Singularisation and Identification

Identification means knowing which person is meant. Singularisation means knowing what kind of person is meant. The details are more complex. How do the two concepts relate?

Philip Glass, 23. Januar 2018, http://www.datalaw.ch/singularisation-and-identification/
Citation: Philip Glass, Identification and Singularisation, www.datalaw.ch, Rz.

The two concepts

[1] According to the Message of the Federal Council regarding the latest revision of the federal data protection law, a natural person is *bestimmbar* (determinable), if she can be identified directly or by indirect means.[1] The definition is an unlucky one, as a determinable person need not be the same as an identified person. A particular person may be thought of as an abstract position (e.g. "the perpetrator", "the butler" or "the gardener"), whereas an identified person is always a concrete real person (e.g. "the perpetrator", turning out to be "the butler" disguised as "the gardener" who is person X.). In Swiss law, this distinction is drawn by differentiating between the terms of "absolute determinability" and "relative determinability" (i.e. "Bestimmbarkeit"[2]): data that can only be assigned to a certain individual using further information that is not readily and generally accessible.[3] Here it is important to keep in mind that assigned data is regarded as personal data regardless of whether the information it carries is true in regard to that certain individual.

[2] What the Message of the Federal Council does not clarify, however, is what is meant by "identifying" a person.[4] It does enumerate methods of identification, such as indirect identification (using contextual information about a person such as ID, GPS data, specific aspects of the persons physical, physiological, genetic, psychic, economic, cultural or social identity), direct identification using one piece of data such as a telephone number, house address, social security number or finger print or by matching information.[5] The take away here is that identification must not necessarily involve a persons *name*.[6]

[3] The Duden defines *identifizieren* (the German word for "to identify") as *genau wiedererkennen* (to recognise exactly). The process of **identification** serves to determine by recognition, who is meant or whom one is dealing with. The purpose of this usually is to get into (actual or legal) contact with a certain person or to enable a third party to do so. This is effectuated by *identifiers* such as an official, preregistered name or designation or internally assigned tags or labels.

---

[1] BBl 2017 6941, 7019); being a slightly more precise version of the Message regarding the first version of the federal data protection law: Botschaft zum Bundesgesetz über den Datenschutz (DSG) vom 23. März 1988, (BBl 1988 II 413, 444).

[2] BSK DSG-Blechta, Art. 3 Rz. 10 f.; Beat Rudin, in: Baeriswyl/Pärli (Hrsg.), Stämpflis Handkommentar zum DSG, Art. 3 N. 12.

[3] David Rosenthal, Handkommentar zum Datenschutzgesetz, Zürich 2008, Art. 3 N. 19 f.

[4] David Rosenthal, Personendaten ohne Identifizierbarkeit?, digma 4/2017, fn. 17.

[5] loc. cit. fn. 1.

[6] Ditto David Rosenthal, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 199 f.; regarding EU law, see Article 29 Working Party, WP 136 4/2007, p. 16; regardig German law see Dammann, in: Simitis (Hrsg.), BDSG, 7. Auflage, Baden-Baden 2011, § 3 DSG N. 22.

[4] A purpose-oriented interpretation of the legal term of identification additionally takes into account the function of identification within the context of data protection law and refines the question from that perspective. The purpose of data protection law is the protection of personality rights of the persons about whom data is processed. The legal mechanisms to further that protection are the prerequisite of consent to violations of the right of personality in civil data protection law and in the domain of public data protection law the prohibition of abuse of personal data in art. 13 paragraph 2 of the Swiss Federal Constitution as well as the right to informational self-determination.

[5] From the point of view of data protection law, the significant criteria thus seem to be that identification can form a basis from which the rights of a person can be infringed by the means of data processing. To clarify this, I would like to introduce the distinction between internal and external identification. By *internal* identification I mean the assignment of identifiers within an information system (an internal indicator). This regularly pursues the external purpose of establishing rights and obligations of contracting parties (e.g. entering a contractual agreement), of modifying such rights and obligations (e.g. terms of service) or to terminate them (e.g. termination of a contract) - or to get into actual contact with that person for other reasons;[7] to make that person *reachable*.[8] In order to realise this second step, an identification outside of the internal information system is required (such as an address or a credit card number). It locates and determines the person outside the information system concerned. I would therefore like to call this second form of identification *external identification*.

[6] The term **singularisation** has been used for some time, however it's content does not seem very clear. Recently, DAVID ROSENTHAL attempted to interpret it. He understands "singularisation" as the view that data must be considered as personal data, and privacy must be respected, even if it can't be determined whose information is embodied in the data.[9] He regards the concept as massively flawed.[10]

[7] As far as can be determined, the data protection concept of singularisation finds its original legal basis in Recital 26 of the GDPR which states as follows: "To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as *singling out*, either by the controller or by another person to identify the natural person directly or indirectly" (p. 16). Before that, the article 29 working party stated in opinion 136 4/2007: "At this point, it should be noted that, while identification through the name is the most common occurrence in practice, a name may itself not be necessary in all cases to identify an individual. This may happen when other "identifiers" are used *to single someone out*. Indeed, computerised files registering personal data usually assign a unique identifier to the persons registered, in order to avoid confusion between two persons in the file"[11] (the German language version of the opinion uses the term "singularisiert werden" in lieu of

---

[7] The same holds true for public administration.

[8] See SOLON BAROCAS/ HELEN NISSENBAUM, Big Data's End Run around Anonymity and Consent, in: Lane/Stodden/Bender/Nissenbaum, Privacy, Big Data and the Public Good - Frameworks of Engagement, Cambridge University Press 2014, p. 45 ff.; "Even when individuals are not "identifiable", they may still be "reachable", may still be comprehensibly represented in records that detail their attributes and activities, and may be subject to consequential inferences and predictions taken on that basis"; HELEN NISSENBAUM, The Meaning of Anonymity in an Information Age, The Information Society, 15:141-144, 1999; "If, in previous eras, namelessness, that is choosing not to reveal one's name, was the best means of achieving unreachability, it makes sense that namelessness would be protected. However, remaining unnamed should be understood for what it is: not as the end in itself of anonymity, but rather, the traditional means by which unreachability has been achieved".

[9] DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 198.

[10] DAVID ROSENTHAL, Der Vorentwurf für ein neues Datenschutzgesetz: Was er bedeutet, in: Jusletter 20. Februar 2017, in the Version of the supplement to the consultation response of 4 April 2017, p.7.

[11] Article 29 Working Party, WP 136 4/2007, p. 16.

"singling someone out").

[8] In the field of sociology, the concept of singularisation is being debated with different terms being applied to the same or very similar phenomenon. There seems to be unity in the broad terms though. The process of singularisation is set in relation to the ever-increasing resolving power of information technology, which in terms constitutes the threat of a "granular society".[12] The perceived goal is not primarily surveillance, but the fact that "systems observe their environment, that is, distinguish and designate phenomena there".[13] In the process, a "multiple self" emerges (from the point of view of the individual), which can be viewed, isolated and compared from different perspectives.[14] The driving force is an increasing "institutional-technological" interest and an increasing ability to visualise and automate uniqueness.[15]

[9] According to these descriptions, "singularising" can thus be understood as pointing out and producing uniqueness or assigning an internal identifier. From the information technology point of view of the data processor, this means an internal identification (see above, para. 5). Singularisation is therefore when data is merged, and the resulting records are marked with the purpose of recognising them within the information system. Without the tagging of merged records, there is no singularisation in this sense. If untagged records contain unidentified personal data, this data would be classified as anonymous personal data.

[10] In conclusion, a singularised data record is an *informational model of a person* that does not necessarily correspond to a real person but is usually assigned as a personal record or created to that end. According to the view represented here, these singularised data sets are of legal importance if they are marked in a way so as to be recognised as such by the information system. The affected persons are not (yet) externally identifiable but may still be "reachable".

The interplay between singularisation and identification

[11] Common to the terms of identification and singularisation is that they crystallise a person from a mass of people based on data handling or processing. Identification does this by using identifiers, i.e. predetermined tags or characteristics that identify a person to other people as that specific person.[16] In a legal context - such as between legal entities and administrative bodies or between contracting parties - this is predominantly the name, supplemented by other characteristics such as date of birth to authenticate uniqueness if needed (see para. 1).

[12] Frequently there is a blending of identification and singularisation: among several people of the same name, further characteristics or tags are used to designate the proper person. The "singling out" mentioned in recital 26 is thus not only an indication ("Indiz"[17]), but the most important prerequisite

---

[12] CHRISTOPH KUCKLICK, Die Granulare Gesellschaft, Wie das Digitale unsere Wirklichkeit auflöst, Berlin 2014.

[13] ANDREAS RECKWITZ, Die Gesellschaft der Singularitäten, Berlin 2017, p. 253.

[14] ANDREAS RECKWITZ, Die Gesellschaft der Singularitäten, Berlin 2017, p. 255; on "digital doubles" see UELI MÄDER, Selbstorganisation durch soziale Kooperation, digma 2/2015, p. 63; from a data legal perspective PHILIP GLASS, Die rechtsstaatliche Bearbeitung von Personendaten in der Schweiz, Zürich 2017, p. 213.

[15] ANDREAS RECKWITZ, Die Gesellschaft der Singularitäten, Berlin 2017, p. 73 f.

[16] See DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 200, who, in the context of identification, speaks of a person having to be "in some way previously known" in order to establish an identifying link to anonymous data (translation by the author).

[17] DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 202.

for (internal and external) identification: personal data always refers to a single person as a matter of necessity. The name is probably especially useful because it provides a great differentiation performance *and* is a label the wearer identifies with.

[13] Finally, the distinction can be grasped more precisely by asking the question of *what* is being singularised. While identification always pursues the goal of recognition of a particular human being, the process of singularisation refers to the *naming of unique information patterns* - that is, the description of a distinctive node of information links within a database or information network. Such a "logical place" does not necessarily have to be attributed to a certain human being or even be attributable at all. Thus, for example, within information networks, various information machines can be singularised according to "addresses", which in turn can be assigned to persons. In this case, however, the process of singularisation concerns first and foremost a machine which, in certain cases, is assumed by experience to be used by a particular human being.[18] This brings us to the risk aspect of determinability.

Determinability as risk factor

[14] As mentioned in paragraph 1, "absolute determinability" as it is understood in data protection law, means the same as "identification". Therefore, the broader concept of "relative determinability" must be assumed to mean the possibility or probability of identification. As far as personal data is concerned, the probability of identification is usually very high (ideally 100%), as the purpose of processing this data lies in the link to the person concerned. With regard to pseudonymised or anonymised personal data, however, the probability of identification should ideally be at 0%. Because the identification is not covered by the purpose of the data use in question, it represents a separate danger of infringement of civil or constitutional rights not covered by that purpose. Therefore, the probability of identification describes the risk of identification.

[15] According to the Message of the Federal Council regarding the original Data Protection Act, a person is identifiable "if, although not clearly identified by the data alone, identification can be inferred from the circumstances, that is, the context of an information" (as translated by the author). There follows an important clarification of this *legal risk*: "However, not every theoretical possibility of identification suffices for the determinability". The Message then clarifies that determinability is not to be assumed if the effort involved makes it practically worth no one's time and effort.[19] The risk of identifiability affects both the person concerned (risk of personal injury) and the data processor (risk of unlawful data processing). However, the Swiss data protection laws have to be interpreted with regard to their defined purpose of protecting the personal rights of the persons of whom personal data is processed.[20] The rights associated with that protection therefore tend to have a higher weight within the frame of legal risk assessment.

[16] In his juxtaposition of the two concepts of singularisation and identification, ROSENTHAL defines

---

[18] To illustrate this, look up *amiunique.org*; see also the reference to cookies and browser apps in DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 200.

[19] BBl 1988 II 413, 444; the restriction was taken directly from a comment on German data protection law (fn. 30); regarding the legal aspect of risk in data protection law, see PHILIP GLASS, Die rechtsstaatliche Bearbeitung von Personendaten in der Schweiz, Zürich 2017, p. 138 ff.

[20] See Art. 1 DSG-CH: "Dieses Gesetz bezweckt den Schutz der Persönlichkeit und der Grundrechte von Personen, über die Daten bearbeitet werden", which unofficially translates as "The purpose of this law is to protect the personality and fundamental rights of persons of whom data are processed".

the clarification of the legal risk of identification so that in the absence of typical identifiers (such as the name) in a data set, the data set is considered "not absolutely determinable" and therefore does not fall within the scope of data protection law. The test for defining the scope of the law should then focus on two aspects, namely whether **a)** in addition to the database (containing singularised data) the data processor is provided with a so-called reference data set (containing identifiers) with which a *match* is generated and thus a singularised data set can be identified and if **b)** the processor has an interest in the identification or in the effort involved.[21] According to the view taken here, this form of identification denotes an external identification. ROSENTHAL adds that he leaves the question open whether the reference data test is suitable in every case and that there are several points to be clarified: the period for the availability of reference data, a conceivable "half-life" of data (due to decreasing interest in the identification of the data over time) and what it means that a person is "known".[22]

[17] Of these open questions, the one concerning the property of "being known" is of particular interest: it refers to the problem of internal identification as well as the transition to external identification and thus to determinability. The latter constitutes a legal risk (see para. 15). Consequently, the two aspects of the possibility of a match and the interest in identification must not be included in the assessment as mere cumulative conditions and the existence of personal data denied if one of them is not met. Rather, they must be weighted as cumulative *risk factors* for the possibility of external identification.

[18] From a risk perspective, the identifiability of the person in the sense of data protection law is therefore to be assumed if **a)** data are marked in a way that is intended to make individual persons recognisable within an information system (internal identification) and if **b)** due to the *context space*[23] of these singularised data sets there is a (realistic) possibility that the data processor or third parties enter into an actual or legal relationship with persons concerned (external identification).

[19] Of the possible constellations, the one where the processor already has a reference data record is merely the most obvious variant. Declaring it to be a test for the processing of personal data[24] restricts the responsibility of the data processor for the protection of the personal rights of those affected to an absolute minimum. For example, the determinability of a personal data can also be assumed if the person who has made an internal identification has sufficient resources and/or technical means for external identification but is not necessarily interested to do so, or conversely, has relatively modest resources but a very high interest in identification. Further examples of risk factors that should be mentioned at this point are the availability of reference data to third parties, the interest of third parties in the singularised data and the associated risk of unlawful access to the data by third parties who have a reference database.[25]

The power of differentiation of the concept of singularisation

[20] The term singularisation is useful in that it describes records that are prepared in a manner that allows for internal identification. A corresponding real individual does not even need to actually exist,

[21] DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 202.

[22] DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 201.

[23] PHILIP GLASS, Die rechtsstaatliche Bearbeitung von Personendaten in der Schweiz, Zürich 2017, p. 125 ff.

[24] DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 200 f.; "Referenzdaten-Test", cautioning that this test may not be suitable in all cases (see above, N. 15).

[25] DAMMANN, in: Simitis (Hrsg.), BDSG, 7. Auflage Baden-Baden 2011, § 3 DSG Rz. 26.; third parties ("Dritte") may especially include employees of the data pocessor.

but may consist of the user account of an information system used by several people, such as a *guest account* on a private PC. It is also possible that the personal nature of the data is not apparent to the data processor due to encryption by third parties.

[21] In this sense, the legal concept of singularisation describes a necessary information technology tool and condition for the determination and thus identification - in the sense of a "exact recognition" - of a human being.[26] It appears suitable as an "indication" for an (external) identification as the existence of "singularised" data forms an informational point of reference with regard to a specific person - and thus the risk of external identifiers being linked to that person by the data processor or third parties.[27]

Addendum

[22] The real problem does not concern the concepts discussed here, the argument about them is merely a symptom. The real problem is that the scope of data protection law is determined by the concept of the personal data, and thereby restricted by risk in a way that makes it a plaything of the nuances of the individual case. In consequence, the law may not apply to some processing of data relating to persons - but does apply to other forms of processing of the same data. Here a legal problem is regulated at the conceptual level, which should be solved more meaningfully by a material regulation on the processing of anonymous personal data.

[23]There should be a clear distinction between the scope of the law and the rights and duties of the data processors and those persons over whom data are processed, as well as the balance of interests between them. In the private sphere, this would correspond to the personality rights of the Civil Code, which require a weighing of interests in order to assess violations of law in the absence of consent (which is specified as a possible justification for processing personal data in Art. 13 DSG). The balance of interests should be struck within the law and not determine whether it applies at all. Here I agree with ROSENTHAL when he says that legal certainty is critically undermined if the case-by-case impact of data processing has a decisive influence on whether data protection law is applicable or not.[28]

---

[26] Thus, according to the view expressed here, the example of singularised data given by ROSENTHAL, of a man who can be seen slipping on a banana in a video gone viral but who cannot be clearly recognised, is not an example of singularisation. These are anonymous personal data, as they obviously represent information from a particular person (see note 1). However, the data are not clearly assignable and apply to all persons who have the characteristics recognisable in the video (stature, clothing, possibly haircut, etc.). The example of the online newspaper, which uses cookies to produce anonymous dossiers of its readers over the years (op. Cit., P. 198), according to the view expressed here, shows a case of singularisation and internal identification. Since the newspaper tailors the offer to the respective machine(s) with cookies, the risk of external identification of the respective users is constantly being increased over the years. Personal data are finally present if and as far as the data records are de-anonymised or at least that risk is significantly increased (e.g. by selling the records to third parties that have resources, and/or a high degree of motivation to de-anonymise the data).

[27] The European Commission cites the following categories of data as examples of personal data: name and surname; private address; e-mail addresses in the form of name.surname@firm.com; identity card number; gps data; IP-address; cooky-identifier; advertising identifier of a cellular phone; data available in a hospital or doctor's office that could lead to the unambiguous identification of a natural person.

[28] DAVID ROSENTHAL, Personendaten ohne Identifizierbarkeit?, digma 4/2017, p. 200.