Working translation from German to English

Memorandum

JENTIS GmbH

SPIRIT LEGAL®

Author

Attorney Tilman Herbrich (CIPP/E)

Date of the document

19.07.2023, v1.1

Project:

Evaluation of the basic concept "Synthetic Users" under data protection law

Executive Summary

From the consideration of the basic concept "Synthetic Users " [1]it follows that companies can use the **JENTIS Data Capture Platform** ("DCP") and the "JENTIS Essential Mode" [2]as a fallback solution to configure first-party tracking in such a way that the application of the exception regulations from the requirement for consent for end device access (e.g. § 25 para.2 TTDSG) can be reconciled with supervisory authority positions.

By applying **mathematical procedures** based on user data for which consent has been given and on reduced technically required or strictly necessary user data collected via the JENTIS Essential Mode, JENTIS enables a statistical analysis of up to 100% of the usage behaviour of digital offerings in a reduced scope **without querying for consent.**

Basis of evaluation

1. Impact of industrial developments and consent rate on analytical methods

The measurability of the usage behaviour of digital offerings averages 30% of the target audience. Given an average consent rate and the negative impact of **ad-blockers** and browser **tracking-blocking systems** by default such as Safari (Intelligent Tracking prevention "ITP") and Firefox (Enhanced Tracking Prevention , "ETP") on measuring usage behaviour using third-party and first-party tracking, the importance of reliable measurement results is growing. Since Safari (v16+) also blocks third-party server-side cookies in certain cases that are set through a first-party domain, the inability to analyse the usage behaviour of digital offerings becomes more acute.

2. Generation of Synthetic Users with JENTIS Synth - User Engine & Server Side Tracking

On average, up to 70% of user behaviour cannot be measured due to a lack of consent or blocking of user browsers, so that only the data of 30% of users is actually used as a basis for decision-making by the company. JENTIS enables the generation of synthetic users for valid pseudonymisation of usage data when JENTIS Essential Mode is activated . [3]For this purpose, instead of the 70% real, but missing users for analysis, synthetic users, i.e., artificial users without personal or end device reference, are generated in order to measure these together with 30% of the real users and thus ensure 100% visibility in the analysis tool.

[1]Basic concept "Creation of synthetic users as a result of the combined processing of consented and stricly necessary data within the meaning of ePrivacy and the GDPR for the purpose of online marketing .", JENTIS GmbH, April 2023 (see appendix).
[2]Information on the JENTIS Essential Mode can be found here .
[3] *Ibid.*

The **JENTIS Synth User Engine** will regenerate the missing data in the non-consent user area using historical data and live tracking data (both from consent and non-consent users) using the statistical method from mathematics "imputation". In this way, incomplete data sets can be completed through functional relationships. After applying the statistical method of imputation, all **raw data** from **users without consent is completely deleted** using JENTIS technology .

For the **generation of synthetic users,** the artificial data sets are combined with the data sets of the real consent users to generate new behavioural data from non-real, artificial users. The data set, consisting of real and synthetic data, can be passed on to analysis or marketing tools for analytical purposes via the **JENTIS server-side tracking technology as a technical pre-filter.**

Legal Evaluation

In European data protection law, the **concept** of **data synthesis**[4] has become firmly established as a **Privacy Enhancement Technology** (PET). [5]According to the European Data Protection Supervisor (**EDPS**) and the European Union Agency for Cybersecurity (**ENISA**), synthetic data represent a PET and in this sense an **additional protective measure** for **data transfers** - also to third countries without an adequate level of data protection.[6]

1. Definition of Synthetic Data

Synthetic data is also called "fake data " or "artificial data". Regardless of the terminology, synthetic data are fundamentally artificially generated data, produced from the original data and preserving the statistical properties of these original data, without however having any reference to an identified or identifiable person. [7]

2. Classification of synthetic data under data protection law

Due to the risks of re-identification of affected individuals, [8]**synthesised data** generated from real data subjects is generally not classified as anonymous data, but pseudonymous data. In particular, when synthetic data has sufficient structural equivalence with the original data set or share essential properties or patterns that allow assignment to real users, pseudonymisation within the meaning of Art. 4 No. 5 GDPR is assumed.[9]

ENISA refers to various studies, according to which even with a completely anonymised data set **of** traffic data (mobility data) three to four known data points are sufficient to carry out a re-identification.[10]

The **Norwegian Data Protection Authority** issued a fine in June 2021 against an association due to an accidental publication of user data due to a lack of a test procedure for a cloud solution. In the reasoning , the

[4]On the development and the methods, see EU Commission , JRC Technical Report, 2022, p. 12 ff.
[5] EDPS, techsonar 2021-2022, p. 10 ; ENISA, Data Protection Engineering, 2022, p. 17 .
[6] EDPS, techsonar 2021-2022, p . Hintze/ Emam , Can synthetic data help organizations respond to ' Schrems II' ?,
[7] EDPS, techsonar 2021-2022, p . López/ Elbil , European Law Blog, On synthetic data: a brief introduction, 2022 .
[8] EDPS, technographic 2021-2022, p .
[9] López/ Elbil , European Law Blog, On synthetic data: a brief introduction, 2022 ; cf. on the risks of re- identification in general Art. 29-Data Protection Working Party, WP 216, Opinion 5/2014 on Anonymisation Techniques, p. 11 and on synthetic data Stadler/ Oprisanu / Troncoso , Synthetic Data – Anonymization Groundhog Day, 2022, p. 4 ff.
[10]See ENISA, Data Protection Engineering, 2022, p. 10 ; see, for example , Gieselmann , How common methods for anonymising data fail, 2019 .

authority points out that the use of synthetic data could have prevented the data protection incident and that its use is therefore always recommended.[11]

3. Data protection assessment of the generation of synthetic users

(1) The generation of synthetic users as described in the JENTIS basic concept[12] " Synthetic Users" can be classified as a robust measure of pseudonymisation. Due to the derivation of behavioural data from real users who have given their consent and the deletion of raw data of users who have not given their consent, the process of **generating synthetic users** can only be classified as **pseudonymisation** according to Art. 4 No. 5 GDPR.

The **synthesis of real raw data** such as the Client-ID or User-ID assigned by third parties is to be classified **under the same conditions** as the **formation** of **hash values** from real raw data as pseudonymisation according to Art. 4 No. 5 GDPR. As long as the [13]**artificial values used** in place of Client-IDs and User-IDs are **irreversible** for third parties, the **collision resistance** of the processed data parameters is **ensured and** the user's **IP address** has been **replaced** by artificial values, a **GDPR-compliant pseudonymisation** is assumed considering the unanimous assessment of hash values in the absence of contrary case law.

(2) In the process of generating synthetic users, there is **no further direct access** to **or immediate storage** of the information in **users' terminal device resources.** . Therefore, this process, as a subsequent processing phase, falls exclusively under the regulations of the GDPR, in accordance with the **ECJ** decision on the **one - stop -shop procedure [14]**, the **legal justification** for the TTDSG [15]and the view of the **EDPB**[16].

(3) As far as user data from consent users is collected and used for the generation of synthetic users, both processing operations are based on the user's consent. The data of consent users is measured as accurately as possible and the data set is used to determine predictors that are unique in order to extrapolate consenting users. The predictors determine the data parameters that are collected from the non-consent users. Corresponding raw data is completely deleted before the data is passed on to other systems.

Data collected from **non-consent users** can be qualified as excluded from the consent on the basis of the **exemption** under Art. 5 para. 3 s. 2 ePrivacy Directive [17]and §25 para.2 TTDSG[18]provided this data (also

---

[11] Norwegian Data Protection Authority (Datatilsynet), Press release of 15 June 2021 .

[12]Basic concept "Creation of synthetic users as a result of the combined processing of consented and strictly necessary data within the meaning of ePrivacy and the GDPR for the purpose of online marketing .", JENTIS GmbH, April 2023 (see appendix).

[13] See on hashing as a valid measure for pseudonymisation Schwartmann/Weiß, Draft for a Code of Conduct on the use of GDPR compliant pseudonymisation, 2019, v1.0, S. 26; ENISA, Pseudonymisation techniques and best practices, 2019, S. 33; ENISA, Data Pseudonymisation: Advanced Techniques & Use Cases, 2021, S. 12; Artikel 29-Data Protection Working Party, WP 216, Opinion 05/2014 in Anonymisation Techniques, S. 20.

[14] ECJ , ruling , v. 2021-06-15 - C-645/19 - One-Stop-Shop, paragraph 74 .

[15] BT Drs . 19/27441, p. 38 .

[16] EDPB, Opinion 5/2019 on the interaction between the ePrivacy Directive and the GDPR , p. 23.

[17]Art. 5 para. 3 sentence 2 ePrivacy Directive: "This shall not prevent any technical storage or access for the sole purpose of carrying out or **facilitating** the transmission of a communication over an electronic communications network, or as strictly necessary in order to provide an information society service explicitly requested by the subscriber or user.""

[18] Section 25 (2) TTDSG: "(2) Consent according to paragraph 1 is not required,

1. if the sole purpose of storing information in the end-user's terminal equipment or the sole purpose of accessing information already stored in the end-user's terminal equipment is to carry out the transmission of a message over a public telecommunications network, or

referred to as "predictors") is not persistent and is limited to what is technically required or strictly necessary to upscale users with consent (through imputation). [19]

In the context of JENTIS technology for synthetic users, a **"predictor"** refers to a technical (typically non-personal) **metric** that the algorithm requires in order to bundle ("cluster") users before imputing the data of non-consenting users with the data of consenting users **to create the synthetic user.**

Examples of a **predictor** used for clustering users include:

- the browser being used;
- session duration;
- start time of the session;
- number of page views;
- whether a specific product is in the shopping cart;
- scroll rate on the 1st page.

Predictors are automatically extracted by the mathematical model based on the data processed from users who haven given explicit consent. The qualitative test is applied to determine the predictors that would allow accurate cluster formation. There is a technical capability to exclude certain data points from being classified as predictors.

The transient **processing** of **predictors** for the creation of groups consisting of synthetic users to ensure reach measurement for the flawless delivery of digital offerings such as websites can be **supported by exemption** in § 25 (2) No. 2 TTDSG. As part of the necessity test for the resilience of § 25 (2) No.2 TTDSG in accordance with the EDPS "Necessity Toolkit"[20], the criteria of the Data Protection Conference (DSK) from the guidance of telemedia providers[21] were fully taken into account.

Key criteria for determining the **strict necessity** according to the DSK (Data Protection Conference) are:

- **Time of storage**: The reading of predictor data starts from the moment the user accesses the website.
- **Duration of storage**: The duration of storage of predictor data is not persistent and only lasts for the duration of imputation and the formation of clusters for the non-consent users. The exact period depends on the website's user traffic. The period can range from a few minutes to several hours. Predictor data is read from the end device and stored until synthetic users are created based on the clusters determined by the predictors. Following this process, all raw data of non-consent users is deleted.
- **Data content**: Predictor data consists of a very limited set of technical metrics, carefully filtered by an algorithm. Predictors are automatically extracted based on the data processed from users who have given explicit consent. These predictors are used to allow as accurate clustering as possible. There's a

2. if the storage of information in the end user's terminal equipment or the access to information already stored in the end user's terminal equipment is strictly necessary in order for the provider of a telemedia service to provide a telemedia service explicitly requested by the user."

[19] *See* : imputation .

[20] Assessment of the necessity of measures that restrict the fundamental right to the protection of personal data: A Toolkit, April 11, 2017

[21] DSK, Orientierungshilfe für Anbieter:innen von Telemedien, v1.1, p. 27 - p. 30.

technical capability to exclude certain data points from being classified as predictors. Predictors can, for example, include the following non-personal end device information: session duration, start time of the session, number of page views, whether a specific product is in the cart, and scroll rate on the 1st page.

- **Readability of the information**: The data is only read by the telemedia service provider (1st Party) for a very limited period and then permanently deleted. No third-party providers are used for creating synthetic users. Third parties never have access to the user's end devices at any time.

The reach measurement for the demand-oriented and error-free presentation of digital offerings such as websites or apps requires a usage analysis based on the entire target audience and is explicitly mentioned by the DSK as an example of permissible reach measurement (Guidance for telemedia providers (v 1.1, p. 28 f.)).

(4) Due to the performed pseudonymisation, the legal basis according to Art. 6 para. 1 s. 1 lit. f) GDPR (legitimate interest) can be reasonably applied for the transfer of the synthesised data to third parties during server-side tracking.

(5) The following measures are recommended for the lawful use of synthetic users:

- Configuration of the JENTIS DCP when JENTIS Essential Mode is activated by customers;
- Documentation of the balancing of interests according to Art.6 para. 1 s 1 lit. f) GDPR [22] by a so-called LIA (Legitimate Interest Assessment).
- Adaptation of consent texts in CMP regarding the use for the creation of synthetic users;
- Transparent information in the privacy policy on the website;
- Conclusion and documentation of a data processing agreement with JENTIS.

[22] EDPB, Guidelines on Transparency under Regulation 2016/679, WP 260, rev.01, Annex .