

Deliberation 2022-030 of March 10, 2022 Commission Nationale de l'Informatique et des Libertés Nature of the deliberation: Opinion Legal status: In force Date of publication on Légifrance: Tuesday April 26, 2022 NOR: CNIX2211305V Deliberation n° 2022-030 of March 10, 2022 providing an opinion on a draft decree on the procedures for the implementation by the digital regulation expertise center of computerized and automated processing allowing the collection of publicly accessible data made available to the public by platform operators (request for opinion no. 22000888) The National Commission for Computing and Liberties, Seizure by the Ministry of the Economy, Finance and Relaunch of a request for an opinion concerning a draft decree on the terms of implementation by the center of expertise the digital regulation of computerized and automated processing allowing the collection of publicly accessible data made available to the public by the operators of e platform, Having regard to Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals 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, hereinafter "GDPR"); Considering the law n° 78-17 of January 6, 1978 modified relating to data processing, files and freedoms; Having regard to law n° 2021-1382 of October 25, 2021 relating to the regulation and protection of access to cultural works in the digital era, in particular its article 36; Having regard to Decree No. 2020-1102 of August 31, 2020 creating a service with national jurisdiction called "Digital Regulation Expertise Center" (PEReN); After hearing the report of Mr. Bertrand du MARAIS, commissioner, and the observations of Mr Benjamin TOUZANNE, Government Commissioner, WHEREAS THE FOLLOWING CONTEXTUAL ELEMENTS ARE RECALLED: ° 2020-1102 of August 31, 2020 and placed under the joint authority of the ministers responsible for the economy, culture and digital technology. PEReN's missions include: providing support to independent administrative authorities and independent public authorities involved in the regulation of digital platforms. In this case, the intervention framework is defined by agreements; to provide its expertise in the context of research work commissioned by these services, by carrying out exploratory or scientific studies. Since its establishment, the PEReN points out difficulties, in the exercise of its activities, related to the framework for collecting publicly accessible data useful to its work and the refusal of certain digital platforms to cooperate with it. He was thus denied by a platform operator access to his programming interface (API) made available to researchers, while the general conditions of use (CGU) of the services of digital platforms explicitly prohibit the automated collection of data outside of such APIs ("web harvesting", or "webscrapping"). Consequently, the legislator authorized, by article 36 of law n° 2021-1382 of October 25, 2021, this service to collect data publicly accessible online without the operators of the platforms being able to oppose a

refusal to it access to existing services for making this data available or a ban on automated collection through the T&Cs.

independent administrative authorities and independent public authorities authorized to use the expertise of PEReN within the framework of agreements (article 1 of the draft decree); and on the other hand, to regulate automated processing allowing the collection of publicly accessible data on online platforms. This processing will take place within the framework of experiments, that is to say in this case a research program, aimed at designing or evaluating technical tools intended to help the public authorities benefit from it, in their mission of regulating operators online platforms. While the Commission has no comments to make on Article 1 of the draft decree, Articles 2 et seq. concerning the procedures for implementing the experiments call for the following comments.

MAKES THE FOLLOWING COMMENTS ON THE DRAFT DECREE: Concerning the general economy of the processing of personal data envisaged On the purpose and operation of the experiments As a preliminary point, it seems necessary to note that the experimental activities of the PEReN involve the implementation not not just one but several automated processing of personal data, each experiment having its own characteristics. The draft decree aims to define a regulatory framework for several processing operations with common characteristics. The purposes of this processing, which must fall within the purpose specified by the draft decree, will be defined later in the context of the data protection impact analyzes (AIPD) which must be carried out for each experiment, as well as as provided for in Article 7 of the draft. With regard to these elements, the Commission takes note of the Government's commitment to modify the second sentence of Article 5 of the draft decree as follows: "These processing operations make it possible to carry out experimental activities aimed at designing or evaluating technical tools intended for the regulation of the operators mentioned in I of Article L. 111-7 of the Consumer Code or having as their object the reflection on this regulation ". Each experiment will be broken down into several phases. Initially, under the provisions of the fifth paragraph of article 36 of the law of October 25, 2021, the PEReN may, on its own initiative, carry out an experiment, including within the framework of the expression of a need formulated by an administrative authority. Prior to the collection, the PEReN must send a notification to the online platform operator concerned. The latter has a period of six weeks to formulate its observations on the preservation of the security of its services and, if necessary, to provide the information necessary for the use of its data provision interfaces (or API , for "Application Programming Interface"). A phase for determining the technical criteria should make it possible to guide the collection operation by selecting the content necessary for the objectives of the tool that PEReN will be responsible for designing. Data collection will then be activated on all platforms previously targeted and notified, in particular by techniques for extracting

content from sites, via scripts or automated programs or the use of APIs. The data collected will then be subject to automated processing in order to separate the information to be kept (those which correspond to the categories of data forming the previously defined data model) from those which must be immediately deleted (i.e. data not necessary for purpose and sensitive data). The data stored at the end of this operation will then be used for the previously defined purpose. With regard to these elements, whatever the end use of the tool intended to be designed or evaluated, the Commission considers that the

The experiments carried out by the PEReN will fall under the GDPR when their sole purpose is to develop technical tools with a view to improving reflection on the regulation of online platform operators. The legal basis of the data processing will be the execution of its mission of public interest, in accordance with article 6.1.e of the GDPR. In accordance with Article 36 of Law No. 2021-1382 of October 25, 2021, the processing will be under the responsibility of PEReN, placed under the authority of its head of department. On the legitimacy of the experiments and the guarantees provided for by the draft decree

The mere fact that the data is accessible on the Internet does not authorize anyone to collect and use it for any purpose, the lawfulness of each processing operation having to be assessed on a case-by-case basis. In addition, the nature of such processing is likely to infringe certain fundamental rights and freedoms, including freedom of expression and freedom of opinion; . Such attacks can only be accepted if sufficient safeguards are provided. Thus, particular attention must be paid to the principle of transparency as well as to the principles of data minimization and data protection by design. In this case, the experiments are justified under Article 36 of Law no. ° 2021-1382 of October 25, 2021 cited above, for the legitimate objective of strengthening the means of State services intervening in the field of regulation of operators of digital platforms, and which come up against the refusal of access to existing services provision of this data or a ban on automated collection through the contracts that bind these operators to their users. PEReN underlined during the exchanges that some of the experiments will have their own interest, independently of a need expressed by a State service or a project to create a specific tool. However, with regard to experiments which will have the specific purpose of designing a specific tool for a public authority, the PEReN and the beneficiary authorities must be particularly vigilant to ensure that their use may itself be lawful within the framework of the procedures and missions of the authorities, and this with constant law or by changing the legal framework. In particular, if the tools made available to the beneficiary authorities lead them to carry out an automatic analysis of categories of publicly accessible data, it will be necessary to question the legality of this method of investigation or study, which involves often massive processing of personal data, particularly with regard to the constitutional requirement of proportionality of invasions of

privacy. The draft decree already provides for several guarantees that the Commission deems essential. the draft recalls that, in accordance with article 36 of the law, only data publicly accessible on the online platforms of the operators concerned may be collected. The Commission notes that, according to the clarifications provided by the PEReN, the reference to "publicly accessible content" excludes content subject to restricted access, in particular when it requires a prior access request or connection validation step (for example, a closed group on a social network or an online messaging service). Furthermore, these data must be destroyed at the end of the experimental work, at the latest, at the end of a period of nine months after their collection. Moreover, if the draft decree authorizes the creation of dedicated accounts on the platforms concerned, it specifies that these cannot be used by PEReN agents to enter into contact with the persons concerned. In addition, it prohibits any reuse of data for purposes other than those provided for in the context of the experiment that justified the collection and excludes the transmission of data to third parties other than a service subcontractor. Finally, in order to ensure transparency on the data processing implemented in the context of the experiments carried out, the draft decree specifies the content of the annual report which must be sent to Parliament and the Commission. Given these guarantees , and subject to the following observations, the Commission considers that the processing envisaged is proportionate.

Concerning the provisions of the draft decree

On the procedures for collecting and using data

As a preliminary point, the Commission notes that no processing of personal data will be implemented during the phase of determining the technical criteria of each experimental activity. Firstly, the Commission draws the Government's attention to the need to put in place safeguards appropriate to ensure that all personal data publicly accessible on online platforms is not collected and thus ensure the proportionality of the collection.

First of all, the Commission considers that the phase for determining the technical criteria must make it possible to target the collection by ensuring that only information whose type and nature is proportionate to the purposes previously defined in the context of the DPIA are collected. Then, in accordance with the clarifications provided by the Ministry, the Commission invites the Ministry to complete the draft decree in order to specify that the collection will only be carried out on samples of random data and low amplitude; this amplitude will have to be documented within the framework of each AIPD. the collection operation triggered on the basis of previously determined technical criteria may, in practice, lead to the collection of data not necessary for the previously defined purposes as well as special categories of personal data which could constitute sensitive data (political opinions, religious or philosophical convictions, state of health or sexual orientation, etc.). 6° and 7° of article 7 of the draft decree provide that this data will be immediately and automatically deleted. their collection, to proceed with the immediate

and automated deletion of data considered irrelevant for the exploitation phase, in the absence of technical processes allowing sufficient sorting prior to collection. The methods available to PEReN to identify irrelevant data downstream of collection, such as the selection of content by category (posts, comments, etc.) do not allow for exhaustive sorting. Consequently, the Commission interprets the provisions of the draft decree as implying that the PEReN must take care to delete any irrelevant data as soon as it is identified as such in the context of the operation. sensitive and infringement data, their sensitivity makes it necessary to implement additional safeguards, upstream of the collection operation. by nature, because of the purposes pursued, to the collection of such data. The Commission also recommends limiting the risk of collecting such data by excluding searches that may lead to their collection (when selecting keywords, for example). On the categories of data and technologies used Article 6 of the draft decree provides for the exclusion of any use of a facial recognition system during the selection phase of the content to be collected, in accordance with the provisions of Article 36 of the law n° 2021-1382 of October 25, 2021. Other biometric recognition systems may however present some of the known flaws relating to facial recognition systems, such as the presence of biases due to a lack of representativeness of the training data which can entail risks of discrimination for individuals. The Commission therefore recommends that particular vigilance be given to experiments dealing with biometric data such as the voice template or any other biometric template of persons. In any case, no processing of biometric data for the purpose of uniquely identifying a natural person should be implemented. In addition, PEReN indicated that algorithms for identifying physical traits whose purpose is not no personal identification could be used by the device. The Committee recommends that these techniques be subject to a specific analysis relating to the presence of bias and the risk of discrimination of people during the planned experiments, in particular for those intended to offer tools to the authorities empowered to use the PEReN. informing the persons concerned Article 10 of the draft decree provides that the persons concerned are informed through written announcements published on the PEReN website. The Ministry thus intends to derogate, by application of the provisions of Article 14.5.b of the GDPR, to the obligation to inform data subjects individually; such an obligation would, in fact, require disproportionate effort and would lead to the collection or retention of identification data, whereas these will not always be collected and, when they are, will be subject to rapid deletion or, at a minimum, of a pseudonymisation. The draft decree provides for appropriate measures to protect the rights, freedoms and legitimate interests of the persons concerned, in accordance with Article 14.5.b of the GDPR: general information will be delivered on the PEReN website presenting in particular the various experiments, the sources of collection and the rights of individuals related to the

protection of personal data concerning them. On this point, the Commission insists on the need to provide information that can be understood by as many people as possible in order to allow Internet users to become aware of the precise conditions under which the personal data concerning them are likely to be collected at any time. Other measures are provided for by the draft decree, such as carrying out a DPIA for each experiment or even the application, where appropriate, of pseudonymization techniques. On the scope of the exercise of the right of opposition Article 10 of the draft decree provides that the right of opposition will be exercised under the conditions provided for in Article 21 of the GDPR. The Commission considers that, to ensure its effectiveness, the right of opposition must be able to concern all experiments intended to be implemented without it being necessary, for the persons concerned, to renew the exercise of their right for each processing of personal data carried out within the framework of the experiments. Therefore, it invites the ministry to complete the draft decree to clarify the scope of the right of opposition. On the conservation of logs The 5° of article 7 of the draft decree provides for the procedures for logging traces and specifies that these information will be kept for three years. The Commission considers that this retention period is excessive in view of the planned retention period of nine months for the personal data collected in the context of the processing. Indeed, the conservation of logging traces after the deletion of the data concerned by this logging does not seem useful beyond a transitional period allowing their exploitation. The Commission therefore recommends that the traces be kept for a maximum period of one year after their logging to take account of this transitional period. In addition, the ministry indicated that no automated processing of log traces was currently in place. The Commission strongly recommends carrying out an automatic trace control, in order to detect any abnormal behavior and to generate alerts, if necessary. The other provisions of the draft decree, in particular those on retention periods, do not call for no additional comments from the Commission.

MAKES THE FOLLOWING OBSERVATIONS REGARDING THE CONDITIONS OF IMPLEMENTATION:On the identification data declared by account holdersArticle 6 of the draft decree provides that the data of identification declared by account holders are likely to be collected. According to the PEReN, this data may represent useful information for certain experiments, from which predictive variables can be derived (for illustrative purposes, the presence of numeric characters in the identifier of a social network account is an indicator of the automatic character of an account). However, the ministry specified that, when the purpose of the experiment allows it, these data must be pseudonymised. Where appropriate, the Commission recommends that pseudonymisation be carried out immediately after data collection and that cross-checking, on several online platforms, from the identifiers collected, be possible only after having conducted an analysis documented in the

DPIA in order to balance the interest of this overlapping against the risks for the rights and freedoms of the persons concerned.

On the concrete methods of implementing the exercise of rights Article 10 of the draft decree provides that all the rights enshrined in the GDPR will be exercised with the head of the PEReN. Firstly, the Commission notes that when the PEReN is unable to identify the data subject, Article 11 of the GDPR will make it possible to exclude the application of these rights without prejudice to the possibility, for the persons concerned, to provide additional information for the purpose of re-identification for the purpose of exercising their rights. Specifically of the right of opposition, the PEReN plans to create a "refusal list" which would allow data subjects to oppose the collection of their data on online platforms by providing their identification data on the platforms upstream of the experiments. The Commission considers that the implementation of such a list contributes to ensuring the effectiveness of a right of prior opposition, subject to the following observations. First of all, it recommends the implementation of technical and organizational measures to ensure compliance with the principles of data minimization and data protection by design. Thus, in order to limit the impact that the accidental disclosure or loss of such a list could have for the benefit of an attacker, it recommends that PEReN keep only the cryptographic fingerprints of the identifiers of the persons concerned, by deleting any link between the identifiers of the same person in order to prevent the overlapping of identifiers used by the same person on the platforms, as well as the immediate deletion of the raw data. Then, this information, which remains personal data, must be kept in compliance with the principle of limitation of storage. Given the fact that the experiments carried out by the PEReN are intended to be implemented indefinitely over time, it recommends offering the persons concerned the possibility of objecting for different longer or shorter periods. For example, the data subject will be able to choose from among different retention periods predetermined by the PEReN (for example, from a few months to a few years). In any event, the persons concerned must be informed of the retention period in order to ensure transparency on the processing of data implemented within the framework of the "refusal list". Finally, the Commission recalls that the right of opposition must be able to be exercised at any time and that the PEReN must provide for procedures for deleting the data stored in the database. On system security The PEReN indicated that the data collected could be accessible authorized agents outside the network of its supervisory ministry. The Commission recommends the implementation of a double authentication mechanism. The PEReN also indicated that the use of "artificial intelligence" algorithmic processing was planned. The Commission underlines that in view of the difficulties that this type of algorithms can pose in terms of explainability, their use on personal data must be the subject of a specific review in order to identify their potential limits before

deployment of the tool being tested. The Commission thus recommends, when this would not harm experimentation, that the algorithms used be published and subjected to a peer review process. access to data and those relating to the technical mastery of collection tools (the latter allowing in particular the selection of the categories of data to be collected and those to be deleted immediately after collection). The President Marie-Laure DENIS