

Omar Ali Fdal  
Statice GmbH  
Eisenacher Str. 1,  
10777 Berlin  
Germany

Paris, February 13<sup>th</sup>, 2023

N/Réf.: AGE/BPS/CS231015

Dear Mr. Ali Fdal,

You have requested an opinion from the Commission nationale de l'informatique et des libertés (CNIL) concerning the work carried out by the company Statice on the evaluation of the residual risks of re-identification for synthetic datasets. The set of methods evaluated is to be published under the name *Anonymeter* in the upcoming Privacy Enhancing Technologies Symposium (PETs), which will take place in Lausanne in 2023.

First, the technology experts department of the Commission welcomes the approach consisting of a practical privacy risk assessment based on the *singling out*, *linkability* and *inference* criteria defined in the opinion 05/2014 on Anonymisation Techniques of the WP29. After analyzing the provided pieces of documentation, we have not identified any reason suggesting that the proposed set of methods could not allow to effectively evaluate the extent to which the aforementioned three criteria are fulfilled or not in the context of production and use of synthetic datasets.

Second, the technology experts department would like to commend the approach taken by Statice of going through the peer review process and academic publication of its work, as well as the publication of the corresponding source code. Indeed, the independent evaluation, wide dissemination and continuous improvement of the presented tools are the best way to ensure their robustness and adoption.

Third, while *Anonymeter* appears to be an efficient tool for the evaluation of the residual risks of re-identification for synthetic datasets, the technology experts department would like to point out that other indicators found in the scientific literature could usefully be associated to it. As an example, privacy risk assessment performed on well-chosen examples (such as outliers) could turn out to be complementary to the global assessment methods proposed by Statice.

Furthermore, the services of the Commission would like to draw your attention to several key points:

- The generation of a synthetic dataset based on personal data constitutes a processing of personal data. It is therefore necessary to ensure that it complies with the regulations in force on data protection, namely the General Data Protection Regulation (GDPR) in the European Union.
- The anonymity of a synthetic dataset can only be determined on a case-by-case basis, i.e. for each generated dataset, and should therefore not be assumed from analyses performed on other datasets coming from the same provider or data synthesis tool.
- The results produced by the tool *Anonymeter* should be used by the data controller to decide whether the residual risks of re-identification are acceptable or not, and whether the dataset could be considered anonymous. However, as a solution provider, the company Statice should also

provide practical elements describing precisely how to use the tool and interpret the results obtained (such as examples, tutorials, thresholds, etc.).

- The scientific fields of data anonymization and re-identification are in perpetual evolution. Existing methodologies should therefore be regularly reassessed to take into account the latest developments.

In conclusion, the technology experts department believes *Anonymeter* is a valuable tool, relevant in the context of personal data protection. They encourage further testing in various contexts by researchers and institutions, in order to confirm its usefulness and reliability, and to refine the assessment metrics and acceptable thresholds.

Yours sincerely,

Bertrand Pailhès,  
Director of technology and innovation