www.jusletter-it.eu

Valentin Conrad

Web data collection by Swiss actors in a data protection perspective

Data scraping, data crawling or collecting data through an API are methods which enable to collect personal data online. These techniques create significant problems in terms of data protection. We provide here a short analysis of the phenomenon pursuant to the GDPR and the FADP.

Category: Articles
Region: Switzerland; EU
Field of law: Data Protection

Citation: Valentin Conrad, Web data collection by Swiss actors in a data protection perspective, in: Jusletter IT 23 May 2019

ISSN 1664-848X, jusletter-it.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77

Table of contents

A. Introduction

[Rz 1] In our practice, we noticed that Swiss researchers increasingly tend to gather information from the Web through Application Programming Interface (API), data crawling or data scraping, because the Web is an inexhaustible source of data. The marketing sector probably also uses these methods in order to acquire substantial information about future or existing customers.

[Rz 2] Data collection on the Web may cover different practices and techniques, such as data scraping, data crawling or data mining. Data scraping is the process of automatically requesting a web document and collecting information from it, whereas data crawling refers rather to the process of locating information on the Web, and indexing the content in a database (such as Google).[1] Whereas, data mining is more related to the use of algorithms in order to explore large amounts of data and to find links between those data elements.[2] Data mining is therefore the phase following the data collection.

[Rz 3] From a legal point of view, and for the purpose of this publication, we will consider all these techniques enabling data collection from the Web as a single process (hereinafter referred to as «data harvesting»), because all these processes may imply the processing of personal data and the creation of a copy of the website's content. Legally speaking, data harvesting may also be a challenge from an intellectual property, a contractual[3], a civil and criminal liability[4], or an unfair competition point of view if we look into foreign cases law[5]. However, in the scope of this article, we will restrict ourselves to briefly analyzing compliance of data harvesting in regards to lawfulness and with the obligation to inform, as well as with the concept of further use. Our

1 ProWebScraping, Web Scraping Vs Web Crawling: http://prowebscraping.com/web-scraping-vs-web-crawling (consulted on 13 March 2019); or see EduTech Wiki, Web Scraping: https://edutechwiki.unige.ch/fr/Web_scraping (consulted on 13 March 2019).

2 Swiss Federal Council Message related to the modification of the copyright law of 22 November 2017 (BBI 2018 559 p. 594).

3 Judgement of the European Union Court of Justice C-30/14 *Ryanair Ltd* of 15 January 2015.

4 United States District Court, Northern District of California, in Case No. 17-cv-03301-EMC, HIQ LABS Inc. v LinkedIn Corporation of 14 August 2017.

5 United States District Court, Northern District of California, in Case No. 17-cv-03301-EMC, HIQ LABS Inc. v LinkedIn Corporation of 14 August 2017; BGH, Judgement of 30 April 2014 – I ZR 224/12.

analysis will focus on the European Union General Data Protection Regulation, as of 28 May 2018 («GDPR»), in part B, and on the Federal Act on Data Protection, as of 19 June 1992 («FADP»), in part C.

B. EU Regulation: GDPR

[Rz 4] Before entering into the analysis itself, we should consider how Swiss entities could be subject to the GDPR when they are using data harvesting techniques in the course of their activities.

1. Territorial scope

[Rz 5] GDPR applies only if one of the two following connecting factors is fulfilled. The first criterion is the establishment of an organization in a European Union (EU) country. The establishment implies the effective and real exercise of an activity through stable arrangements (art. 3 par. 1 GDPR cum Recital 22). There must exist an inextricable link between the establishment and the data processing. For instance, if an EU-based company sells advertising, thus contributing to making the search engine owned by another foreign company economically profitable, such an advertising company will be regarded as an establishment because there is an inextricable link between the company which runs the search engine and the activities of the advertising company.[6] Regarding data harvesting activities, everything will depend on whether or not the entity in question has a seat or an establishment in an EU Member country.

[Rz 6] The second criterion is twofold. If a person does not have an establishment in the EU, the GDPR may nonetheless be applicable if one of these conditions is met: (a) the organization offers goods or services to EU data subjects, irrespectively of whether a payment is required by the data subject; or (b) the organization monitors the behavior of EU data subjects (art. 3 par. 2 GDPR). This second criterion is independent from the legal notion of citizenship, residence, or other types of legal status. The determining factor is the location of the concerned data subjects.[7]

[Rz 7] The first hypothesis (a) requires to ask whether it is apparent that the controller envisions offering services to data subjects in one or more Member States of the Union by the accumulation of sufficient evidence, such as the accessibility of a website, email address or contact details, use of language, the possibility of ordering/accessing goods or services, the use of currency, or the express mention that EU customers have the right to use the services, etc. (art. 3 par. 2 lit. a GDPR cum Recital 23). Different factors must be taken into consideration.[8] Nevertheless, for data harvesting, no particular service is usually directly rendered; therefore, this scenario is not necessarily relevant.

[Rz 8] The second hypothesis applies to the monitoring of the data subjects' behavior. First, the monitored behavior must take place in the European Union. Secondly, the specific monitoring

6 Ruth Boardman, Territorial and Material Scope of the General Data Protection Regulation, in European Data Protection, Law and Practice, IAPP Publication (USA), 2018, p. 76; Judgement of the European Union Court of Justice C-131/12 *Google Spain SL* of 13 May 2014, consid. 55 and seq.

7 European Data Protection Board, Draft Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) – Version for public consultation, Adopted on 16 November 2018, p. 13.

8 Ibidem, p. 15.

technology is irrelevant, because the tracking activities can occur through the use of different technologies; therefore, internet monitoring is not the sole situation than can trigger the application of art. 3 par. 2 GDPR. Thirdly, the proof of the data controller's intention is not necessary: it is sufficient if the data controller chooses a specific purpose, which implies behavioral analysis or profiling techniques.[9] By automatically collecting personal data, data harvesting may be a means to monitoring data subjects, but it will depend on the concrete purpose of the data controller (for example, is the data controller specifically interested in the behavior of European online users?). In our view, the latter situation may often trigger the indirect application of the GDPR for data harvesting activities.

2. Lawfulness: «regular» personal data

[Rz 9] If we assume that data harvesting involves the collection of personal data from European citizens, in the sense defined in art. 3 par. 2 of the GDPR, one tricky area is the lawfulness of processing. First, it will depend on whether the data collection applies to «regular» personal data (art. 6 GDPR) or rather to special categories of personal data (art. 9 GDPR). It should be noted that the lawfulness of the processing depends on the grounds on which the data controller relies, and not on the question of whether an infringement occurs, because the GDPR assumes that a personal data processing implies an invasion of privacy.

[Rz 10] For instance, if we consider the hypothesis in which a controller wants to study the diet plan of Twitter or Instagram users with the intention of analyzing the food preferences of a certain panel of people, it is safe to assume that he/she will try to collect information based on public posts published on these social media platforms. To rely on consent would be practically impossible, because the data controller does not have the direct contact with data subjects and the collection of thousands of statements of consent from users living all around Europe would be disproportionate. In our view, there remain two relevant grounds for processing:

i. Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller (art. 6 par. 1 lit. e GDPR); or
ii. Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party (art. 6 par. 1 lit. f GDPR).

[Rz 11] Regarding the first option, the task carried out in the public interest should be enshrined in a national law (art. 6 par. 3 GDPR cum Recitals 10 and 45 of the GDPR) and this national law should be sufficiently precise in order to ensure a lawful and fair processing (Recital 45 of the GDPR).

[Rz 12] Regarding the second option, three cumulative conditions must be demonstrated: (i) the pursuit of a legitimate interest by the data controller or by the third party to whom the data is disclosed; (ii) the processing of personal data is strictly necessary for the purposes of the legitimate interests pursued; and (iii) fundamental rights and freedoms of the person concerned by the data protection do not take precedence.[10] It is furthermore important to take into consideration the

9 Ibidem, p. 17–18.

10 Judgement of the European Union Court of Justice C-13/16 *Valsts policijas Rgas reiona prvaldes Krtbas policijas prvalde* of 4 May 2017, consid. 28.

reasonable expectations of data subjects based on their relationship with the data controller, and the concerned data subjects should reasonably expect (further) processing of their personal data (Recital 47 of the GDPR). In our view, it is questionable whether a data controller may justify its processing by claiming that personal data contained in a public post is a strong indication of the awareness and will of the data subjects. At least, each situation must be analyzed thoroughly.

[Rz 13] Incidentally, another ground may be available if the concerned Member State in the EU enables data controllers to process personal data for scientific research purposes or statistical purposes (art. 6 par. 2 cum art. 89 GDPR). To our knowledge, Member States did not really use this possibility as of now (excepting few Member States).

3. Lawfulness: special categories of data

[Rz 14] We highlighted three potential legal justifications for data harvesting involving «regular» personal data. Now, if we change the facts, and if the processing relates to the political opinions of Twitter or Instagram users, the concerned personal data will be more sensitive and the respective processing requires a special justification. As a rule, processing of political opinions is prohibited, unless an exception applies (art. 9 GDPR). Among these exceptions, you have notably:

i. processing relates to personal data which are manifestly made public by the data subject (art. 9 par. 2 lit. e GDPR);
ii. processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (art. 9 par. 2 lit. i GDPR).

[Rz 15] Probably, in regards to the first option, the provision would need to be construed restrictively.[11] When new data is created from the published personal data, which brings new information about an individual, a separate legal basis would be required.[12] In fact, for Jakob Zanol, art. 9 par. 2 lit. e GDPR may require legitimate interests from the controller and, consequently, art. 6 GDPR may be applicable in addition.[13] This opinion goes too far in our view, and mixes up the concept of lawfulness and the principles contained in art. 5 GDPR (in particular the principle of purpose limitation).

[Rz 16] Regarding the second option, a national implementation is necessary and it requires putting in place appropriate safeguards, such as sufficient technical and organizational measures, in order to ensure respect for the principle of data minimization notably (art. 89 GDPR). For instance, some Member States ask for the adoption of a code of conduct[14], the prior approval of the

11 Jakob Zanol, Öffentlich gemachte Daten und Datenschutz, in: Jusletter IT 21. Februar 2019.

12 Dietmar Jahnel, Handbuch Datenschutzrecht: Grundrecht auf Datenschutz, Zulässigkeitsprüfung, Betroffenenrechte, Rechtsschutz, Jan Sramek Verlag (Wien), 2010, § 4/24; Markus Kastelitz/Walter Hötzendorfer/Christof Tschohl in: Rainer Knyrim (Ed.), Der DatKomm. Praxiskommentar zum Datenschutzrecht, DSGVO samt DSG und Nebenbestimmungen, Manz Verlag (Wien), 2018, Art. 9 § 41.

13 Jakob Zanol, Öffentlich gemachte Daten und Datenschutz, in: Jusletter IT 21. Februar 2019.

14 Art. 186 of Belgian law regarding the protection of natural persons in relation with personal data processing, as of 30 July 2018.

supervisory authority[15], or the realization of a data protection impact assessment[16], etc. Because Member States have the power to legislate in this field, and because they have a certain appreciation margin, different requirements exist or will continue to exist in Europe. Therefore, in the research field, important legal differences remain, despite the entry into force of the GDPR, and the national legislation will keep on playing a key role.

4. Challenge with the duty to inform

[Rz 17] Data harvesting is particularly challenging if information must be given to the concerned data subjects. By automatically collecting personal data on social media, information is not obtained from the data subjects directly. Especially with social media, the amount of data may be important, and the number of concerned data subjects too. Sometimes, it is even difficult to know if account holder's details are referring to a true person (the user can use pseudonyms, or is just a fake account). Automatically, it triggers practical complications with the duty to inform bestowed on the data controller.

[Rz 18] In principle, the data controller shall furnish information to data subjects (art. 14 par. 1 GDPR). It means that the data controller must provide numerous details, such as the purposes of the processing, the categories of personal data in question, or the source from which the personal data originate (art. 14 par. 1 and 2 GDPR). A recent case showed us that a supervisory authority can be severe when the principle of transparency is not respected.[17] Therefore, when a data controller deviates from the principle of transparency, he/she must be sure to comply with the principle of accountability (art. 24 GDPR), and must be cautious to inform the data subjects if he/she wishes to take advantage of the exemptions. In this respect, art. 14 par. 5 indeed provides for several exceptions. The controller does not have the obligation to inform, notably when such information is impossible to provide (i), involves a disproportionate effort in particular when processing for archiving purposes in the public interest, for scientific or historical research purpose or for statistical purposes (ii), or also if such information is likely to seriously impair or render the achievement of the research-related processing impossible (iii).[18] In such cases, the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests (art. 14 par. 5 lit. b GDPR).

[Rz 19] According to the Working Party 29 (now called European Data Protection Board), in the first hypothesis, the controller shall demonstrate that he/she cannot objectively inform the concerned individuals. In our case, this exemption cannot apply, because there is no objective ground that prevents to give information. The mere fact that a database has been compiled by a data controller using more than one source is not enough for raising an impossibility to in-

15 Art. 10 par. 3 of the Danish Act on supplementary provisions to the regulation on the protection of natural persons with regard to the processing of personal data on the free movement of such data of 23 May 2018.

16 Art. 31 in fine of the Finlandese Protection of Privacy Law.

17 David Vasella, Polen: DSGVO-Busse von EUR 220'00 (Verletzung der Transparenz): http://datenrecht.ch/polen-dsgvo-busse-von-eur-220000-verletzung-der-transparenz/ (consulted on 28 March 2019). The concerned Polish data controller collected personal data from public registers and used it for commercial purposes and informed only 90'000 people out of a total of 6 million affected people.

18 Article 29 Working Party, Guidelines on transparency under Regulation 2016/679, adopted on 29 November 2017, § 58.

form.[19] For the second hypothesis, it requires to demonstrate that the provision of information to the data subjects would involve a disproportionate effort, notably because of the number of data subjects, or the age of the data. The disproportionate effort must result from the fact that the personal data has been collected through an intermediary.[20] Furthermore, the controller must carry out a balancing test, which weighs the effort of the controller and the effects on the data subjects. This assessment must be documented and must result in the implementation of appropriate measures. One appropriate measure may be to render the information publicly available on the controller's website or in a newspaper, and also other important measures could be the realization of a Data Protection Impact Assessment, the application of pseudonymization techniques, or the minimization of the collected personal data.[21] This scenario would be relevant in the case of data harvesting, because it is often performed on social media or on websites containing numerous information from a lot of people from all around the world, which renders the provision of information much more difficult and costly. Of course, it calls for a careful analysis and for a balance of interests as described above. Thirdly, if the objectives of the processing are seriously impaired by the provision of information, the controller would not need to inform data subjects. To rely on this exception, the controller has to demonstrate that informing individuals would nullify the objectives of the processing.[22] For instance, we could imagine a research project where the information of data subjects can create biases that nullify the scientific results.

[Rz 20] Besides, the data subjects maybe already have received the information, but for data harvesting activities, and regarding the extent of information to be furnished, this exception could be difficult to put forward (art. 14 par. 5 lit. a GDPR).

5. Further use?

[Rz 21] Data harvesting can occur after an initial collection of personal data. Therefore, data harvesting may be seen as further use of personal data.

[Rz 22] A possible antagonism of purpose between the initial data collection and the subsequent data processing «should be allowed only when the processing is compatible with the purposes for which the personal data was initially collected» (Recital 50 of the GDPR). In such a case, no separate legal basis is required (Recital 50 of the GDPR).[23]

[Rz 23] Further processing for archiving purposes in the public interest, or for scientific, historical research purposes or statistical purposes «shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes» (art. 5 par. 1 lit. b in fine GDPR).

[Rz 24] Particularly, when the processing relies on a ground other than the data subject's consent or on a Union or Member State law, the controller shall determine whether its purpose is compatible with the purpose for which the personal data is initially collected, by taking into account different criteria, as set forth in art. 6 par. 4 GDPR.

19 Ibidem, § 60.

20 Ibidem, § 62.

21 Ibidem, § 64.

22 Ibidem, § 65.

23 Interpretation almost confirmed in the EDPB, Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (70.1.b), adopted on 23 January 2019, § 31; however, as mentioned in the document, further opinion / guideline is required.

[Rz 25] The context at the time of collection and the information provided initially are important. As a reminder, the purpose must be specified, explicit and legitimate (art. 5 par. 1 lit. b GDPR). If the initial purpose is vague or general, such as for instance «improving users' experience», «marketing purposes», «IT-security purposes» or «future research», it will usually not meet the criteria of being «specific».[24]

[Rz 26] In our view, the GDPR is, however, unclear about who is able to make further use of the personal data: is it the initial controller, or is it also applicable to a new data controller? In our view, only the initial data controller could raise the possibility of art. 6 par. 4 GDPR (pursuant to a literal interpretation), but another opinion is possible.[25]

[Rz 27] In fact, to rely on further use for justifying data harvesting is difficult. First, to our knowledge, no particular guidance or case-law is really helpful. Secondly, the fact of considering all the criteria set out in art. 6 par. 4 GDPR implies more or less a «perfect» initial collection of the personal data and presupposes a low risk for the rights of the data subjects. Finally, data harvesting occurs often when a new data controller wants to process personal data of the initial data controller (for instance, the corresponding social media platform). The new data controller should hence not have the right to claim for further use of the personal data in order to justify the new data processing. He/she should seek its own legal justification.

C. Swiss law: FADP

[Rz 28] Swiss entities must comply with the FADP, since the act applies to the processing of data pertaining to natural persons (art. 1 FADP). Nonetheless, for data harvesting activities, the collection of data would be made on foreign websites, and the personal data in question would not necessarily pertain to Swiss citizens. Accordingly, does Swiss law still apply? In Switzerland, the effect theory applies.[26] If the concerned personal data does have «a close connection» with Switzerland, the FADP is applicable. It should be the case in data harvesting activities, because the data controller would have its seat in Switzerland, and the data processing (notably the storage for example) takes place in Switzerland.

[Rz 29] A total revision of the law remains a matter of discussion, but the Swiss Parliament decided not to review the draft too precipitously. Discussions have been postponed. We probably cannot expect to have an entry into force of any potential revision before 2021.[27] In this text, we will be referring to the «new FADP», even if we do not know at this stage if the Parliament will accept the draft as it is.

---

24 Article 29 Working Party, Opinion 03/2013 on purpose limitation, adopted on 2 April 2013, p. 16

25 Lionel Maurel, Données personnelles et recherche scientifique : quelle articulation dans le RGPD ?: https://scinfolex.com/2018/07/18/donnees-personnelles-et-recherche-scientifique-quelle-articulation-dans-le-rgpd/ (consulted on 23 April 2019); he considers that researchers are therefore not obliged to collect the data themselves on the basis of consent. For him, it is conceivable that this type of research partnership involves the establishment of agreements under which the data provider and the research team identify themselves as joint data controllers for the processing.

26 BGE 138 II 346 consid. 3.

27 See Federal Council, Object n°17.059 regarding the total revision of the Federal Act on Data Protection and other laws: https://www.parlament.ch/en/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20170059 (consulted on 21 March 2019).

[Rz 30] We will use the same structure as before, i.e. analyzing the lawfulness, the challenge with the duty to inform, and the concept of further use. In Swiss law however, the lawfulness will be different for private entities and federal bodies. Additionally, we will determine to which extent the new draft of the law will change the legal regime.

1. Lawfulness: private entities

[Rz 31] Beforehand, it is necessary to check whether an infringement of any personality rights has occurred. Three constitutive acts may affect the concerned data subject: (1) a general principle has been breached; (2) no consent has been obtained; (3) sensitive data or profiling data was disclosed to a third-party (art. 12 par. 2 FADP). Other breaches are possible, notably when the personality of the data subject is affected (art. 12 par. 1 FADP). Nevertheless, there is «no breach of privacy if the data subject has made the data generally accessible and has not expressly prohibited its processing» (art. 12 par. 3 FADP). For this, it would be necessary for the data subject to have made their data generally accessible with knowledge and will or to have made it accessible to a third party. Just tolerating the act of a third party without contributing actively is not enough (typically for public registers).[28] Therefore, in the context of data harvesting, if the concerned personal data has been made willfully and knowingly public by the data subject, no infringement will be considered to have taken place, and it means that data harvesting activities would be seen as valid. It will nonetheless require to analyze for which purpose the data subject publishes the personal data, and especially whether the publication was made willfully and knowingly. Otherwise, it means that the processing of this public personal data would lead to a breach of the data subject's personality.

[Rz 32] FADP allows a presumption: in the case of a personality breach, the data processing will be deemed illegal. This presumption can be rebuttable when there are specific justifications. However, the reversal would be admitted exceptionally, because such a justification must be accepted restrictively[29], particularly (and exclusively?) when the justification is stemming from an overriding private or public interest. FADP lays down that an infringement is unlawful, unless the processing may be justified by the consent of the victim, an overriding public or private interest, or by the law (art. 13 par. 1 FADP). Three legal grounds are then available in Swiss law, and the reference to «the law» implies that the justification may be found in another sectorial regulation.

[Rz 33] In order to ascertain whether an overriding interest exists, a balance test is required. FADP contains a non-exhaustive list of private interests that could prevail (art. 13 par. 2 FADP). Among these interests, the processing for purposes not relating to a specific person, in particular for the purposes of research, planning and statistics, is admissible (art. 13 par. 2 lit. b and e FADP). This justification targets processing whereby the identity of the data subject does not play any role, and which could have been performed on pseudonymized or anonymized data.[30] Subject to the condition that research results are published on a format that does not allow the identification

28 Judgement of the Federal Administrative Court A-4232/2015 of 18 April 2017, consid. 5.4.1.

29 BGE 136 II 508, consid. 5.

30 Federal Data Protection and Information Commissioner, Data protection and research in general: https://www.edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/statistik–register-und-forschung/recherche/protection-des-donnees-et-recherche-en-general.html (consulted on 20 March 2019).

of data subjects and also provided that general principles are respected, personal data harvesting may be lawful for research or statistical purposes for instance.

[Rz 34] The new FADP law would use exactly the same methodology and systematic approach. No major change is currently planned.[31] The only change pertains to the justification in case of data processing for research purposes notably. New requirements would be necessary in order to claim it: (i) the personal data must be anonymized whenever it is possible; (ii) the sensitive personal data must be communicated to a third party only in a form that does not allow the data subjects to be identified; (iii) and the results must be published in a form that excludes re-identification of data subjects.

2. Lawfulness: federal bodies

[Rz 35] For federal public bodies, the overall processing is examined in light of the principle of legality, which demands to have a simple legal basis (art. 17 par. 1 FADP) or a formal legal basis («a law») in case of processing of sensitive personal data (art. 17 par. 2 FADP).[32] Regarding the latter, the specific legal act must provide which sensitive data is concerned, and must contain at least schematically the purposes of the data processing to ensure that the processing is recognizable for the concerned data subject.[33] Those requirements should also be applicable irrespective of whether sensitive personal data is processed, because a legal basis needs to be sufficiently specific and clear. A formal legal basis is nevertheless not necessary if the federal public body absolutely needs to accomplish a legal task clearly identified in a law (i)[34], if the Federal Council decided it specifically (ii), if the concerned data subject has given its consent in a particular case, or if its personal data was disclosed publicly and no opposition by the data subject has been raised for the data processing (iii) (art. 17 par. 2 FADP).

[Rz 36] Generally, federal public bodies do not need to gather the consent of data subjects if they want to process the personal data.[35]

[Rz 37] If a federal body wants to process personal data for purposes not related to specific persons (such as for research or statistical purposes), art. 22 FADP would not constitute an adequate legal basis pursuant to art. 17 FADP, but this provision offers more flexibility if the conditions are met. Three conditions are imposed: the data must be rendered anonymous, as soon as the purpose of the processing allows to do it (i); a recipient of the personal data may disclose the personal data only with the consent of the federal body (ii), and the results are published in a manner that does not allow the identification of the data subject (iii) (art. 22 par. 1 lit. a to c FADP). This legal basis reduces the regulatory burden in some aspects, notably regarding the purpose of processing (art. 22 par. 2 lit. a FADP) and the necessity to have a formal legal basis for the processing of sensitive personal data (art. 22 par. 2 lit. b FADP).

31 See notably art. 26 par. 3 and art. 27 of the new Federal Act on Data Protection (FADP; CC 235.1).

32 BGE 131 II 413, consid. 2.3 et 2.5.

33 BGE 137 I 167, consid. 9.1.1.

34 Agnès Hertig-Pea, La protection des données personnelles est-elle efficace ?, Helbing Lichtenhahn (Neuchâtel), 2013, p. 106.

35 BGE 131 II 413, consid. 2.5.

[Rz 38] Research institutions, which are attached to the Federal Institutes of Technology Domain (ETH Domain), such as EPFL, ETHZ, PSI, WSL, Empa, or EAWAG, can process personal data on the basis of a specific law (Federal Act on the Federal Institutes of Technology of 4 October 1991, ETH Act).[36] They are notably allowed to process regular and sensitive personal data for research purposes (art. 36c and seq. ETH Act).

[Rz 39] The new FADP takes up the fundamental principle of legality: a federal body is only entitled to process personal data if there is a legal basis (art. 30 and seq. of the new FADP). The two distinct regimes have been clarified: an ordinary regime for regular personal data and a regime for sensitive personal data[37]. For regular personal data, the principle of legality applies in full. For sensitive data, a law in the formal sense is generally required, but a law in the material sense may exceptionally be sufficient if the processing is essential for the performance of a task clearly defined in a formal law (i) and if, in addition, the processing is not likely to involve specific risks to the personality and fundamental rights of the concerned person (ii). The second condition is new and sets an additional precaution.[38]

3. Challenge with the duty to inform

[Rz 40] Again, we must distinguish between obligations for private entities and federal bodies. As mentioned before, data harvesting is often performed on social media, where the data controller can find a sufficient amount of data. In some circumstances, the obligation to inform becomes an unrealistic obligation, due to the quality of the data, the number of the data subjects, the information's fatigue, etc. Of course, in Swiss law, the principle of transparency is also a key element in data protection. But, in some cases, exceptions may be raised.

[Rz 41] The new FADP will harmonize the obligations between private entities and federal bodies. The legal regime would change slightly but the main ideas would remain.[39] Currently, the duty to inform is limited to what is essential, and the legal provisions are rather sparse and limited in scope. The current dichotomy between regular personal data and sensitive personal data will be removed and replaced by a single provision dealing with the duty to inform when collecting personal data (art. 17 of the new FADP). In addition, it will remove the distinction between a federal body and a private person. Regarding the exceptions to the duty to inform, the text hereafter would remain valid for data harvesting activities.

a. For private entities

[Rz 42] Currently, for private entities, there is no provision obliging them to inform data subjects about specific information if the personal data is solely «regular». The duty to inform is stemming from the general principle of fairness and the principle of purpose limitation (art. 4 par. 3 and

36 For the scope of the law, see notably the Swiss Federal Council Message related to the partial revision of the ETH Act of 27 February 2002 (BBI 2002 3251 p. 3263).

37 Explanatory report on the preliminary draft federal law on the total revision of the Data Protection Act and on the amendment of other federal laws, Office fédéral de la justice, December 2016, p. 68.

38 Ibidem.

39 Swiss Federal Council, Message related to the total revision of the Federal Act on Data Protection and other laws of 15 September 2017 (BBI 2017 6565 p. 6668).

4 FADP). The processing must be apparent, and the intended use of the personal data must have already been specified at the time of data collection.[40] However, since sensitive personal data is concerned, a specific provision requires to inform the data subjects about the identity of the data controller (i), the purposes of the data processing (ii), and at least the categories of recipients (iii) (art. 14 par. 2 FADP). Exceptionally, the data controller can set aside this duty if:

i. the data subjects already received the information;
ii. a formal legal basis enables it;
iii. a third party's overriding interest can be raised;
iv. its own overriding interest requires it, provided no personal data is communicated to a third-party; or
v. provided the personal data have been collected through a third-party, a legal basis covers this further processing (1), or finally the provision of information is impossible (2) or requires disproportionate efforts (3).
(art. 14 par. 4 and 5 cum art. 9 par. 1 and 4 FADP)

[Rz 43] Besides, sectorial laws could also provide for more specific rules (for instance, the Federal Act on Research involving Human Beings of 30 September 2011).

[Rz 44] Applied to data harvesting, probably two exceptions could be applicable. First of all, it could happen that data subjects were already informed thanks to privacy policies for instance. Furthermore, whereas the data subject published the information itself, there would be a presumption that the person is already informed.[41] Secondly, and principally, the data controller could pretend that the provision of information would lead to disproportionate efforts, because the personal data originated from social media and from thousands or hundreds of thousands of users. According to the explicative report regarding the pre-project for a total revision of the FADP, this exception must be deemed restrictively: the data controller must deploy all the efforts that one could expect from him to fulfill his duty to inform.[42] In our opinion, a balance test should be performed when a data controller considers how to inform the data subjects. For instance, in some situations, an individual information would require a large amount of work, although a general publication in the concerned website or in a newspaper could reach a sufficient percentage of the users and could demonstrate the good faith of the data controller.

b. For federal bodies

[Rz 45] The duty to inform has been more precisely defined when personal data is processed by a federal body. The federal body has the obligation to inform the data subject of any data collection (art. 18a par. 1 FADP), and must ensure that the data subject received certain information (art. 18a par. 2 FADP). If the data is not collected directly from the data subject, the latter must be informed at the latest at the time of recording[43] or, in the absence of recording, at the time of the first communication to a third party (art. 18a par. 3 FADP).

40 BGE 138 II 346 consid. 9.1.

41 Federal Department of Justice and Police, Explicative report regarding the pre-project for the total revision of the Federal Act on Data protection and other laws of 21 December 2016, § 8.1.3.2.

42 Ibidem.

43 The notion of recording does not only include the technical act of recording the data collected, for example in a computer system; it includes any subsequent act after the collection that prepares the use of the data (Swiss Fed-

[Rz 46] The federal body is released from its obligation if the data subject has already been informed (i), or, provided the personal data was not directly collected from the data subject, if the recording or the communication is expressly laid down in a law (ii), as well as the provision of information is impossible (iii) or requires disproportionate efforts (iv) (art. 18a par. 4 FADP). In the latter case, given the circumstances, the federal body must take steps that can be reasonably expected and its behavior would be examined in accordance with the principle of good faith.[44]

[Rz 47] In the context of data harvesting, the concerned data subjects may have already received the information, and if this is the case, the data controller must prove it. Otherwise, if the data controller cannot objectively inform the concerned data subjects because it collects personal data from a third-party website, it would need to demonstrate why it was impossible to inform those data subjects, or why it would require a disproportionate effort. In our view, a balance test must be done between the costs and efforts that need to be furnished by the data controller and the consequences on the rights and freedoms of the data subjects. The data controller would need to choose the option which allows at best to inform the data subjects and to minimize the risks for data subjects.

[Rz 48] In addition, a specific law may be applicable. In the ETH Domain for instance, the institutes, such as EPFL or ETHZ, will be obliged to inform the concerned data subjects about the personal data processing and the collection in connection with a specific research project (art. 36e ETH Act). Particularly, when personal data is collected through a third party (a social media platform for example), the concerned institution must ensure that the third party duly informed the data subjects (art. 36e par. 2 ETH Act). They should even demand a written confirmation from the third party stating that the concerned individuals have been duly informed.[45] Another question is the possibility to claim for the same exceptions contained in the FADP. As a rule, *lex specialis derogat lex generali*, but the ETH Act is not really precise and the Message of the Federal Council is rather succinct about the matter. In our view, there is no reason to be more restrictive in the field of research in comparison with other federal mandates. According to the law however, there is no place for exceptions in the field of research within the ETH Domain.

4. Further use?

[Rz 49] In the FADP or in the new FADP, there is no specific provision or case law that addresses this concern. Of course, the law enshrines the principles of «good faith»,[46] «purpose limitation» or «recognizability»[47] (art. 4 par. 3 and 4 FADP), but the concept of further use is absent in the FADP. As a rule, a data controller cannot change the purpose of the data processing without informing the concerned data subjects beforehand. Otherwise, for private entities, the processing will affect the personality of the data subjects (art. 12 par. 2 lit. a FADP). However, the purpose

eral Council Message related to the partial revision of the FADP in relation with the adoption of the additional protocol of Convention 108, BBI 2003 1915 p. 1943).

44 Swiss Federal Council Message related to the partial revision of the FADP in relation with the adoption of the additional protocol of Convention 108 (BBI 2003 1915 p. 1944).

45 Swiss Federal Council Message related to the promotion of education, research, and innovation during the years 2017 to 2020 (BBI 2016 2917 p. 3082).

46 Philippe Meier, Protection des données, Stämpfli (Berne), 2011, p. 264.

47 Ibidem, p. 276–277; BGE 136 II 508 consid. 4; BGE 138 II 346 consid. 9.1 and consid. 11.

modification may be justified by the law or by an overriding interest.[48] We can say that there is no concept of further use and the analysis of the processing's lawfulness must be done (*supra* C.1). For the federal bodies, the conclusion is the same, since their processing is subject to the principle of legality (*supra* C.2). Nevertheless, there is one exception for federal bodies when they process personal data notably for research or statistical purposes, because the concerned federal body has no obligation to observe the purpose limitation's principle (art. 22 par. 2 lit. a cum art. 4 par. 3 FADP). It means that a federal body can process personal data for a different purpose than those that had been communicated at the time of the collection.

[Rz 50] Finally, a data controller cannot pretend to justify further use without performing the overall standard legal analysis. A notable exception is available however for federal bodies when they rely on art. 22 FADP. The latter legal basis may be raised by a federal body in case of data harvesting activities for research or statistical purposes.

D. Conclusion

[Rz 51] In conclusion, the activities of data harvesting are particularly challenging in terms of data protection. Most of the time, the controller would need to perform balance tests between relevant interests, first for assessing the lawfulness, and, secondly, for determining its obligation to inform data subjects. Moreover, data harvesting activities must comply notably with principles of fairness, transparency, purpose limitation, accuracy, and security (art. 5 par. 1 GDPR cum art. 4, art. 5 and art. 7 FADP). Of course, in Swiss law, the particularity lies in the fact that the federal bodies are strictly subject to the principle of legality regarding data harvesting activities. And to make matters even worse, data harvesting activities can breach contractual obligations, intellectual property rights, or competition rules. It will especially be hard not to infringe terms of use of social media sites for example. Usually, social media platforms require their approval before automatically collecting data, unless their own API is used.[49] Data harvesting is probably an easy thing to do, but people need to be very cautious. Data harvesting is very popular in the research field. Consequently, Universities should be proactive and should issue guidelines or instructions to researchers.

[Rz 52] Besides, one curious thing in the GDPR is the possibility to process sensitive data when the concerned personal data has been manifestly made public by the data subject, whereas this possibility does not exist for «regular» personal data. In Swiss law, whatever the kind of personal data, the fact that personal data is made public may be an argument for considering the processing as lawful for private entities, if all conditions are fulfilled.

Valentin Conrad is a Legal Counsel and works for the Ecole Polytechnique Fédérale de Lausanne (EPFL). He studied law at the University of Geneva and the University of Neuchâtel (Switzerland). After short experiences in the watchmaking and pharmaceutical industries, he joined his current

48 Philippe Meier, Protection des données, Stämpfli (Berne), 2011, p. 283.

49 According to their terms of use (consulted on 17 March 2019), Instagram and Facebook seems to prohibit data scraping and data crawling, whereas Twitter authorizes data crawling subject to the provisions of the robots.text.file.

employer as a Legal Counsel. He is now specialized in data protection issues in the field of research. The author wishes to thank Mr. Ruwan Schneuwly for his careful rereading.