

CNS 50/2021

Opinion in relation to the consultation made by a Department on whether to provide data on the number of students enrolled in the first year for each degree and Catalan public university, classified according to the secondary school of origin (with indication of the municipality, county and name of the center) when the number of students is less than three people may violate statistical confidentiality.

A letter from the Department is presented to the Catalan Data Protection Authority in which, before complying with resolution 720/2021, of July 30, of the Commission for Guaranteeing the Right of Access to Public Information (hereinafter GAIP) and deliver the information to the claimant, it is considered, in order to guarantee respect for student data protection, whether to provide data on the number of students enrolled in the first year for each degree and Catalan public university, classified according to the secondary school of origin (with indication of the municipality, county and name of the school) when the number of students is less than three people may violate statistical confidentiality.

Having analyzed the query, which is not accompanied by any document, in view of the current applicable regulations and in accordance with the report of the Legal Counsel, the following is ruled:

I

(...)

II

The Department explains to the consultation that the Commission for Guaranteeing the Right to Access to Information Pública asks you to provide the following information to a person:

"the ratio of the number of students enrolled in the first year of the different university degrees of the Catalan public universities, who have accessed via University Entrance Examinations (PAU) or Higher Degree Training Courses (CFGS), in the 2020 academic year -2021.

The report must include:

- 1) Number of students disaggregated by each degree of each university. The name, surname or other personal data of the students must not be included, but simply the number of students per degree at each university.
- 2) Ownership of the post-compulsory secondary education center of the student's origin: whether it was a public, chartered or private school.

3) The municipality or county of this center of origin and, if possible, its name center

In the event that there is no data available from the 2020-2021 academic year, they request the same data from the last available academic year."

On the occasion of this request, the Department asks "if providing this data classified according to the secondary school of origin with indication of the municipality, region and name of the center when the number of students is less than three people may violate confidentiality statistical, in the sense that this disaggregation may allow the immediate identification of the interested parties, or lead by its structure, content or degree of disaggregation to the indirect identification of these."

Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, relating to the protection of natural persons with regard to the processing of personal data and the free movement of such data and for the which repeals Directive 95/46/CE (hereafter RGPD) considers "personal data": all information about an identified or identifiable natural person ("the interested party"); Any person whose identity can be determined, directly or indirectly, in particular by means of an identifier, such as a number, an identification number, location data, an online identifier or one or more elements of identity, shall be considered an identifiable physical person physical, physiological, genetic, psychological, economic, cultural or social of said person; (Article 4.1 GDPR).

In this regard, recital 26 of the RGPD states that: "The principles of data protection apply to all information relating to an identified or identifiable natural person (...)
To determine whether a person is identifiable, it is necessary to take into account all the means that the controller or any other person can reasonably use to directly or indirectly identify the natural person, such as singulation. In determining whether there is a reasonable likelihood that means will be used to identify a natural person, all objective factors, such as the costs and time required for identification, must be considered, taking into account both the technology available at the time of the treatment as technological advances. (...)."

However, recital 26 also specifies that "(...) Therefore, the principles of data protection do not apply to anonymous information, that is to say to information that does not relate to an identified or identifiable natural person , nor to the data made anonymous in such a way that the interested party is not identifiable or ceases to be so. Consequently, this Regulation does not affect the treatment of this anonymous information, even for statistical or research

In this sense, it is necessary to take into account Opinion 4/2007 "on the concept of personal data", of the former Working Group 29 (hereafter GT 29) on the definition of personal data, which determines:

"For the purposes of the Directive, "anonymous data" can be defined as any information relating to a natural person that does not allow identification by the person responsible for the treatment of the data or by any other person, taking into account the set of means that can reasonably be used by the person responsible for the treatment or by any other person, to identify said person. "Anonymized data" will be, therefore

therefore, the anonymous data that previously referred to an identifiable person, whose identification is no longer possible. Recital 26 also refers to this concept when it says that "the principles of protection will not apply to those data made anonymous in such a way that it is no longer possible to identify the interested party". Once again, the decision on whether the data allows the identification of a person and on whether the information can be considered anonymous or not depends on the specific circumstances of each case, so a case-by-case analysis must be carried out, in which special attention must be paid to the extent to which it is reasonably possible to use the means to identify that person, as described in recital 26. This is particularly important in the case of statistical information, since despite that this can be provided in the form of aggregated data, the original sample may not be sufficiently wide and other data may allow the identification of physical persons.

On the other hand, article 25 of Law 23/1998, of 30 December, on statistics of Catalonia, of application to the statistical activities regulated by this Law, establishes the following in relation to statistical secrecy:

"1. Statistical secrecy protects all individualized data of a private, personal, family, economic or financial nature, used to prepare the statistics, obtained both directly from the reporting person and from administrative sources, with the exception of data that has been made public through registers public or publications not declared illegal or against which no legal proceedings have been opened.

2. By virtue of the obligation referred to in paragraph 1, the individual data of mandatory communication cannot be made public or communicated to any person or entity, not even to public administrations, except for institutions or entities that are also bound by the obligation of statistical secrecy and exclusively for the purpose of being used for statistical operations."

And article 26 of this Law provides that statistical secrecy is violated "by the communication of data from which individual information can reasonably be deduced".

In the same sense, article 13.2 of Law 12/1989, of May 9, on the public statistical function provides that "personal data referring to physical or legal persons that, or allow the identification immediate identification of the interested parties, or lead by their structure, content or degree of disaggregation to the indirect identification of the latter."

The decision on whether the data allows the identification of a person and on whether the information can be considered as anonymous or not, depends on the specific circumstances of each case (in particular the number of members of the group, the number of cases within of each group, and of the information that can be related and of the context in which we find ourselves) and is included in the risk analysis that the data controller must carry out prior to the processing of the data. Therefore, it will be the Department, when providing this information, as responsible for the treatment, in the process of anonymizing the data, who must assess the risks and the possibility of re-identification afterwards of the holders of the data taking into account the specific circumstances of each case.

In any case, in the case being analyzed, in which the data sets, when it comes to the grouping by centers, may be relatively low in some cases, it cannot be ruled out that a certain attribute may end up being known of most group members. Thus, if in a certain center only one student has enrolled in a Degree in a Catalan public university, once this student is known (and this can be done easily, apart from the student himself, for example, people from his environment more or less close) it can be easily inferred that the rest of the group members do not meet this attribute. This information will not allow you to know in which Degree the rest of the students have enrolled, but they have not enrolled in Catalan public universities. This could also be applicable to cases where the number of known persons is more than one but is small. Especially in an environment such as a group/class or educational center, where it can be relatively easy for its members to know directly by other means the identity of a certain number of students less than three who have enrolled in some Grade of a public university in Catalonia.

In this sense, the Opinion 05/2014 on anonymization techniques of WG 29 which formulates recommendations for the management of anonymization techniques taking into account the residual risk of identification inherent in them, highlights that the concept of identification does not only entail the possibility of recovering the name or address of a person, but also includes the "potential identifiability by singularization, linkage or inference", that is to say, it is not only the possibility of direct identification of the person, but also indirect.

Therefore, it cannot be ruled out that there is a certain risk that, once a person has been identified as a student of a secondary school, the attribute that he has not enrolled in a public university can end up being inferred about him Catalan

In any case, it should also be noted that this is a low risk not only from the point of view of its probability (it would only occur in centers with a low percentage of students enrolled in a Catalan public university degree) , but also in its severity, given that the information that can be inferred would not be information that is part of special categories of data, that is particularly sensitive, or that could give rise to discrimination, especially considering that the information that could be inferred would not even be that the people subject to inference have not enrolled in a university, but only that they have not enrolled in a Degree at a Catalan public university.

III

In accordance with what has been explained, in certain cases with a low number of students from a center who have enrolled in a Degree from Catalan public universities, it cannot be guaranteed that by delivering the information at the level central to maintaining confidentiality on this attribute at an individual level.

Different techniques can be applied to try to reduce the risks of re-identification:

First of all, it should be borne in mind that the resolution of the GAIP establishes that the information must be provided at the municipality and county level. The inclusion of the name of the center indicates that it should only be included "if possible". That is to say, from the point of view of data protection regulations, if the dissemination of information at center level allows re-ide

very few cases, it would be necessary to provide the information at county and municipality level. This would significantly reduce the possibilities of re-identification, which would be limited to a large extent to those municipalities that only have a single secondary school.

Beyond this technique of generalization or aggregation of results, other techniques can also be applied to reduce risks.

At the outset, one possibility would be to facilitate the information by ranks. Thus, instead of providing the exact number of students in each school, information can be provided by ranks that are relevant. For example, information can be provided indicating the rank to which each center belongs (center with between 1 and 5 students who have enrolled in a Degree at a Catalan public university, center with between 5 and 10 students, between 10 and 20, etc.).

Another possibility would be to offer the information in percentages. In this case, it is true that if you know the exact information about the number of students in each center you can end up obtaining the absolute number of students included in the percentage, but the exact number of students in each center is not easily available to anyone.

conclusion

The regulations for the protection of personal data do not prevent the delivery of the requested information if, in cases where there is a risk that the requested information can be attributed to identifiable persons, the techniques set out in this report are applied, in particular facilitating the information by ranks.

Barcelona, November 10, 2021