Swiss Legal Framework for De-identification of Health-Related Data

Date	Zurich, December 19, 2020
To	Dr. Katrin Crameri and Dr. Julia Maurer, SIB Swiss Institute of Bioinformatics
From	Luca Dal Molin and Vera Rentsch, Homburger AG

Table of Contents

I. Background

1 The SIB Swiss Institute of Bioinformatics (**SIB**) is an academic not-for-profit organization whose mission is to lead and coordinate the field of bioinformatics in Switzerland and to foster excellence in data science to support progress in biological research and health. Its data science experts join forces to advance biological and medical research and enhance health.

2 Among other things, SIB is engaged in an initiative under the leadership of the Swiss Academy of Medical Sciences that aims at building an infrastructure to enable the nationwide use and exchange of health-related data for research: the Swiss Personalized Health Network (**SPHN**). In particular, it is contemplated that hospitals and other medical institutions and professionals (collectively the **Institutions**) may share health-related data with third party researchers (the **Researchers**) through the SPHN, so that the Researchers can use such data collected by the Institutions for the Researchers' own research purposes.

3 Such health-related data contemplated to be shared for research purposes may consist of any type of information, in structured and unstructured form, and it may include, in particular, identifying information regarding specific patients and information regarding their health, clinical conditions and treatments.

4 In order to facilitate the sharing of such health-related data between Institutions and Researchers and in order to comply with applicable regulation, it is contemplated to de-identify the health-related data (i.e., to anonymize or pseudonymize it; cf. below) prior to its sharing among the Swiss research community. SIB and the Institutions desire to establish guidelines in order to ensure that such de-identification is performed in accordance with Swiss law requirements.

5 In this context, we have been asked by SIB to outline the Swiss legal framework applicable to the de-identification of health-related Data.

II. Outline of the Relevant Swiss Legal Framework

A. Introduction

6 Typically, health-related data includes data relating to identified or identifiable persons: For instance, a patient's medical record may contain the patient's name, date of birth, gender and other identifying information, and combines this with information on the patient's medical conditions and treatments. Under the Swiss Federal Act on Data Protection (the **DPA**), information relating to an identified or identifiable person constitutes personal data, as such term is defined in article 3(a) DPA (**Personal Data**). Substantially any processing of such Personal Data is governed by the DPA[1] or, as the case may be, cantonal data protection laws.[2] Thus, health-related

[1] Aricle 2(1) DPA.

[2] If cantonal public authorities or private persons acting in performance of cantonal public tasks process personal data, such processing is governed by cantonal data protection laws apply. Where we refer to cantonal data protection laws in the following, we refer to the the Information and Data Protection Act of the Canton of Zurich (the **IDPA-ZH**) by way of example, which includes a definition of personal data that is essentially the same as in the DPA (Praxiskommentar IDG-

data will typically contain Personal Data and, if so, its processing for research as well as for other purposes will, in principle, fall within the scope of the DPA.

7 In addition, the use of health-related data for the purpose of research concerning human diseases and concerning the structure and function of the human body is governed by the Human Research Act (the **HRA**). Hence, in order for the contemplated sharing of health-related data to comply with Swiss law, the provisions of the HRA and its implementing ordinances (most importantly, the Human Research Ordinance, **HRO**) have to be complied with.

8 Besides the DPA and the HRA, there are a number of other rules under Swiss law to be taken into account when processing health-related data, including, for instance, rules governing secrecy obligations applicable to medical professionals (such as article 321 and 321bis of the Swiss Penal Code). While these laws contain provisions relating to the processing of Health-Related data in general and their disclosure to third parties in particular, they do not specifically address the issue of de-identification.

9 Hence, given that this present Memorandum shall focus on the de-identification of health-related data, we will in the following focus on the requirements applicable under the DPA (cf. Section C below) and the HRA (Section D). But first, we will briefly outline the terminology used in this Memorandum (Section B).

B. Terminology

10 In this Memorandum, we will use the following terms to distinguish whether certain data is, or is not, linked to an identified or identifiable person:

- *Personal Data* (as defined above) is data that relates to an identified or identifiable person;[3] thus, the person with access to such Personal Data will be able to directly or indirectly identify the person concerned;
- *De-identified data* is data for which the link to an identified or identifiable person has been removed so that the person with access to de-identified data (but not to the source data) is, in principle, not able to identify the person concerned. De-identified data may be anonymized or pseudonymized data;
- *Anonymized data* is data for which the de-identification is, in principle, irreversible, because no key or code exists to re-link the data to an identified or identifiable person;[4]
- *Pseudonymized data* (or coded data) is data for which the de-identification is, in principle, reversible because there is a key or code to re-link the data to an identified or identifiable person.[5] In the HRA, the term "coded data" is used for pseudonymized data. Given that the

RUDIN, § 3 n. 13 ff.). The IDPA-ZH does not contain rules pertaining to de-identification that go beyond those on the federal level, which is why we only include limited references to the IDPA-ZH in the following.

[3] Cf. Article 3(a) DPA.

[4] Handkommentar DSG-ROSENTHAL, Art. 3 n. 35.

[5] Cf. Article 26 HRO.

German term for "coded data" (i.e., *"verschlüsselte Daten"*) used in the HRA is misleading as it may be misinterpreted to refer to "encrypted data", we propose to uniformly use the term "pseudonymized data" instead of "coded data" in this Memorandum. Such term is also more frequently used in literature regarding the DPA.[6]

C. Relevant Rules under the Data Protection Act

11 The DPA applies to any kind of processing of Personal Data by private or federal bodies. Cantonal data protection laws on the other hand are applicable to the processing of Personal Data by cantonal public authorities or private persons acting in performance of cantonal public tasks.

12 As mentioned above, *Personal Data* is data relating to an identified or identifiable person (article 3(a) DPA, § 3(3) IDPA-ZH). Data is deemed to relate to an *identified* person if it is linked to a specific person.[7] To assess whether or not data is deemed to relate to an *identifiable* person requires a relative approach: It is to be assessed whether the person having access to the data at issue (which may be the intended recipient or an interested third party[8]) is reasonably able, taking into account the means and data available to such person and the potential interest in identifying the data subject, to determine to which specific person the data relates.[9] It is important to note that not every theoretical possibility to identify a person is sufficient for data to be deemed relating to an identifiable person; instead, it has to be considered whether the person with access to the data would, considering general life experience, reasonably be able and willing to identify the data subject.[10]

13 *Processing* of Personal Data is broadly defined and includes any operation with Personal Data, irrespective of the means applied and the procedure, in particular the collection, storage, use, revision, disclosure, archiving or destruction (article 3(e) DPA, § 3(5) IDPA-ZH).

14 The *processing of anonymized data* does not fall within the scope of the DPA because anonymized data does not constitute Personal Data.[11] The same holds true for the *processing of pseudonymized data* by persons who do not have access to the key to re-identify the data: From their perspective, pseudonymized data does not constitute Personal Data given that they

[6] Cf. SHK HFG-Rudin, Vor Art. 32–35 n. 9 ff.

[7] Handkommentar DSG-Rosenthal, Art. 3 n. 19; BSK DSG-Blechta, Art. 3 n. 9.

[8] Cf. BSK DSG-Blechta, Art. 3 n. 11.

[9] Handkommentar DSG-Rosenthal, Art. 3 n. 20; BSK DSG-Blechta, Art. 3 n. 10.

[10] Handkommentar DSG-Rosenthal, Art. 3 n. 24; BSK DSG-Blechta, Art. 3 n. 11; Decision of the Swiss Federal Court of February 26, 2018, 4A_365/2017, E. 5.

[11] BSK DSG-Blechta, Art. 3 n. 12; Handkommentar DSG-Rosenthal, Art. 3 n. 3; Praxiskommentar IDG-Rudin, § 3 n. 18.

are, in principle, unable to link the data to an identified or identifiable person.[12] [13] In contrast, the processing of pseudonymized data by persons with access to the key to re-identify the data constitutes a processing activity within the scope of the DPA.[14]

15 The *process of de-identifying* Personal Data as such also constitutes a processing activity. Thus, it falls within the scope of the DPA, even if the subsequent processing of the de-identified data may not fall within the scope of the DPA.[15]

16 The DPA sets forth a number of general processing principles that need to be complied with when processing Personal Data. Most importantly, these include the following:

- Lawful processing: The processing of Personal Data has to comply with Swiss law (article 4(1) DPA);
- Transparency: The collection of Personal Data and especially the purpose for the processing must be evident for the data subjects (article 4(4) DPA). This requires that data subjects are informed about the processing of their Personal Data, unless they have other reasonable means of understanding how Personal Data relating to them is processed;
- Purpose limitation: The Personal Data may only be processed for the purpose (i) indicated at the time of the collection, (ii) evident from the circumstances or (iii) provided for by law (article 4(3) DPA);
- Proportionality: The processing of Personal Data must be carried out in good faith and be proportionate, *i.e.* one may only collect and process such Personal Data as is necessary to achieve a legitimate purpose. This principle also requires the controller to retain Personal Data only for as long as it is necessary with respect to the purpose of the data processing (article 4(2) DPA);
- Accuracy: Personal data must be accurate and, where necessary, kept up to date (article 5(1) DPA). Personal data that is incorrect or incomplete in view of the purpose of its processing must not be processed; and

[12] Thus, it is to be considered anonymized data from the point of view of the person without having access to the key; SHK DSG-RUDIN, Art. 3 n. 14. Cf. article 26 HRO, where it is explained that health-related personal data is considered correctly pseudonymized if it can be qualified as anonymized from the point of view of a person without access to the re-identification key. Thus, it is not sufficient for pseudonymization to replace direct identifiers; indirect identifiers must also be taken into account. The view that pseudonymized data does not constitute Personal Data is confirmed by the Federal Council Dispatch of September 15, 2017 on the new data protection act, where it is pointed out that the DPA does not apply to pseudonymized data if re-identification by third parties is impossible (Federal Gazette 2017, p. 7019) and de facto anonymization exists if data is passed on pseudonymized, but the key remains with the person passing on the data (Federal Gazette 2017, p. 7083).

[13] We may note that the GDPR builds upon a slightly broader definition of "personal data", which considers it sufficient for data to be qualified as personal data if an unknown individual can be "singled out", i.e., identified as specific and distinct from others, even if there is no possibility for linking the singled out data to a specific person. By way of example, IP addresses are usually considered Personal Data under the GDPR, even if the person processing such IP addresses has no means of identifying the real person behind such IP address, simply because the unknown individual is singled out and thus distinct from others. Thus, under the GDPR it is argued that pseudonymized data continues to be personal data (cf. recital 26 GDPR; this view has already been expressed by the article 29 working group in Opinion 05/2014 (p. 20).

[14] SHK DSG-RUDIN, Art. 3 n. 14; Handkommentar DSG-ROSENTHAL, Art. 3 n. 36.

[15] Handkommentar DSG-ROSENTHAL, Art. 3 n. 63.

— Security: Those who process Personal Data have to implement and maintain adequate technical and organizational measures to ensure data security*, i.e.* to prevent unauthorized processing of such Personal Data (article 7(1) DPA).

17 Regarding the procession of sensitive personal data, such as data relating to health, the following additional requirements must be observed:

— Express information duty: The data subject has to be informed, if sensitive personal data is collected, this also applies where the data is collected from third parties (article 14 DPA);

— Requirement to justify disclosure of sensitive personal data: The disclosure of sensitive personal data to third parties must be justified by the data subject, by an overriding private or public interest[16] or by law (article 12(2)(c) DPA); and

— Express consent: where the consent of the data subject is required, such consent must be given expressly and voluntarily on the provision of adequate information.

18 Further, the DPA privileges research activities. Article 13(2)(e) DPA specifically mentions the processing of Personal Data for purposes not relating to a specific person, in particular for the purposes of research, planning and statistics, and the publication of the results in a manner that the data subjects may not be identified, as an eligible ground for justification by overriding interests.

D. Relevant Rules under the Human Research Act

19 The HRA applies, *inter alia*, to the use of health-related personal data in the area of research concerning human diseases and the structure and function of the human body. The following definitions are relevant to the HRA:

— *Research concerning diseases and the structure and function of the human body* is defined as the method-driven search for generalizable knowledge on the causes, prevention, diagnosis, treatment and epidemiology of impairments of physical and mental health in human beings resp. on human anatomy, physiology and genetics, and non-disease-related research concerning interventions and impacts on the human body (article 3(a)–(c) HRA);

— *Health-related personal data* means information concerning the health or disease of a specific or identifiable person,[17] including genetic data (article 3(f) HRA);

[16] We may note that professional secrecy obligations under article 321 or 321bis SCP may apply and restrict disclosure in addition. In view of the scope of this present Memorandum, we do not further address these obligations. We may further note that also the IDPA-ZH imposes additional restrictions on the disclosure of sensitive personal data: disclosure without express consent of the data subject and without authorization by law is only possible in individual cases if this is indispensable to avert imminent danger to life and limb or if the necessary protection of other essential legal interests is to be given greater weight (§ 17(1) IDPA-ZH).

[17] Note that the German version of the HRA uses the same term as the German DPA: *"bestimmte oder bestimmbare Person"*, while the English convenience translations of the HRA and the DPA provided by the Federal Government differ: The HRA uses *"specific or identifiable person"*, whereas the DPA uses *"identified or identifiable person"*. The English translations are provided for information purposes only and have no legal effect in Switzerland. In the official French and Italian

- *Genetic data* means information on a person's genes, obtained by genetic testing (article 3(g) HRA);
- *Anonymized health-related data* means health related data which cannot (without disproportionate effort) be traced to a specific person (article 3(i) HRA);
- *Coded health-related data* means health-related data linked to a specific person via a code (article 3(h) HRA); as mentioned, we will in the following use the term *pseudonymized* rather than coded.

20 Specifically, the HRA addresses the further use of health-related data for research (i.e., its use for a research purpose other than the purpose for which it was originally collected; cf. article 24 HRO). The further use of health-related personal data for research is only permitted under certain conditions. The requirements vary depending on the type of data at issue, i.e., whether it is genetic data or other health-related personal data:

- *Non-genetic health-related personal data*: As a matter of principle, informed consent is required for non-genetic health-related information to be further used for research purposes (article 33(1) HRA). In case non-genetic health-related data is pseudonymized, no consent is required; instead, such pseudonymized data may be used for research purposes if the data subject has been informed in advance about the contemplated further use and has not objected (opt-out possibility) (article 33(2) HRA). No consent or information requirement is stipulated for the use of anonymized non-genetic health-related personal data.
- *Genetic data*: The use of genetic data for further research purposes is subject to informed consent (article 32(1) HRA). Such informed consent is also required in order to further use pseudonymized genetic data for research purposes (article 32(2) HRA). Further, the anonymization of biological material and genetic data is permitted only if the data subject has been informed in advance and not objected to the anonymization (opt-out possibility) (article 32(3) HRA).

21 A comparison of article 32 and 33 HRA shows that consent is not required to anonymize non-genetic health-related data: While article 32(3) HRA expressly provides that biological material and genetic data must not be anonymized except upon prior information and with the informed consent of the data subject, article 33 HRA with respect to non-genetic health related data remains silent on the issue.[18]

22 Further, the HRA delegates the power to define requirements regarding the correct and secure anonymization and pseudonymization to the Federal Council (article 35 HRA). This refers to articles 25 and 26 HRO (cf. Section III.B below).

translations, the terms differ as well (HRA: *"personne déterminée ou determinable"* resp. *"persona determinata o determinabile"* vs. DPA: *"personne identifiée ou identifiable"* resp. *"persona identificata o identificabile"*). According to the explanations of the Federal Council on the HRA, the term health-related personal data used in the HRA shall have the same meaning as health-related personal data under the DPA (Federal Gazette 2009, 8095). Thus, in our best assessment, there is no difference between the terms used in the HRA and the DPA. Cf. SHK HFG-Rudin, Art. 3 n. 43 ff.

[18] Cf. the Federal Council's explanations on the HRA, Federal Gazette 2009, p. 8122.

E. Hierarchy Between the Data Protection Act and the Human Research Act

23 The HRA is viewed as including sector-specific data protection law.[19] As *lex specialis* to the DPA or, if applicable, the cantonal data protection laws, it may supersede the DPA on a case-by-case basis: If the HRA conclusively regulates a specific issue, only the HRA applies.[20] Otherwise, where the HRA does not conclusively address an issue, the general principles stipulated in the DPA remain relevant also with regard to the processing of Personal Data in the context of research.[21] Thus, while the HRA may take precedence in specific cases, in general, the DPA and the HRA apply side-by-side.

24 The issue of the HRA potentially taking precedence over the DPA could arise for the anonymization of non-genetic health data: While the HRA stipulates that genetic data must not be anonymized without giving the data subject the opportunity to object, it remains silent on the issue with respect to non-genetic health-related personal data. Arguably, the silence of the HRA may be interpreted in such way that the anonymization (which, as a matter of fact, is a processing activity under the DPA) and subsequent re-use of anonymized non-genetic health data for research purposes is permitted even if the data subject has not been specifically informed and has not been given the opportunity to object.[22]

III. Focus on De-Identification of Health-Related Data

A. Introduction

25 The term "de-identification" is not used in Swiss law. The purpose of de-identifying data is to break the link between the data and the identified or identifiable data subject to which such data relates. Thus, de-identifying data aims at creating de-identified data from Personal Data. The term "de-identification" is used to describe the process of removing those data points that, alone or in combination, bear the reference to an identified or identifiable person. Properly done, de-identification results in de-identified, i.e., anonymized or pseudonymized data.[23]

26 In the following, we shall discuss the rules applicable under Swiss law to the process of the de-identification of health-related data.

B. Swiss Law Requirements Applicable to the De-Identification of Health-Related Data

27 Neither the DPA nor the IDPA-ZH do specifically address the process of de-identifying data and they do not stipulate any specific methods or guidelines regarding how to de-identify Personal

[19] See article 1 HRA, which provides that the HRA shall protect, *inter alia*, the personality of human beings in research. Cf. SHK HFG-Brunner, Vor Art. 56–61 n. 4; SHK DSG-Rudin, Art. 2 n. 6.

[20] SHK HFG-Brunner, Vor Art. 56–61 n. 5; Rosenthal David/Jöhri Yvonne, Handkommentar DSG, Art. 2 n. 3.

[21] SHK HFG-Brunner, Vor Art. 56–61 n. 5 and 23.

[22] Cf. SHK HFG- Rudin, Art. 33, n. 18.

[23] See Section II.B above. C.f. https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/.

Data.[24] Requirements regarding the process of de-identification can only be inferred with a view to the definition of Personal Data[25] and the purpose of the de-identification, which is to convert Personal Data into anonymized or pseudonymized data. In order to do so, the de-identification has thus to remove any personal references included in the original Personal Data in such a way that it is no longer reasonably possible to link the data to an identified or identifiable person.[26]

28 The HRA refers to the rules of the HRO regarding the anonymization and pseudonymization of health-related data:[27]

- According to article 25(1) HRO, *anonymization* of health-related data requires that all items which, when combined, would enable the data subject to be identified without disproportionate effort, must be irreversibly masked or deleted. In particular, this requires the name, address, date of birth and unique identification numbers must be masked or deleted (article 25 HRO).
- According to article 26(1) HRO, health-related data is correctly pseudonymized if, from the perspective of a person without access to the key, it constitutes anonymized data. Thus, in substance, the requirements to pseudonymize data are the same as those to anonymize data, except that a key is retained separately from the pseudonymized data.

29 As already mentioned, it is important to note that whether or not data is anonymized is determined using a relative approach: The existence of a merely theoretical risk that data may be linked to an identified person is not sufficient in order for data to be considered Personal Data (see n. 12 above). This is established under the DPA and also holds true under the HRA, which expressly defines anonymous data as *"data which cannot (without disproportionate effort) be traced to a specific person"* (article 3(i) HRA). Thus, in order for data to constitute anonymous data, it is not required that the identification of a specific person based on the data is absolutely impossible.[28] Whether or not a *"disproportionate effort"* is needed to identify the data subject is assessed on a case-by-case basis, considering all relevant circumstances at hand.[29] It may be argued that data may be considered anonymous if considerable criminal energy or the use of sophisticated technical means and skills would be necessary to restore the personal reference.[30] Further, even if (re)identification would be theoretically possible according to purely objective criteria, but the subjective interest or individual possibilities to actually do so are missing,

[24] As mentioned above, however, the process of de-identification is a processing activity that has, as such, to comply with the general processing principles. We do not address these any further here.

[25] I.e., data relating to an identified or identifiable person (article 3(a) DPA).

[26] Cf. n. 12 above.

[27] Article 35 HRA.

[28] SHK HFG-Rudin, Art. 35 n. 6 ff.

[29] Decision of the Swiss Federal Court of February 26, 2018, 4A_365/2017, E. 5.

[30] Erläuternder Bericht über die Verordnungen zum Humanforschungsgesetz (21. August 2013), p. 70.

the data is still considered sufficiently anonymous.[31] Thus, we believe it reasonable to also factor in contractual mitigations (i.e., contractually agreed prohibitions to re-identify and to attempt to re-identify data) on the one hand, and technical measures preventing third party access to data on the other hand.

30 However, scientific and technological progress increases the risk of re-identification of persons from anonymized data sets.[32] When several data sets are merged it cannot be excluded, that formerly anonymized data is re-identified.[33] These developments and associated risks have to be taken into account for the de-identification of health-related personal data, especially when choosing the technical method for de-identification.[34] In view of technological progress, certain authors go as far as to argue that the technical effort required to re-identify data is progressively no longer considered disproportionate.[35]

31 In our best assessment, when assessing the risk of re-identification, reasonably anticipated technical developments that are expected to become available to the persons with access to the data in the near future should be taken into account; in contrast, we believe that it is not required to account for speculative future developments that may not become reality within a reasonable timeline.

32 Overall, it is to be concluded that Swiss law, case law and legal doctrine provides limited specific guidance on how to de-identify health-related data. In our view:

— not every theoretical risk of re-identification is sufficient for data to be considered Personal Data, but we note that there is no definition of what a *theoretical* risk is as opposed to a *practically relevant* risk;

— in addition to the objective risk of re-identification it is further to be considered whether the person at stake, be it the intended recipient or an interested third party, has a relevant interest in re-identifying the data at issue, so that factors besides the mere technical possibility are to be factored into the assessment; and

— the technical progress may increase the risk of re-identification, which means that whether or not data is de-identified will have to be periodically re-assessed.

33 In addition, we believe it is important to note that anonymization as such is an important pillar of data protection, even if there will always remain a theoretical risk of re-identification. Despite potential residual risks of re-identification, anonymization serves to significantly reduce the risks

[31] Decision of the Swiss Federal Court of February 26, 2018, 4A_365/2017, E. 5; BGE 138 II 346, E. 6.1; BGE 136 II 508, E. 3.2.

[32] Humanforschungsgesetz (HFG): Ergebnisse der Evaluation und weiteres Vorgehen Bericht des BAG, S. 5; Cf. CICHOCKI MICHAL, Big Data und Datenschutz: Ausgewählte Aspekte, Jusletter IT, 21. Mai 2015, n. 14.

[33] See: https://www.edoeb.admin.ch/edoeb/en/home/data-protection/Internet_und_Computer/online-services/explanatory-notes-on big-data/explanatory-notes-on-big-data.html. EPINEY ASTRID, Big Data und Datenschutzrecht, Jusletter, 27. April 2020, n. 15 ff.

[34] WEBER ROLF H./OERTLY DOMINIC, Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics? Jusletter IT, 21. Mai 2015, n. 18.

[35] CICHOKI MICHAL, Big Data und Datenschutz: Ausgewählte Aspekte, Jusletter IT, 21. Mai 2015, n. 18.

associated with the processing of Personal Data. We believe it to be inappropriate to dismiss the concept of anonymization simply because of the fact that there remain residual risks of re-identification.

C. Feasibility of Risk-Based and/or Rule-Based Approach

1. Introduction

34 In view of the limited guidance available under Swiss law, it is worthwhile considering approaches taken by legislators outside of Switzerland. Thus, we will in the following briefly discuss the risk-based and rule-based de-identification procedures under the U.S. Health Insurance Portability and Accountability Act (the **HIPAA**)[36] (Section 2 below) and the relevant rules under the EU General Data Protection Regulation (the **GDPR**) (Section 3) and assess their feasibility from a Swiss law perspective (Section 4).

2. Overview: Risk-Based and Rule-Based De-Identification

35 We understand that the HIPAA in principle provides for two methods that can be used to de-identify protected health information[37]: (1) a formal determination by a qualified expert ("Expert Determination"), or (2) the removal of specified identifiers as well as absence of actual knowledge by the medical professionals that the remaining information could be used alone or in combination with other information to identify the individual ("Safe Harbor").[38] Protected health information is de-identified when it does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual.[39]

36 The *"Expert Determination"* method is based on the use of statistical methods that have alter information in such a way that individuals are no longer identifiable. The application of statistical and scientific principles and methods need to determine that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information. Further, the methods and results of the analysis that justify such determination need to be documented. The de-identification process is roughly divided into three steps:

— Evaluation of the re-identification risk of the data;

[36] As Swiss lawyers, we do not assess or advise on HIPAA or any other U.S. law from a U.S. law perspective and this Memorandum is not to be understood as such advice. We only refer to concepts under HIPAA as outlined in public sources to assess these from a Swiss law perspective.

[37] We understand that PHI is considered to include information, including demographic information, which relates to the individual's past, present, or future physical or mental health or condition, the provision of health care to the individual, or the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual. Section 262 HIPAA, resp. Section 1171(6) Social Security Act.

[38] Cf. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard.

[39] Section 164.514(a) of the HIPAA Privacy Rule.

- Determination of the applicable method to reduce the evaluated risk and application of the method;
- Evaluation of the re-identification risk based on the de-identified data; the risk should have been reduced to a "very small" risk.

37 There is no definition for when a detected risk is "very small". Therefore, the expert needs to define an acceptable "very small" risk based on the ability of an anticipated recipient to identify the data subject, which is dependent on many factors, which an expert will need to take into account while assessing the risk from a data set. This is because the risk of identification that has been determined for one particular data set in the context of a specific environment may not be appropriate for the same data set in a different environment or a different data set in the same environment.[40] Thus, the "Expert Determination" method is a case-by-case assessment, considering the data set and the specific environment at hand.

38 The *"Safe Harbor"* method on the other hand provides for a list of specific identifiers of the individual, or relatives, employers or household members of the individual that need to be removed or changed. In addition to the removal of the 18 identifiers, the medical professionals must not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information. Actual knowledge means clear and direct knowledge that the remaining information could be used, either alone or in combination with other information, to identify an individual who is a subject of the information. This is the case when medical professionals conclude that the remaining information could be used to identify the individual (e.g., the profession is none of the above-mentioned identifiers, if the profession of the individual is listed as Federal Counsel, this information in combination with additional data would lead to an identification). No distinction is made between data entered into standardized fields and information entered as free text. An identifier listed in the Safe Harbor standard must be removed regardless of its location in a record if it is recognizable as an identifier.[41]

39 Both methods recognize that even when properly applied, there is still a certain risk, that the de-identified data can be re-identified.

3. GDPR

40 The GDPR contains certain provisions regarding pseudonymized data.

- Pseudonymization under the GDPR means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (article 4(5)).

[40] Cf. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard.

[41] Further information: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard.

- Recital 26 GDPR specifies that 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. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

41 It follows that data is considered anonymized from a GDPR perspective if it is reasonably unlikely that any means are available and used to identify the individual. Thus, the approach chosen under the GDPR with regard to the anonymization of Personal Data is substantially the same as the one under Swiss law.

42 The GDPR does not provide for specific rules or technical measures that need to be taken into account for anonymization resp. pseudonymization.

43 In Finland, where the GDPR applies, more specific guidance has been issued by the Finnish Social Science Archive (the **FSD**). Similar to the U.S. "Safe Harbor" method, several identifiers were established, which may have to be removed to de-identify data. However, the approach of the U.S. "Safe Harbor" method has been taken much further, as 42 characteristics (not exhaustively) were defined, which are categorized as direct identifier, strong indirect identifier and indirect identifier. Direct identifiers contain information that is sufficient on its own to identify an individual. Strong identifiers may be used to identify an individual easily. Information that on its own is not sufficient to identify an individual but when linked to other available information, could be used to identify a person, is categorized as indirect identifiers. If the combination of different indirect identifiers or the combination of indirect identifiers with additional outside information pertains to more than one person and a single person cannot be identified with reasonable effort, the data is de-identified.

44 In addition to listing the identifiers, a recommendation is made as to which method can be used to de-identify these identifiers. However, it would be wrong to assume, based on these explanations, that Finland simply applies a more detailed "Safe Harbor" method. The list of identifiers is provided as a guide, as the number of indirect identifiers and their level of detail affect the de-identification choices. It is emphasized that the removal of identifiers is rarely sufficient to obtain de-identified data and that there is no de-identification technique that is applicable to all data types. The de-identification method used should be chosen depending on the specific case, the data used and the intended purpose.

45 As a guideline to assess the choice of de-identification technique and the robustness of the outcome, the FSD provides for the following three questions that are adapted from EU's article 29 working group (Opinion 05/2014):

- Singling out an individual: Can you still single out any individual in the data after anonymization?
- Linkability: Can you link records relating to an individual to another dataset or information from external sources and thus identify the individual?

— Inference: Can you infer that certain information concerns a specific individual? Can you infer the original values of altered or removed values?[42]

4. Swiss Law Assessment

4.1 Introduction

46 In the absence of concrete requirements for de-identification in Swiss law, we will discuss in the following to what extent the application of the "Expert Determination" method and/or the "Safe Harbor" method is possible and whether the de-identified data generated by these methods meet the Swiss law requirements regarding the de-identification of Personal Data. It must again be noted that although Swiss law does not provide for any requirements for the de-identification process, the resulting data is only successfully de-identified if the establishment of a personal link is cannot be restored without disproportionate effort.

4.2 Swiss Law Assessment of the Rule-Based Safe Harbor Approach

47 Under the "Safe Harbor" method, identifiers need to be removed or transformed even if an identification based on a certain identifier is not possible. On the other hand, possible or indirect identifiers that are not listed but might – in combination with other information – lead to an identification do not need to be removed unless the medical professionals have actual knowledge about this possibility. No analysis of the remaining risk of re-identification seems to be conducted. This leads to an uncertainty as to whether the transformation of the identifiers is sufficient to de-identify the data.[43]

48 Additionally, the 18 identifiers of the "Safe Harbor" method are criteria that we understand are considered sensitive under U.S. Law. It must therefore be considered, which of these characteristics are also considered as identifiers under Swiss Law and where an adoption is reasonable. Furthermore, it must be taken into account that for an effective de-identification, also those features, which can identify a person in the context of further information, must be considered.[44] [45]

49 Therefore, the mere application of the "Safe Harbor" method, in particular the direct application of the 18 identifiers to Swiss circumstances, does not per se lead to a result that meets the Swiss law requirements for de-identified data. While there may be scenarios where the removal of the 18 identifiers is sufficient for the data to constitute anonymized data from a Swiss law perspective, there may be other scenarios where this is not the case. Thus, while the determination

[42] See https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/. Cf. WEBER ROLF H./OERTLY DOMINIC, Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics? Jusletter IT, 21. Mai 2015, n. 18.

[43] See https://gate250.com/tc2/SafeHarborvstheStatisticalMethodWhitePaper.pdf p. 3.

[44] STÜRZER MATTHIAS/KARJOTH GÜNTHER, Werden Patietendaten anonymisiert? digma 2017, p. 176.

[45] Looking at article 30 CRO, it seems that the legislator has recognized that in certain constellations it is not enough to remove certain identifiers. However, it is questionable whether the additional requirements are sufficient to guarantee that the data will not be re-identified.

of identifiers and the identification of these identifiers within a data set can support technical procedures, the simple removal of the direct identifiers does not necessarily result in the data being re-identified only with disproportionate effort.[46]

50 The more elaborated approach of the FSD where not only direct and strong indirect identifiers but also general indirect identifiers are determined, likely gives a more differentiated output. It takes into account that, depending on additional information, also indirect identifiers can result in re-identification of a person. The FSD has also recognized that the identifiers cannot be enumerated exhaustively, but depending upon the case to be assessed, also further characteristics need to be considered as identifiers.

51 From a Swiss perspective, there is nothing to prevent the creation of a similar list of identifiers, as the HRO itself does (not exhaustively) list certain identifiers, which must be made unrecognizable or deleted to de-identify Personal Data (article 26(2) HRO). However, as pointed out, the mere removal of these identifiers does not lead to de-identified data, if it cannot be excluded that a re-identification is possible without disproportionate effort.[47] Thus, any rule-based approach will have to be combined with a risk assessment, in order to satisfy Swiss law requirements.

4.3 Swiss Law Assessment of the Risk-Based Expert Determination Approach

52 When using the "Expert Determination" method, the risk of re-identification is assessed on a case-by-case basis. The expert conducting the de-identification needs to consider the factors facilitating the re-identification by a recipient in order to determine the level of re-identification risk.[48] The risk of re-identification is finally determined on the basis of the data that has been de-identified, and since suitable de-identification techniques can be selected to mitigate these risks, a higher level of protection or better use of the data can be achieved.[49] In order to select the appropriate technique, contextual information is necessary, such as the planned use of the data, recipients of the data and possible background knowledge of the recipients.[50]

53 A risk-based approach is in our view in line with Swiss law, given that the success of the de-identification depends on the assessment of the risk of re-identification on a case-by-case basis. In the risk assessment within the framework of an expert determination, those parameters can be taken into account that are relevant to the individual case, including the technical possibilities

[46] SHK HFG-RUDIN, Art. 35 n. 6; WIRTH FELIX, JOHNS MARCO, MEUERS THIERRY, PRASSER FABIAN, Anonymisierung medizinischer Daten Innovative medizinische Forschung benötigt qualitativ hochwertige Daten. Können diese sicher anonymisiert werden?, digma 2020, pp. 75.

[47] SHK HFG-RUDIN, Art. 35 n. 6.

[48] See https://gate250.com/tc2/SafeHarborvstheStatisticalMethodWhitePaper.pdf p. 3.

[49] STÜRZER MATTHIAS/KARJOTH GÜNTHER, Werden Patientendaten anonymisiert? digma 2017, p. 177.

[50] WIRTH FELIX, JOHNS MARCO, MEUERS THIERRY, PRASSER FABIAN, Anonymisierung medizinischer Daten Innovative medizinische Forschung benötigt qualitativ hochwertige Daten. Können diese sicher anonymisiert werden?, digma 2020, pp. 75.

for re-identification. However, it should be noted that due to technological evolutions, the risk assessment is an ongoing process. Consequently, the risk assessment would have to be renewed regularly to ensure the ongoing de-identification of the data.[51]

D. Conclusion

54 Swiss law does not provide for specific methods or processes that are to be applied in order to de-identify Personal Data, including health-related data. It only defines what de-identified data is, and it does so on an abstract level: In order to assess whether data is de-identified, it needs to be taken into account whether there is a reasonable risk that a person with access to the data could and would re-identify the data, considering all relevant circumstances.

55 Therefore, the sole application of the "Safe Harbor" method does not per se result in de-identified data as that term is to be understood under Swiss law. However, the development and use of a list of identifiers to be removed from a data set in the process of its de-identification can be helpful to provide guidance as to which data in particular, but not exclusively, must be removed or modified. A reasonably flexible list of identifiers may therefore serve as a starting point to de-identify data.

56 In addition to such rule-based approach, however, a risk assessment is needed in order to provide for a de-identification meeting Swiss law requirements. Such risk assessment will have to take into account the specific context of the individual case, because whether or not information is de-identified depends on a case-by-case assessment.

* * * * *

This memorandum has been prepared solely for SIB Swiss Institute of Bioinformatics in connection with the request referred to above. As such, it may not be relied upon by any other person or for any other purpose.

We have issued this memorandum as of the date hereof and we assume no obligation to advise SIB Swiss Institute of Bioinformatics of any changes in fact or in law that are made or brought to our attention hereafter.

[51] EPINEY ASTRID, Big Data und Datenschutzrecht, Jusletter, 27. April 2020, n. 12.