anonymous" requirement under ePD Art 9(1) first alternative may be eliminated. In such case, the ePR may create new legal routes for NSIs to obtain mobile location data from MNOs, pre-process it and, ultimately, produce official statistics.

Controls guaranteeing the proportionality and necessity of the processing	Acceptable	Corrective controls
Purposes: specified, explicit and legitimate	+	–
Basis: lawfulness of processing, prohibition of misuse	–	1) include the Module A in the Solution, thus ensuring that the pseudoymised mobile location data is effectively "made anonymous" as per ePD Art 9(1) first alternative, or 2) create a new legal basis under EU law and/or national law of the relevant Member State, which would allow pre-processing of (pseudonymous) mobile location data in pseudonymous form (e.g. by means of the Module A) before it is further processed for producing official statistics (e.g. by means of the Solution).
Data minimization: adequate, relevant and limited	+	–
Data quality: accurate and kept up-to-date	+	–
Storage durations: limited	+	–

[46] European Commission. Shaping Europe's digital future. Proposal for an ePrivacy Regulation. – Internet: https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation (13.05.2021).

3.2. Data subject's rights

3.2.1 Explanation

Considering that the "making anonymous" of pseudonymised mobile location data by means of the Solution is further processing of such mobile location data, it is presumed that all data subject's rights apply until the moment when anonymisation is complete. That moment arrives when the pseudonymous mobile location data has been temporally summarised and aggregated in within the Trusted Execution Environment (P5.3 on Figure 8).

In this light, most of the controls for data subject's rights would have to be applied by the MNO as part of the Primary Processing in their pre-existing business processes. In the Secondary Processing, much of the pseudonymised mobile location data processed in the Privacy Processing is re-used. Therefore, the NSI and MNO need to agree to what extent the Primary Processing controls must be implemented in case of Secondary Processing, which controls must be added or revised in case of Secondary Processing and in which cases an exemption provided in the GDPR may be applied.

Because the legal basis of "making anonymous" of pseudonymised mobile location data by means of the Solution for producing official statistics remains to be specified in the future (see Section 3.1.1.2 above), it is difficult to provide an overview of all applicable controls at this stage. In order to facilitate understanding of the potential controls to be applied, consent of a Subscriber will be used as an example of how a legal basis for personal data processing can determine the controls.

3.2.1.1 Transparency

Controls for the right to information	Implementation	Implementation justification or justification why not
Presentation of the terms & conditions for use/confidentiality	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Possibility of accessing the terms & conditions for use/confidentiality	+	Will remain accessible as part of the process of maintaining consents from Subscribers for further processing.
Legible and easy-to-understand terms	+	Will need to be assured presented as part of the process of obtaining consent from Subscribers for further processing.
Existence of clauses specific to the device	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing (specification that only the location of mobile phones will be considered).
Detailed presentation of the data processing purposes (specified objectives, data matching where applicable, etc)	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.

Detailed presentation of the personal data collected.	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Presentation of any access to the identifiers of the device, the smartphone/tablet or computer, by specifying whether these identifiers are communicated to third parties	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing (pseudonymised IMSIs are communicated to the Solution's enclave and thereby anonymised).
Presentation of the user's rights (consent withdrawal, data erasure, etc)	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Information on the secure data storage method, particularly in the event of sourcing.	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Arrangements for contacting the company (identity and contact details) about confidentiality issues	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Where applicable, information for the user on any change concerning the data collected, the purposes and confidentiality clauses.	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Regarding transmission of data to third parties: - **detailed presentation of the purposes of transmission to third parties** - **detailed presentation of the personal data transmitted** - **indication of the identity of third-party bodies**	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing (NSI as the third party to whom anonymised mobile location data is transmitted).

3.2.1.2 Consent

Controls for obtaining consent	Implementation	Implementation justification or justification why not
Express consent prior to sharing data with NSI	+	Will be obtained as part of the process of obtaining consents from Subscribers for further processing.
Consent presented in an intelligible and easily accessible form, using clear and plain language adapted to the target user (particularly for children)	+	Will be presented as part of the process of obtaining consents from Subscribers for further processing.
Obtaining parents' consent for minors under certain age	+	Will be obtained as part of the process of obtaining consents from Subscribers for further processing.
After a long period without use, the user must be asked to confirm his/her consent	+	Will be asked as part of the process of managing consents from Subscribers for further processing.

3.2.1.3 Right of access

Controls for the right of access	Implementation	Implementation justification or justification why not
Possibility of accessing all of the user's personal data, via the common interfaces	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Possibility of securely consulting the traces of use associated with the user	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Possibility of downloading an archive of all the personal data associated with the user.	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.

3.2.1.4 Right to data portability

Controls for the right to data portability	Implementation	Implementation justification or justification why not
Possibility of retrieving, in an easily reusable format, personal data provided by the user, so as to transfer them to another service.	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.

3.2.1.5 Right to rectification and erasure

Controls for the rights to rectification and erasure	Implementation	Implementation justification or justification why not
Possibility of rectifying personal data.	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Possibility of erasing personal data.	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Indication of the personal data that will nevertheless be stored (technical requirements, legal obligations, etc.)	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Implementing the right to be forgotten for minors	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Clear indication and simple steps for erasing data before scrapping the device	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Advice given about resetting the device before selling it	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Possibility of erasing the data in the event the device is stolen	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.

3.2.1.6 Right to restriction of processing and to object

Controls for the rights to restriction and to object	Implementation	Implementation justification or justification why not
Existence of "Privacy" settings	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.
Effective exclusion of processing the user's data in the event consent is withdrawn	+	Will be implemented based on agreement between the NSI and MNO, as well as in accordance with electronic communications and data protection laws applicable to the MNO.

3.2.1.7 Contractual relations with processors

Processor's name	Purpose	Scope	Contract reference	Compliance with GDPR Art 28
MNO	Process mobile location data by means of the Solution in order to help the NSI produce official statistics	Sample Use Case	To be specified	The processor: - provides sufficient guarantees to implement appropriate technical and organisational measures, - shall not engage another processor without prior specific or general written authorisation of the controller, - processing by a processor shall be governed by a contract or other legal act under Union or Member State law.

3.2.1.8 Requirements for transfer outside EU

No mobile location data will be transferred outside the EU.

3.2.2 Assessment

The above controls, if deployed in full, are considered sufficient to lower identified technical and privacy risks to a level where residual risks are acceptable.

4. Risks

4.1. Existing or planned controls

4.1.1 Explanation

This section lists security controls applicable to the Solution and its deployment. They correspond to the threat mitigations identified in Appendix. The Solution attacker model. Controls are categorised as follows:

1. controls bearing specifically on the data being processed,
2. general security controls regarding the system in which the processing is carried out, and
3. organisational controls /governance.

Note, that while controls belonging to the first category are enforced by the Solution itself, several controls from the latter two categories depend on the environment where the Solution is deployed. In this risk analysis we assume that all those controls are fulfilled.

4.1.1.1 Controls bearing specifically on the data being processed

ID	Short description	Explanation
M1	Authentication and authorisation procedures for users	Users are authenticated to Sharemind HI server and Task enclaves using client certificates. Only authorised users can invoke the Task enclave. Accepted certificates (with corresponding roles) are managed by Coordinators and cryptographically validated/overseen by Enforcers. User generates her own cryptographic key pair and nobody else has copy of the private key. Residual risk: User is responsible for safeguarding her private key, see M14.
M2	Authenticated logging in Sharemind HI server	Sharemind HI keeps authenticated log of task enclaves' invocations and this log can be audited.
M8	Design for temporary fault tolerance	Design retry functionality into the Solution so that if some resource is not available for the consuming device, then the process can resume after recovery of the situation. For that purpose, ensure transactional (i.e., "all or nothing") logic in processing.
M9	Code auditing and authenticated binaries	All project-specific source code of the Sharemind HI Solution is made available and can be audited against malicious functionality. In each zone (MNO-ND, MNO-VAD, NSI) the sysadmin builds the application binary from the

		source code itself or receives a signed copy of the binary from a trusted party who has audited the application source code. Note, Sharemind HI software as generic product is closed source with all components signed by Cybernetica AS.
M10	Pseudonymisation algorithm review	Design efficient pseudonymization – select short enough pseudonym lifetime, design cryptographically strong pseudonym generation. Do review of the design with experts with full documentation of review results.
M12	Cryptographically strong design	Use of Intel SGX technology and privacy by design principles in general. Cybernetica notifies all Sharemind HI hosts of published Intel SGX vulnerabilities[47] and Intel enforces corresponding mitigations at the Intel Attestation Service.
M16	Statistical disclosure control	The Solution implements statistical disclosure control (SDC) with k-anonymity. The SDC is implemented as part of the analysis enclave with the default k-anonymity threshold of 20 and is modifiable before the analysis code is audited and deployed. This makes the SDC tamper proof even against the Sharemind HI host.

4.1.1.2 General security controls regarding the system in which the processing is carried out

ID	Short description	Explanation
M3	Logging on the server side	Apply best practices of logging where applicable (Sharemind HI internal logging is separately covered by M2) – reliably collect and backup logs, ensure that operations and maintenance system warn about low disk space. It is recommended to collect logs on central logging server that has different sysadmin (M5).
M4	Responsibility assigned to the maintainer of the infrastructure	The maintainer (MNO-ND or MNO-VAD) already has the infrastructure set up and running for its internal processes and the project is not changing any requirements. All aspects relating to the security of existing infrastructure are outside the scope of this project.
M6	Limit the network visibility	Limit network visibility for the Solution components to respective network zone (MNO-ND, MNO-VAD, NSI) only. External visibility is needed for df1..df4 and df8 (IAS) connections.

[47] For an overview, refer to Jaak Randmets. *An Overview of Vulnerabilities and Mitigations of Intel SGX Applications*. 2021. https://cyber.ee/research/reports/

M7	Design and test load capabilities	Carefully analyse the load from the Solution to existing system, plan and perform load testing
M11	Implement alarms	The system shall raise alarms and send to Operations and Maintenance system. The goal is to avoid unnoticed erratic behaviour of the system.
M13	Secure communication channels	Use of secure communication technology (e.g., TLS) throughout the Solution in order to provide mutually authenticated, encrypted and integrity-protected communication channels. Planned control: Use secure communication channels for data exchange between MNO-ND and MNO-VAD. While all internal communication of the Solution is protected, we cannot guarantee the same level of protection for other data flows as it is dependent on particular stakeholders and their infrastructure. However, the risk analysis assumes that all communication channels offer sufficient protection. Residual risk: User is responsible for safeguarding her private key, see M14.

4.1.1.3 Organisational controls and governance

ID	Short description	Explanation
M5	Avoid overlapping of roles.	Avoid access rights of single person to several zones – e. g. sysadmin or analyst in MNO-ND, MNO-VAD, NSI, dev, deployment. Role and zone separation makes it more difficult to combine tampering in one zone or role and exploiting of the tampered results in the next steps of the process.
M14	Personnel training	Planned control: Train end users to safeguard their access credentials (e.g., private keys).
M15	Agreements between stakeholders	Have agreements to supress any malicious (in)activity by the stakeholders of the Solution deployment themselves. This malicious behaviour includes, for example, collusion between the Enforcers, delaying the process by inactivity or restricting access to shared resources.

4.1.2 Assessment

The above controls, if deployed in full, are considered sufficient to lower identified technical and privacy risks to a level where residual risks are acceptable.

4.2. Potential privacy breaches

4.2.1 Explanation

In this section we discuss potential privacy breaches that stem from the use of the Solution. Note, that we only cover new threats that are added by the Solution and not the ones that existed also prior to deploying the Solution. For example, we list the possibility of leaking mobile location data via a malfunctioning component of the Solution, but not via the mobile location database at rest as the latter exists prior to deploying the Solution.

Severity and likelihood scores are given on a five-point scale: Very low, Low, Moderate, High and Very High. These scores are then combined into a final risk score using Table 2 from the NIST Special Publication 800-30 Guide for Conducting Risk Assessments.

Table 2. Finding risk as combination of likelihood and impact (severity). Originally published as Table I-2 in NIST SP 800-30 Guide for Conducting Risk Assessments.

Likelihood (Threat Event Occurs and Results in Adverse Impact)	Level of Impact				
	Very Low	Low	Moderate	High	Very High
Very High	Very Low	Low	Moderate	High	Very High
High	Very Low	Low	Moderate	High	Very High
Moderate	Very Low	Low	Moderate	Moderate	High
Low	Very Low	Low	Low	Low	Moderate
Very Low	Very Low	Very Low	Very Low	Low	Low

Severity scores are given relative to the current use case i.e., leaking raw mobile location data is the worst that can happen. However, it's improbable that somebody's life is in imminent danger because of this.

Moreover, the severity and likelihood scores given in this document are provided for convenience by the team designing the Solution. These should be reviewed and updated as needed by the MNO and NSI specialists. It should be noted that these values can vary for different MNOs and NSIs.

The threat identifiers at the bottom of each following table refer to the threats listed in the technical risk catalogue in Section 6.4.

4.2.1.1 Illegitimate access of data

Risk sources	External attacker or MNO-VAD insider	**R1**
Threats	Attacker with access to pseudonymised mobile location data at MNO-VAD modifies the pseudonymisation component (PSC) deployed at	

	MNO-ND so that it produces pseudonyms with the key known to the attacker.		
Potential impacts	Attacker gains access to mobile location records, linkable over several pseudonymisation periods (see severity for limitations). Thus, attacker learns mobile devices' movement patterns. Note, that the attacker does not learn IMSIs as these are pseudonymised twice.		
Controls	M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of roles. M11 – Implement alarms.		
Severity	Medium	Leaking data this way has a very limited window of opportunity as integrity checks on pseudonyms will fail on such pseudonyms (M11). Therefore, only a few pseudonymisation days' worth of data could be abused.	
Likelihood	Low	Requires a very special access. Attacker has to be someone who can tamper with the PSC deployment while not having access to the PSC component (MNO-ND premises) itself. Otherwise, it would be easier to leak the raw mobile location data available to PSC.	
Risk score	Low	**Threat identifiers**	T_p1

Risk sources	External attacker or MNO-VAD insider. Together with MNO-ND insider with limited access (see likelihood for details).		**R2**
Threats	The MNO-ND insider extracts and leaks periodic pseudonymisation keys from the pseudonymisation component to the attacker.		
Potential impacts	Attacker gains access to mobile location records, linkable over several days. Thus, attacker learns mobile devices' movement patterns. Note, that the attacker does not learn IMSIs as these are pseudonymised twice.		
Controls	M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of roles. M6 – Limit the network visibility.		
Severity	High	The pseudonyms are not tampered with and thus this attack is not easily detected. So, attacker can accumulate linkable movement data over a long period of time.	
Likelihood	Low	Requires a very special limited access for leaking periodic pseudonymisation keys from PSC, but not access to the raw mobile location data itself.	

Risk score	Low	**Threat identifiers**	I_p1, T_p1, I_ds1

Risk sources	External attacker or MNO-ND insider		**R3**
Threats	Attacker modifies PSC so that it leaks raw mobile location data.		
Potential impacts	Attacker gains access to mobile location records, linkable over several days. Thus, attacker learns mobile devices' movement patterns. Note, that attacker does not learn IMSIs as these are pseudonymised twice.		
Controls	M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of roles. M6 – Limit the network visibility.		
Severity	Very high	Attacker can accumulate linkable movement data over a long period of time.	
Likelihood	Very Low	The Solution has mitigations in place against modifying PSC. It is also assumed that the PSC host (MNO-VAD) has monitoring in place to be alerted about suspicious network activity.	
Risk score	Low	**Threat identifiers**	I_p1, T_p1

Risk sources	External attacker or MNO-VAD insider	**R4**
Threats	Attacker uses the data import process (MNO-VAD) to filter data before it is imported into Sharemind HI solution so that it only contains records from certain countries (or set of countries), as this attribute is available in plaintext. The analysis would then be performed for this specific country or set of countries. In principle, an MNO-ND insider could perform similar attack at pseudonymisation process, but they do not have direct access to the analysis report.	
Potential impacts	Attacker gains access to statistics about more granular subset than normal.	
Controls	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND). M5 – Avoid overlapping of roles. M16 – Statistical disclosure control.	

	Proposed controls • Remove the country identifiers from mobile location data at pseudonymisation process at MNO-ND • Add the total number of countries to the analytics enclave log. This discourages the attacker as the attack could be more easily discovered.		
Severity	Low	Attacker learns statistical indicators for roaming subscribers from one or more countries. No individual records leak. The granularity of the indicators is further limited by the statistical disclosure control (SDC) mechanism built into the analytics enclave.	
Likelihood	Moderate	A rogue MNO-VAD employee can do this if it's not detected by the analysis automatically. However, one cannot perform this attack over a long period as it would stand out in the analysis reports. Requires an attacker to access both the data import and analysis output parts.	
Risk score	Low	**Threat identifiers**	T_p2, (T_p1)

Risk sources	External attacker		**R5**
Threats	Attacker gets to run modified task enclave code by persuading or bribing all Enforcers. An insider (e.g., one of the Enforcers) also provides the attacker with the corresponding enclave output.		
Potential impacts	Attacker gains access to mobile devices' movement data over long periods. However, attacker does not get access to IMSI codes.		
Controls	M12 – Cryptographically strong design. M9 – Code auditing and authenticated binaries. M15 – Agreements between stakeholders		
Severity	Very high	Attacker can obtain periodic pseudonymisation keys to obtain raw mobile location data (with some extra effort) or run customised analysis on this data.	
Likelihood	Very low	There are several independent Enforcers in the system working for different stakeholders, each with their own stake in the system.	
Risk score	Low	**Threat identifiers**	T_p4

Risk sources	NSI insider	**R6**

Threats	Attacker modifies NSI input data (i.e., reference area definitions in the report request) in order to leak personal information from analysis report.		
Potential impacts	Attacker learns, for a chosen low-density tile, for each daily subperiod how many subscribers have this tile in their usual environment. This data is equivalent to that of report D', but without SDC applied.		
Controls	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI).		
Severity	Moderate	Applying SDC for one of the subreports for the chosen tile is cancelled. However, with oversight of NSI report queries, this attack can be identified.	
Likelihood	Low	This attack requires the attacker to construct a set of two reference areas that differ only by the chosen tile. The chosen tile must not be in the usual environment of any subscribers that have the rest of the reference area in their usual environment. In addition, there must be an external tile that both the chosen tile and the rest of the reference area have strong connection to. Finding such a set limits the practicality of this attack.	
Risk score	Low	**Threat identifiers**	T_p3, T_ds3

Risk sources	Sharemind HI host (MNO-VAD) or an attacker inside MNO-VAD security perimeter.	**R7**
Threats	Attacker makes use of the fact that Sharemind HI host can, with access to snapshots, fork and roll back analysis enclave states. First, the attacker rolls the analysis enclave state back to where a new report period is just started, and the accumulated subscriber footprint (S) is empty. Second, the attacker chooses a random record one day's periodic update of subscriber footprints (H file), creates a new H' with just this one record with it and imports it to the analysis enclave. The enclave's state (S) now consists of exactly this one record/device. Third, the attacker chooses a record from the next day's H file (with new pseudonyms) and imports it to the analysis enclave. If the size of S increases, the two records belong to different devices, otherwise they belong to the same device. This way the attacker can try all different pseudonyms from the second day's H file to find the one matching the device in H' from the previous day.	
Potential impacts	Attacker is able to connect pseudonyms of a mobile device over several days and thus learns their moving patterns.	
Controls	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). M5 – Avoid overlapping of roles.	

Severity	Moderate	Attacker learns mobile device's moving patterns, but without any prior knowledge on which pseudonym to attack, it is a randomly selected mobile device.	
Likelihood	Low	Matching the pseudonymised mobile location records over several days like this is a very slow process and the attacker can target only one (random) mobile device at a time. Depending on the extent on applicable controls, this act can be caught.	
Risk score	Low	**Threat identifiers**	T_p4

4.2.1.2 Unwanted change of data

There are no privacy breaches identified that originate from data tampering. The Solution implements secondary use of mobile location data and the original mobile location records in MNO-ND database are never modified by the Solution (it only has read-only access to this data).

4.2.1.3 Disappearance of data

Risk sources	External attacker or MNO-VAD insider	**R8**
Threats	Attacker uses the data import process (MNO-VAD) to filter data before it is imported into Sharemind HI solution, removing (some) mobile location records belonging to a specific tile or set of tiles. The analysis would then skip those records. In principle, an MNO-ND insider could carry out similar attack when exporting mobile location records to MNO-VAD. However, this already existing data flow is out of scope for this analysis and MNO-ND insider is less likely to perform such indirect attack as they also have raw mobile location data available.	
Potential impacts	The analysis cannot take the removed mobile location records into account. Applying this attack over a long period (e.g., the whole reporting period) produces biased statistics that may yield discrimination for people that have the removed tiles in their usual environments (e.g., if the removed tiles contain a place of worship or other points of interest).	
Controls	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND). M5 – Avoid overlapping of roles. <u>Proposed controls</u> The pseudonymisation component can encrypt the whole mobile location record, i.e., encrypting also tile coordinates not just the subscriber information (technically, pseudonymisation used by the	

	Solution is encryption as it must be possible to reverse it in the analytics enclave). The pre-processing functionality at MNO-VAD (Module A) must then also be implemented in the analytics enclave as only the latter can decrypt mobile location records.		
Severity	Moderate	Attacker does not gain any additional information and does not learn any movement patterns. However, in the long run, it may have negative impact for a group of people.	
Likelihood	Moderate	The attacker cannot target any individuals. However, this attack is almost impossible to discover if the attacker is careful not to remove too much data so that it would stand out in the statistical reports.	
Risk score	Moderate	**Threat identifiers**	T_p2, (T_p1)

4.2.2 Assessment

As stated in Section 2.2, in this document we analyse the privacy risks compared to an already existing mobile location data analysis process in the MNO. It is expected that deploying the Solution for additional analysis introduces new privacy risks as it adds several new processes and components to the already existing infrastructure. However, considering that all but one of the newly added privacy risks have Low or Very Low scores, we conclude that the added functionality of the Solution outweighs the added risk to privacy.

The only identified privacy risk with Moderate score involves malicious data filtering before it is imported into the Solution (see risk R8). In principle, input data for any similar solution would be vulnerable to a similar attack. The additional proposed controls make it impossible to modify input data at MNO-VAD and thus mitigate this risk.

5. Validation

5.1. Preparation for validation

This document provides a template for a DPIA for deploying the Solution, with as many details filled in as possible. However, final details depend on the exact chosen use case as well as on specific stakeholders with their policies and infrastructure. In order to finalise this template either for internal validation or for obtaining authorisation from the relevant data protection authorities, the following steps are needed:

1. select a statistical analysis use case along with appropriate statistical methodologies suitable for implementing in a real-world scenario by means of the Solution, update Section 2 of this document accordingly,
2. consult stakeholders (e.g. document any advice of the persons in charge of data protection aspects, opinions of data subjects or their representatives).
3. determine the legal basis for the chosen use case, re-evaluate and if necessary, update the legal part of this document (Section 3),
4. conclude an agreement between the NSI and the MNO for implementing the selected use case, specifying the means and purposes of the processing (the Solution), ensuring protection measures to match the requirements for setting up the Solution (see Section 4.1), dividing the roles of controller and processor, and assistance in obtaining consent from Subscribers,
5. re-evaluate and if necessary, update the Solution attacker model (see Section 6), including technical risks, applicability of proposed controls and residual risks,
6. re-evaluate and if necessary, update privacy risks and their likelihood and severity scores (see Section 4.2) according to the chosen use case and deployment environment,
7. provide an action plan for dealing with the additional controls identified during the previous steps (determine the person responsible for its implementation, its cost and estimate timeframe etc),
8. conduct a formal validation procedure with relevant officers and decision makers of the stakeholders (NSI, MNO), providing them with a visual presentation of the controls selected to ensure compliance with the fundamental principles and contribute to data security (visual map of controls) and a visual presentation of the risks, both initial and residual, where applicable (visual map of risks).
9. consult with or obtain authorisation from the relevant data protection authorities,
10. set up and configure the Solution, carry out relevant penetration tests before collecting pseudonymised real-world mobile location data.

When implementing these steps, the following division of tasks can be envisioned between the different stakeholders (see Figure 10 below). The actors marked in “bold” may add previously non-covered ethical or legal risks that can result in rescoping, additional organizational, technical or physical protection measures or discontinuation of the project.

Figure 10 – General division of tasks in the validation process

No	By/From	To	Activities
1.	NSI	**NSI (ethics committee)**	- Describe the envisioned mobile location data processing use case, including the purposes of data processing - Complete the reference instance of DPIA by adding use-case specific purpose and scope (sample size, study period, geography) and present it to the ethics committee - Evaluate if the envisioned mobile location data processing use case is in conformity with the relevant codes of conduct, statistical principles and professional standards applicable in the given field of statistics
2.	NSI	**MNO**	- Propose a cooperation project for carrying out the envisioned mobile location data processing use case (hereinafter "**project**") - Determine if the data required for the project is available
3.	NSI/MNO	**DPA**	- Consult the DPA prior to carrying out the project - Submit the final DPIA, including use-case specific purpose and scope (sample size, study period, geography) - Implement additional requirements set forth by the DPA for carrying out the project
4.	NSI/MNO	DPA	- Test the project, using synthetically generated data - Consult the DPA regarding the test results - Implement changes in the project, where necessary, based on test results
5.	NSI/MNO		- Carry out the project

6.	NSI	General public	- Analyse the statistical results received from the project - Publish the analysis

6. Appendix. The Solution attacker model

6.1. The STRIDE methodology

STRIDE[48] is a threat modelling methodology. It works on a data flow graph of a system in question. The data flow graph represents the systems attack surface. It consists of four types of nodes:

- Process (p), the active part of the system (e.g., code)
- Data Store (ds), data at rest
- Data Flow (df) showing how data moves between processes
- External Entities (i), users interacting with the system.

In addition, the data flow graph can include trust boundaries (usually shown as dashed curved lines) that indicate where data crosses a trust boundary (e.g., moves between different owners or organisations).

For each node, we analyse these possible threats:

	Threat	Violated property	Threat definition
S	Spoofing identity	Authentication	Pretending to be something or someone other than yourself.
T	Tampering with data	Integrity	Modifying something on disk, network, memory, or elsewhere.
R	Repudiation	Non-repudiation	Claiming that you didn't do something or were not responsible; can be honest or false.
I	Information disclosure	Confidentiality	Providing information to someone not authorized to access it.
D	Denial of service	Availability	Exhausting resources needed to provide service.
E	Elevation of privilege	Authorization	Allowing someone to do something they are not authorized to do.

However, all threats are not applicable to all node types. The following table shows which threats apply to which node types:

Element	S	T	R	I	D	E
External entity	x		x			
Process	x	x	x	x	x	x
Data store		x	o[49]	x	x	
Data flow		x		x	x	

[48] Kohnfelder, Loren; Garg, Praerit. *The threats to our products*. Microsoft, 1999.
[49] Repudiation threat only applies to data stores that contain logs.

6.2. Attack Surface of the Solution

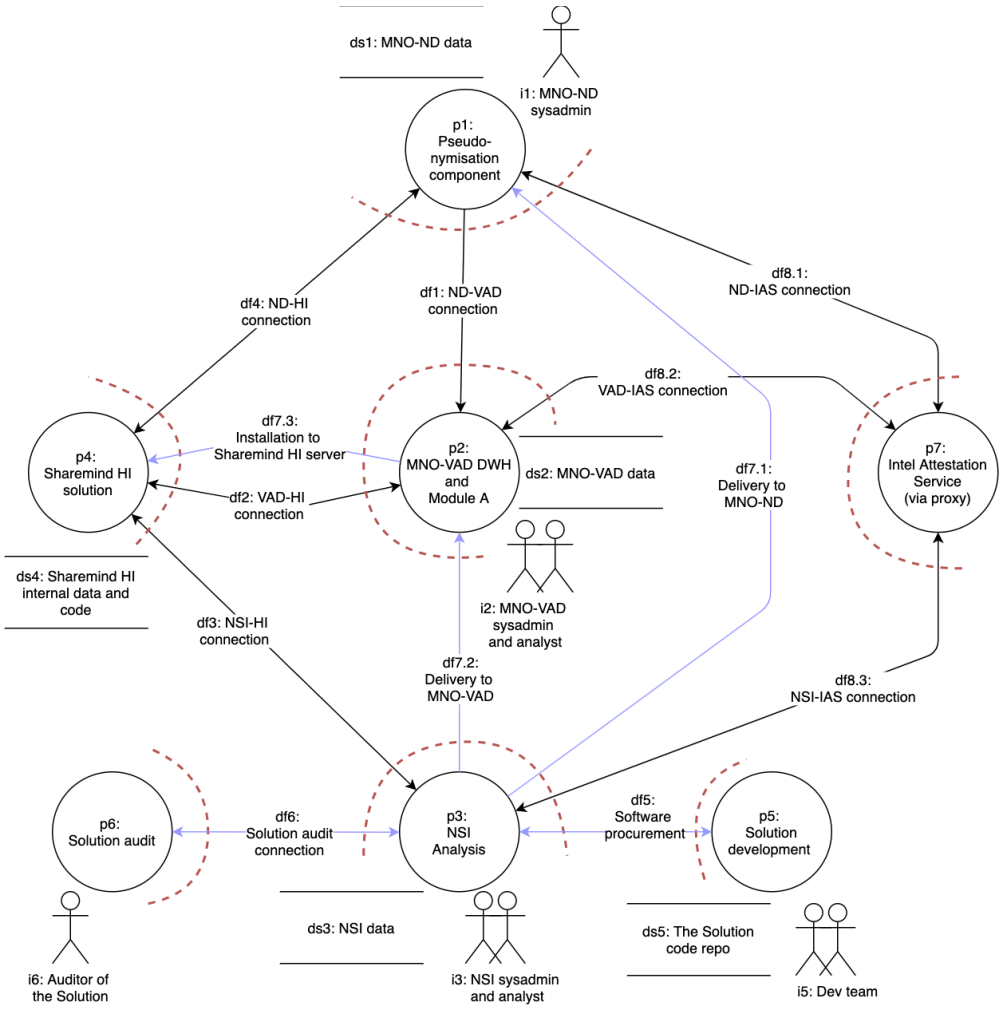


Figure 11 – Attack surface. Violet arrows represent dataflows related to design, audit and deployment of the Solution. Black arrows represent main process of data import and report calculations.

Table 3. Attack surface elements

p1	Data pseudonymization and export application in MNO-ND
i1	The system administrator in MNO-ND
ds1	The data processed and stored in MNO-ND

p2	MNO-VAD data import process for preparing and importing pseudonymised mobile location data into the Solution. This also includes Module A.
i2	The system administrator and analyst in MNO-VAD
ds2	The data imported, stored and processed in MNO-VAD, excluding data in Sharemind HI TEE
p3	NSI scripts for uploading data layers and downloading the results
i3	The system administrator and analyst in NSI
ds3	The input data and results data in NSI
p4	Sharemind HI server
ds4	Data and code inside Sharemind HI solution TEE
p5	Solution development
i5	Development team
ds5	The Solution code repository
p6	Solution audit
i6	Auditor of the Solution
p7	Intel Attestation Service (via proxy, if necessary)
df1	The connection between MNO-ND and MNO-VAD
df2	The connection between MNO-VAD and Sharemind HI server
df3	The connection between NSI and Sharemind HI server
df4	The connection between MNO-ND and Sharemind HI server
df5	The connection between NSI and the development team
df6	The connection between NSI and auditor

df7.1	The connection between NSI and MNO-ND for initial software delivery
df7.2	The connection between NSI and MNO-VAD for initial software delivery
df7.3	Deployment of Sharemind HI at MNO-VAD
df8.1	The connection between Intel Attestation Service and MNO-ND
df8.2	The connection between Intel Attestation Service and MNO-VAD
df8.3	The connection between Intel Attestation Service and NSI

In the following, we analyse each element shown on Figure 11 according to the STRIDE methodology with the following exceptions:

- Individual elements of the Solution development (p5, i5, ds5, df5) are not analysed independently. Cybernetica is an ISO 9001 and 27001 certified company with internal policies on information security and secure software development. For the security of the Solution, it is only important that the development deliverables are signed by Cybernetica and can be audited.
- Individual elements of Solution audit (p6, i6, df6) are not analysed independently. Auditor is a trusted third party with their own internal processes. For the security of the Solution, it is only important that the deliverables are audited.
- The Solution delivery and deployment data flows (df7.x) are assumed to be using secure communication channels. It is also assumed that software packages' signatures are verified on deployment.
- The risk analysis refers directly to the Intel Attestation Service (IAS) in p7, without the proxy service provided by Cybernetica. This is because in the production deployment the IAS proxy is not necessary.

6.2.1 Data details

Table 4. Data assets relation to technical components and BPMN elements in the analysis document. Same element can appear several times if sent over to other zone/trust boundary.

Attack surface element	BPMN element	Description
ds1 (MNO-ND data)		
	D2.2	Protected New periodic pseudonymisation key
	D2.3	New periodic pseudonymisation key
	D2.4	MNO database
	D2.5	Pseudonymised MNO data

ds2 (MNO-VAD data)		
	D2.5	Pseudonymised MNO data
	D2.6	Pseudonymised MNO database
	D3.3	Analysis status
	D3.4	Encrypted statistical analysis results
	D3.5	Statistical analysis results
	D5.1	Pseudonymised footprints for 1 period
ds3 (NSI data)		
	D3.1	Analysis input data NSI
	D3.2	Analysis request
	D3.3	Analysis status
	D3.4	Encrypted statistical analysis results
	D3.5	Statistical analysis results
ds4 (Sharemind HI internal data and code)		
	D2.1	Stored pseudonymisation keys
	D2.2	Protected new periodic pseudonymisation key
	D3.1	Analysis input data NSI
	D3.5	Encrypted statistical analysis results
	D5.2	Reverse pseudonymised footprints for 1 period
	D5.3	Aggregated data
	D5.4	Analysis results
ds5 (Solution code repo)		
	D1.1	Analysis application

6.2.2 Process Details

The following table connects STRIDE data flow diagram processes to BPMN actions in the project analysis document. Note that this document labels process elements with lower-case “p”, whereas in the BPMN, processes are labelled with upper-case “P”.

Table 5. Processes in BPMN vs attack surface process layout.

Attack surface element	BPMN element	Description
p1 (Pseudonymisation component)		
	P4	Perform remote attestation.
	P4.1	Ask for verification quote.
	P4.4	Receive and send verification quote.
	P4.8	Check attestation report and its signature.
	P1.10	Request Dataflow Configuration.
	P1.13	Perform verification.
	P1.14	Approve Dataflow Configuration.
	P1.15	Send signed Dataflow configuration
	P2.1	Ask Sharemind HI for next pseudonymisation key.
	P2.3	Download the periodic pseudonymisation key.
	P2.4	Decrypt the new periodic pseudonymisation key.
	P2.5	Generate new pseudonymisation key.
	P2.6	Compile pseudonymised mobile location data.
	P2.7	Send pseudonymised MNO data.
	P2.10	Delete new periodic pseudonymisation key.
p2 (MNO-VAD data import process and Module A)		
	P1.8	Install the application and start Sharemind HI server.
	P1.9	Send verification commencement to Enforcers
	P1.10	Request Dataflow Configuration
	P1.11	Extract dataflow configuration
	P1.12	Send dataflow configuration.
	P1.13	Perform verification.
	P1.14	Approve Dataflow Configuration.

	P1.15	Send signed Dataflow configuration
	P1.16	Store Dataflow Configuration approval.
	P4	Perform remote attestation.
	P4.1	Ask for verification quote.
	P4.3	Send verification quote.
	P4.4	Receive and send verification quote.
	P4.8	Check attestation report and its signature.
	P2.8	Store pseudonymised MNO data.
	P2.9	Perform internal analysis on pseudonymised mobile data.
	P3.3	Receive NSI inputs and analysis request
	P5	Calculate analysis results
	P5.1	Calculate pseudonymised footprints for 1 period (Module A)
p3 (NSI)		
	P1.1	Develop an analysis application.
	P1.2	Send analysis application.
	P1.6	Compile deployment package.
	P1.7	Send Deployment Package to MNO-VAD
	P4	Perform remote attestation
	P4.1	Ask for verification quote.
	P4.4	Receive and send verification quote.
	P4.8	Check attestation report and its signature.
	P3.1	Prepare the analysis request
	P3.2	Upload NSI inputs and analysis request
	P3.4	Wait for analysis completion
	P3.5	Download and decrypt analysis results
p4 (Sharemind HI TEE)		
	P2.2	Generate next pseudonymisation key.

	P4.2	Compile and sign verification quote.
	P5.2	Load and reverse pseudonymise footprints for 1 period
	P5.3	Calculate aggregated values
	P5.4	Compute specified reports, apply SDC
	P5.5	Encrypt analysis results
p5 (Development controlled by NSI)		
	P1.1	Develop an analysis application.
	P1.2	Send the analysis application.
p6 (Solution Audition)		
	P1.3	Assess privacy risks of the application.
	P1.4	Compile the fingerprint set of the analysis application.
	P1.5	Send the fingerprint set.
p7 (Intel Attestation Service)		
	P4.5	Verify quote and its signature.
	P4.6	Produce and sign attestation report.
	P4.7	Send attestation report.

6.3. Adversary classification

In this document by “attacker” we mean one of the following adversaries:

- Dedicated individual, including an insider
- Large corporation or dedicated competitor
- User error

This document does not model these adversaries:

- Nation state
- Law enforcement
- Global adversary

Law enforcement has legitimate access to mobile location data under the law. Adding nation state or global adversaries to the threat model adds too many details and makes the risk assessment difficult to follow. If necessary, these attacker types can be analysed separately.

A global adversary is an entity with full view of the network, unlimited power and sufficient resources.

6.4. Technical Risks Catalogue

Columns in the risk catalogue:

1. Target – item from Table 3
2. Attack type – STRIDE classifier
3. Threat – threat description
4. Security measures/controls and residual risks – analysis of security controls and residual risks. Descriptions of security measures are given in Section 4.1.

Note that individual threats are referred to as "<attack type>_<target>" throughout the document. For example, S_p1 stands for spoofing attack on the pseudonymisation component.

Target	**Attack type**	**Threat**	**Security measures/controls and residual risks**
p1	S	Attacker replaces the pseudonymisation application with a fake copy.	M9 – Code auditing and authenticated binaries.
p1	T	Attacker reverse engineers and modifies the pseudonymisation application.	M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of roles. Prevents attacker in ND to take advantage of distorted data sent to VAD and NSI M11 – Implement alarms. Alarms on corrupted or severely distorted data distribution improve detection of malicious activities. <u>Residual risks:</u> Attacker modifies/replaces the application after it has been deployed Despite the countermeasures set up by MNO-VAD the attacker obtains access to output data to exploit the results of tampered processing.

p1	R	User claims not to have invoked a certain functionality in the pseudonymisation application.	M2 – Authenticated logging in Sharemind HI server. M3 – Logging on the server side (MNO-ND). Residual risks: Logging is not sufficiently detailed and does not provide sufficient evidence. With sufficient privileges the measure M3 can be disabled by tampering. This in turn is mitigated by M5 – Avoid overlapping of roles.
p1	I	Pseudonymisation component leaks mobile location data.	M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of roles. M6 – Limit the network visibility.
p1	I	Pseudonymisation component leaks periodic pseudonymisation keys.	M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of roles. M6 – Limit the network visibility.
p1	I	Attacker bypasses countermeasures and records the pseudonymization keys provided by Sharemind HI which are supposed to be deleted immediately after use.	M5 – Avoid overlapping of roles. Role separation makes it more difficult to exploit the results of tampering in next steps of process. Residual risk: Attacker obtains access or has accomplice in MNO-VAD zone of the system to exploit tampered processing. See corresponding privacy risk in Section 4.2.1.1.
p1	D	Attacker renders the pseudonymisation component unavailable.	Pseudonymisation component should not be made available over the network, it is meant to be a local application (although technically running as a service). We assume that security of end-user workspace is handled by the owner of the infrastructure. M4 –

			Responsibility assigned to the maintainer of the infrastructure (MNO-ND).
p1	E	Elevation of privileges at the pseudonymisation component	N/A, PSC has a single level of privileges. We assume that the mobile location database has sufficient protection against such attack. M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND). Privilege elevation against Sharemind HI server is covered separately, see E_p4.
i1	S	Attacker is impersonating MNO-ND operator when communicating with the Sharemind HI Solution.	M1 – Authentication and authorisation procedures for users
i1	R	MNO-ND operator claims not to have invoked the periodic pseudonymisation process with Sharemind HI-generated key.	M1 – Authentication and authorisation procedures for users M2 – Authenticated logging in Sharemind HI server M3 – Logging on the server side With sufficient privileges the measure M3 can be disabled by tampering. This in turn is mitigated by M5 – Avoid overlapping of roles.
df1	T	Attacker modifies data sent from PSC to MNO-VAD	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND, MNO-VAD). This data flow crosses security boundaries but is still internal to MNO. Thus, only insider attacks are possible. M5 – Avoid overlapping of roles. Prevents attacker in ND to take advantage of distorted data sent to VAD <u>Prevention of residual risks:</u> Use secure communication channel to protect data integrity. See M13 – Secure communication channels.
df1	I	Attacker leaks pseudonymised mobile	M4 – Responsibility assigned to the maintainer of the infrastructure

		location data flowing from ND to VAD.	(MNO-ND, MNO-VAD). This data flow crosses security boundaries but is still internal to MNO. Thus, only insider attacks are possible. Prevention of residual risks: Use secure communication channel to protect data confidentiality. See M13 – Secure communication channels.
df1	D	Attacker denies or limits pseudonymisation component's communication with MNO-VAD	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND, MNO-VAD). This data flow crosses security boundaries but is still internal to MNO. Thus, only insider attacks are possible. M5 – Avoid overlapping of roles. Prevents attacker in ND to take advantage of distorted data sent to VAD.
df4	T	Attacker modifies data exchanged between PSC and Sharemind HI solution.	M13 – Secure communication channels.
df4	I	Pseudonymised mobile location data flowing from PSC to Sharemind HI solution leaks.	M13 – Secure communication channels.
df4	D	Attacker denies or limits pseudonymisation component's communication with the Sharemind HI solution.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND, MNO-VAD). This data flow crosses security boundaries but is still internal to MNO. Thus, only insider attacks are possible. M5 – Avoid overlapping of roles. Prevents attacker in ND to take advantage of distorted data sent to VAD or NSI. M11 – implement alarms. Detect process deviations before data is lost irreversibly Residual risks:

			Sharemind HI solution being unavailable for PSC is an accepted residual risk. It is also possible to pseudonymise mobile location data using a self-generated periodic pseudonymisation key.
ds1	T	Attacker modifies the mobile location database.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND). We assume that mobile location database integrity is protected by the MNO-ND, as this data element exists prior to deploying the Solution. The Solution (more specifically, the pseudonymisation application, PSC) needs only read-only access to the database.
ds1	T	Attacker modifies periodic pseudonymisation key	M11 – Implement alarms. Raise an alarm as soon as possible if corrupted data is identified. Mobile location data pseudonymised with corrupted key raises a verification error. M5 – Avoid overlapping of roles. Prevents attacker in ND to take advantage of distorted data sent to VAD or NSI.
ds1	I	Mobile location database confidentiality is breached (contents leak).	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND). The Solution adds pseudonymisation application as a new component that has access to this data store. However, pseudonymisation application processes mobile location database locally within the security perimeter of MNO-ND. Its output, pseudonymised mobile location data (part of ds2) is considered less sensitive and is covered separately. M6 – Limited network visibility.
ds1	I	Periodic pseudonymisation key leaks	The periodic pseudonymisation key is used only internally by the pseudonymisation component and

			deleted after use. M6 – Limit the network visibility. <u>Residual risks:</u> Residual risks are accepted as-is. As PSC is internal component for MNO-ND, its operators also have access to the raw mobile location data. Therefore, leaking the periodic pseudonymisation key gives no further gain.
ds1	D	Mobile location database is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-ND). We assume that availability of the mobile location database is guaranteed by the MNO-ND, as this data element exists prior to deploying the Solution. As the Solution needs read-only access to the database, it cannot render the database inaccessible by corrupting the data or misconfiguring the database system. Moreover, it is highly unlikely that it will overload the database system as the data volume is small compared to the database's normal load. M7 – Design and test load capabilities M8 – Design for temporary fault tolerance
ds1	D	Periodic pseudonymisation key is (made) unavailable.	<u>Residual risks:</u> This is an accepted residual risk. As a fallback, it is also possible to pseudonymise mobile location data using a self-generated periodic pseudonymisation key.
p2	S	Attacker creates a fake copy of the data import process at MNO-VAD (DWH, Module A, the script calling HI enclaves). The fake process	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD) – best engineering

		can receive data from MNO-ND and/or run HI enclave services	practices must be used for DWH and Module A. Components created by Cybernetica (scripts invoking the Sharemind HI enclave) are signed. M9 – code auditing and authenticated binaries. Mitigation M5 – Avoid overlapping of roles. Separation of Sharemind HI installer, Sharemind HI runner and Sharemind HI report output consumer roles will reduce exploitation potential of an attack.
p2	T	Attacker reverse engineers and modifies the data import process at MNO-VAD.	The DWH and Module A are external components for this analysis. M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). Components delivered by Cybernetica can be audited or be signed. M9 – Code auditing and authenticated binaries. M5 – Avoid overlapping of the roles. The role deploying (modifying) components in this process shall not have direct access to analytical reports produced by this process. Residual risk: Despite countermeasures set up by MNO-VAD the attacker obtains access to output data to exploit the results of tampered processing.
p2	R	User claims not to have invoked the Sharemind HI enclave.	M2 – Authenticated logging in Sharemind HI server.
p2	R	User claims not to have invoked the DWH export or Module A.	Residual risks: The DWH and Module A are external components for this analysis. Possible mitigations include M3 – Logging on the server side.

			Preferably so that potential attacker cannot also delete or modify the logs. M5 – Avoid overlapping roles.
p2	I	Attacker leaks pseudonymised mobile location data	See I_ds2
p2	D	Attacker makes (parts of) the data import process at MNO-VAD unavailable.	M6 – Limit the network visibility. Protect process components from external parties. This process is internal to MNO-VAD and thus we assume it to be covered my infrastructure management. M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD).
p2	E	Attacker gains higher privileges in the data import workflow.	Locally, there is a single level of privileges in the data import workflow at MNO-VAD. Role based access permissions are enforced by the Sharemind HI server, see E_p4.
ds2	T	Attacker modifies reports calculated by the Solution. Also includes the protected version of the report.	M12 – Cryptographically strong design. The reports are created inside Intel SGX enclaves and written on disk using authenticated encryption. Therefore, tampering raises alarms. M1 – Authentication and authorisation procedures for users. Only authorised users can download the analysis report. M13 – Secure communication channels. The analysis report is downloaded via TLS. M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). After the report has been securely downloaded, the responsibility goes over to MNO-VAD internal procedures.

ds2	T	Attacker modifies pseudonymised mobile location database.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This is part of infrastructure existing prior to deploying the Solution.
ds2	T	Attacker modifies pseudonymised mobile location data. Includes also modification of the footprints produced by Module A.	Module A is out of scope for this analysis. M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). M6 – Limited network visibility. M5 – Avoid overlapping roles. Possible abuse of leaked data is reduced if users accessing pseudonymised mobile location data cannot see the analysis results. Residual risks: In spite of countermeasures, the data can still be modified.
ds2	I	Attacker obtains access and leaks reports calculated by the Solution.	M1 – Authentication and authorisation procedures for users Mitigation M5 – Avoid overlapping of the roles. The role running services in P2 shall not have direct access to analytical reports produced by P2.
ds2	I	Pseudonymised mobile location database confidentiality is breached (contents leak).	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This is part of infrastructure existing prior to deploying the Solution. M6 – limited network visibility. Residual risks: Note that pseudonymised mobile location data is exported from the pseudonymised mobile location database and stored on a disk before importing it to the Sharemind HI solution. In the PoC it is accepted risk that the exported copy on disk might be less protected than the

			DWH. In commercial deployment, such temporary files should be avoided if possible.
ds2	I	Pseudonymised mobile location data leaks. Includes also confidentiality breach of the footprints produced by Module A.	Module A is out of scope for this analysis. M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). M6 – Limited network visibility. M9 – Code auditing and authenticated binaries (for Cybernetica's components only). M5 – Avoid overlapping roles. Possible abuse of leaked data is reduced if users accessing pseudonymised mobile location data cannot see the analysis results. M12 – Cryptographically strong design. Abusing leaked data is difficult if pseudonymisation cannot be reversed without the well-protected key.
ds2	D	Analysis report is (made) unavailable	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). The analysis report is a file on MNO-VAD infrastructure. <u>Residual risks:</u> This is an accepted residual risk as the report can be recomputed.
ds2	D	Pseudonymised mobile location database is (made) unavailable	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This is part of infrastructure existing prior to deploying the Solution.
ds2	D	Pseudonymised mobile location data is (made) unavailable. Includes also availability of the footprints produced by Module A.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This data is a file on MNO-VAD infrastructure. <u>Residual risks:</u> This is an accepted residual risk as the mobile location data can be re-pseudonymised and imported from

			MNO-ND and intermediate data sets can be recomputed.
i2	S	Attacker is impersonating MNO-VAD operator when communicating with the Sharemind HI client application for the data import process.	M1 – Authentication and authorisation procedures for users
i2	R	MNO-VAD operator claims not to have made daily footprint data available for Sharemind HI Solution	M2 – authenticated logging in Sharemind HI server
i2	R	MNO-VAD operator claims not to have downloaded the results from Sharemind HI Solution	M2 – authenticated logging in Sharemind HI server M10 – logging on the server side.
p4	S	Attacker replaces Sharemind HI Solution with a copy with different functionality.	M9 – Code auditing and authenticated binaries. M12 – Cryptographically strong design. Intel SGX technology and remote attestation make it possible to make sure that the user is connected to an authentic copy of the program (enclave) running on genuine platform.
p4	T	Attacker modifies the Sharemind HI solution.	M9 – Code auditing and authenticated binaries. M12 – Cryptographically strong design. Use of Intel SGX technology with enhancements by Sharemind HI (e.g., use of Enforcers and access control) protects against tampering during deployment and also while in use. Neither the platform nor the enclaves can be modified without raising alarms. <u>Residual risks:</u> Sharemind HI host (MNO-VAD) is able to fork Sharemind HI enclave states (make a snapshot of the state

			and run attacks on this copy without affecting the original state or leaving traces in the original audit log). This can at least partly be mitigated by storing the audit log independently from the Sharemind HI host, e.g., at NSI. Furthermore, it can be mitigated by assigning more granular sets of access permissions, so MNO-VAD cannot, for example, both add new data and run the analysis enclave. This required enhancements to M5 – Avoid overlapping of the roles.
p4	R	User claims not to have invoked any of the enclave activities in the Sharemind HI Solution.	All user actions require authentication by a valid user. M1 – authentication and authorisation procedures for users. All user actions are logged on the server side. M2 – Authenticated logging in Sharemind HI server. See repudiation threats on connected processes: R_p1, R_p2, R_p3.
p4	I	Sharemind HI Solution leaks data.	M9 – Code auditing and authenticated binaries. Sharemind HI Solution is audited against malicious behaviour. See also DS4-I <u>Residual risks:</u> Sharemind HI host (MNO-VAD) or attacker with access to Sharemind HI server can deduce something about processed data from side-channel attacks. This can be partly mitigated by special constructs in enclave source code. This may come with performance penalties and is not applied in the PoC.
p4	D	Attacker makes (parts of) the Sharemind HI solution unavailable.	M6 – Limit the network visibility. Protect process components from external parties.

			M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This process is internal to MNO-VAD and thus we assume it to be covered my infrastructure management. <u>Residual risks:</u> Residual risks are accepted. For pseudonymisation, there is a fallback mechanism where MNO-ND itself generates a random periodic pseudonymisation key. The analysis process is usually not time critical.
p4	E	Attacker gains elevated privileges in the Sharemind HI solution.	M1 – Authentication and authorisation procedures for users. M12 – Cryptographically strong design. Use of Intel SGX technology together with Sharemind HI offers cryptographically enforced access control for processes (task enclaves) as well as for data inside the Sharemind HI solution (ds4).
ds4	T	Attacker modifies task enclaves' code.	M12 – Cryptographically strong design. Use of Intel SGX technology together with Sharemind HI role-based access control means that task enclaves can only be modified with the consent of all Enforcers. <u>Residual risks:</u> Attacker persuades/bribes all Enforcers.
ds4	T	Attacker modifies Sharemind HI solution's internal data (generated periodic pseudonymisation keys, analysis request, aggregated data, reverse-pseudonymised mobile location data, audit log)	M12 – Cryptographically strong design. Sharemind HI stored data uses authenticated encryption with integrity checks. Modifying it outside of Sharemind HI raises alarms. Sensitive data is not readily available even to authenticated users.
ds4	T	Attacker modifies Sharemind HI solution keys that are	M12 – Cryptographically strong design. Sharemind HI stored data uses authenticated encryption with

		used to protect the enclave state	integrity checks. Modifying it outside of Sharemind HI raises alarms. The encryption key for encrypting the state is derived from Intel SGX hardware.
ds4	R	User claims she has not performed an operation stated in the audit log.	M2 –Authenticated logging in Sharemind HI server. M12 – Cryptographically strong design. All user operations are authenticated and only Sharemind HI solution itself can add records to the audit log.
ds4	I	Sharemind HI internal data leaks.	M12 – Cryptographically strong design. Sharemind HI stored data uses authenticated encryption with the encryption key only known to the Sharemind HI solution itself and no authenticated users. Sensitive data is not readily available even to authenticated users (unless made available by design). M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). Although all relevant data is encrypted on disk, Sharemind HI server host should take reasonable action to prevent unauthorised access to the disk.
ds4	I	Sharemind HI solution keys that are used to protect the enclave state leak	M12 – Cryptographically strong design. The analytics enclave generates a new encryption key for accumulated subscriber footprints data (the S file) each time its contents change. Therefore, this key lifetime is very short (usually a day). These generated keys themselves are in turn encrypted with a key derived from Intel SGX hardware. The latter cannot be extracted from the CPU.
ds4	D	Sharemind HI solution code or internal data is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This data is hosted on MNO-VAD infrastructure and thus we assume it to be covered my infrastructure management.

			Residual risks: a) Attacker deletes periodic pseudonymisation keys or task enclaves' states and renders all pseudonymised data collected unusable. Sharemind HI deployment has to be reset and all necessary mobile location data has to be re-pseudonymised by MNO-ND. Mitigation. It is possible to create a secure export and import mechanism for periodic pseudonymisation keys to provide backup. This mechanism could use (threshold) secret sharing scheme to avoid the keys being used outside of the enclave. b) If attacker deletes aggregated data (the S file or its encryption keys) or analysis report, these can be recomputed with less effort. Mitigation. The encryption keys could be backed up by a secure export-import system described above. c) Attacker deletes the audit log. All these residual risks can be mitigated by regular backups that the attacker cannot modify. M5 – Avoid overlapping roles
df2	T	Attacker modifies data moving between MNO-VAD and Sharemind HI Solution.	M13 – Secure communication channels. All communication between MNO-VAD and Sharemind HI Solution are transported over a mutually authenticated channel (TLS).
df2	I	Data moving between MNO-VAD and Sharemind HI Solution leaks.	M13 – Secure communication channels. All communication between MNO-VAD and Sharemind HI Solution are transported over an encrypted secure channel (TLS).

df2	D	Attacker prevents MNO-VAD to import data or download report.	M4 – Responsibility assigned to the maintainer of the infrastructure (MNO-VAD). This risk is small as DF2 is MNO-VAD internal connection.
p3	S	Attacker creates a fake copy of the Solution client software with different functionality.	M9 – Code auditing and authenticated binaries.
p3	T	Attacker modifies the Solution client software used by NSI.	The client software has limited capabilities and user rights are enforced by the server side (Sharemind HI Solution). M1 – authentication and authorisation procedures for users. Residual risks: Attacker modifies the application after its deployment. It could lead to leaking or modifying of analysis results or input data (see I_ds3 and T_ds3).
p3	R	User claims not to have invoked: a) ordering a new analysis report with or without uploading NSI input data; or b) downloading the analysis report.	All user actions require authentication by a valid user. M1 – authentication and authorisation procedures for users. All user actions are logged on the server side. M2 – Authenticated logging in Sharemind HI server.
p3	I	The Solution client software at NSI leaks information.	M9 – Code auditing and authenticated binaries.
p3	D	The Solution client software at NSI is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI). The residual risk is accepted as the process at NSI is not time critical.
p3	E	Attacker gains higher privileges in the Solution client software at NSI.	Locally, at NSI there is a single level of privileges for the Solution client software. Client authorisation is enforced by the server side (Sharemind HI

			Solution, see E_p4). M1 – authentication and authorisation procedures for users.
i3	S	Attacker pretends to be authorised NSI user when using the Solution client software.	The Solution client software is a thin client and cannot be abused locally. Client authorisation is enforced by the server side (Sharemind HI Solution). M1 – authentication and authorisation procedures for users.
i3	R	User claims not to have a) ordered a new analysis report with or without NSI input data; or b) downloaded the analysis report.	M2 – authenticated logging in Sharemind HI server M10 – logging on the server side.
ds3	T	Attacker modifies NSI input data.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI). Residual risk: NSI input data is modified by tampering with the NSI analysis process (see T_p3).
ds3	T	Attacker modifies the analysis report.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI). Residual risk: Analysis report is modified by tampering with the NSI analysis process (see T_p3).
ds3	I	NSI input data leaks.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI). Residual risk: NSI input data leaks by tampering with the NSI analysis process (see T_p3).
ds3	I	The analysis report leaks	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI).

			Residual risk: Analysis report leaks by tampering with the NSI analysis process (see T_p3).
ds3	D	NSI input data is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI). Residual risks: NSI input data can be reassembled by NSI, but this postpones the analysis process at the Solution. This is accepted risk.
ds3	D	Analysis report is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI). The analysis report can be re-downloaded from the Sharemind HI Solution.
df3	T	Attacker modifies data moving between the Solution client software at NSI and the Sharemind HI Solution.	M13 – Secure communication channels. All communication between the Solution client software and Sharemind HI Solution are transported over mutually authenticated channel (TLS).
df3	I	Data moving between the Solution client software at NSI and the Sharemind HI Solution leaks.	M13 – Secure communication channels. All communication between the Solution client software and Sharemind HI Solution are transported over an encrypted secure channel TLS.
df3	D	The network connection between the Solution client software at NSI and the Sharemind HI Solution is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (NSI and MNO-VAD). The residual risk is accepted as the process at NSI is not time critical.

p7	S	Attacker fakes the IAS process with different functionality.	M4 – Responsibility assigned to the maintainer of the infrastructure (Intel). M12 – Cryptographically strong design. IAS certificates are pinned to the Intel CPUs with SGX capability. Residual risks: Intel loses control over its private keys. This is accepted risk.
p7	T	Attacker modifies the IAS process.	M4 – Responsibility assigned to the maintainer of the infrastructure (Intel).
p7	R	Intel claims not to have gone through the IAS process.	We assume that Intel has internal procedures and auditing in place to avoid that. M4 – Responsibility assigned to the maintainer of the infrastructure (Intel).
p7	I	IAS process leaks data.	We assume that Intel has internal procedures and auditing in place to avoid that. M4 – Responsibility assigned to the maintainer of the infrastructure (Intel). Residual risks: Intel learns, with the accuracy of an IP, who uses the remote attestation feature of the TEE. This is accepted risk.
p7	D	IAS is (made) unavailable.	This is accepted risk. We assume Intel has motivation and means to reasonably mitigate this. M4 – Responsibility assigned to the maintainer of the infrastructure (Intel).
p7	E	Attacker gains higher privileges at IAS process.	This is accepted risk. We assume that Intel has internal procedures and auditing in place to avoid that. M4 – Responsibility assigned to the maintainer of the infrastructure (Intel).

df8	T	Attacker modifies data moving between the client application and IAS.	M13 – Secure communication channels. All communication between the Sharemind HI client application and IAS is transporter over mutually authenticated channel (TLS).
df8	I	Data moving between the client application and IAS leaks.	M13 – Secure communication channels. All communication between the Sharemind HI client application and IAS is transported over an encrypted secure channel (TLS).
df8	D	The network connection between the client application and IAS is (made) unavailable.	M4 – Responsibility assigned to the maintainer of the infrastructure (Intel and the respective client application host).