PANELFIT

PARTICIPATORY APPROACHES TO A NEW ETHICAL AND LEGAL FRAMEWORK FOR ICT

Guidelines on Data Protection Ethical and Legal Issues in ICT Research and Innovation.

SOCIAL NETWORKS

Social networks for research purposes: ethical and legal requirements regarding data protection

Jose Antonio Castillo Parrilla and Iñigo de Miguel Beriain (UPV/EHU)

Preliminary versions of this document were reviewed by Dr Denise Amram, DPO, affiliate researcher at LIDER Lab - DIRPOLIS Institute, Scuola Superiore Sant'Anna (Italy) and DPO *of Private comparative law at Scuola Superiore Sant'Anna and Prof. Giovanni Comandé, Dirpolis, Sant'Anna School of Advanced Studies, Pisa, Italy.*

This part of the Guidelines was validated by Iñaki Pariente, former director of the Basque Data Protection Agency

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 788039. This document reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

Social media can be described as online platforms that enable the development of networks and communities of users, among which information and content is shared. Additional functions of social networks are personalization, analytics and publishing (mainly via targeting services), allowing either freelance initiatives or wider service offers. Social media allow individuals to create accounts for themselves in order to interact with other users and to develop and broaden connections and networks. Users share data with the with network administrators and with other users for totally different purposes. The content shared by individuals can be created by themselves (user-generated content) or not.[1]

On the other hand, it is important to mention that the main purpose of the data placed in a social network is to allow people to interact, to relate. In fact, users establish two types of relationships: a vertical relationship with the company that owns the network, and a horizontal relationship with other people with whom they want to interact. This relationship can be general (open profiles) or particular (profiles with limited access). Depending on the type of interaction at stake, the legal status of data processing will probably be different.

In general, social networks are optimal for **massive data extraction practices**. Indeed, there are software tools available that can automatically collect web users' data from online public spaces. Furthermore, most social networks enable Application Programming Interfaces, or APIs[2], that simplify software development and innovation and make it possible for applications to exchange data and functionality easily and securely. These circumstances make social networks particularly attractive for some kinds of research, but it also creates demanding challenges in terms of data protection issues.

This part of the Guidelines is aimed at helping ICT researchers or **innovators using personal data obtained from social networks.** It is worth mentioning that we will not address here the use of social networks to collect data (such as, for example, by using Google surveys to get data back on a specified set of questions from real people). This is due to a simple reason: in these cases, the data itself does not come from a social network but through a social network. Indeed, social networks only act as a tool to gather those data. Therefore, these data are not so different to any other data collected by a more traditional way (such as a survey in paper) and, thus, they do not deserve special attention here.

If ICT developers consulting these Guidelines are planning to use AI tools to process data obtained from these networks, they should consult the part of the Guidelines

[1] EDPB Guidelines 8/2020 on the targeting of social media users, p. 3.

[2] See, on APIs: Oscar Borgogno & Giuseppe Colangelo, Data Sharing and Interoperability Through APIs: Insights from European Regulatory Strategy, Stanford-Vienna European Union Law Working Paper No. 38, http://ttlf.stanford.edu; Russell, N. Cameron and Schaub, Florian and McDonald, Allison and Sierra-Pambley, William, APIs and Your Privacy (February 5, 2019). Available at SSRN: https://ssrn.com/abstract=3328825 or http://dx.doi.org/10.2139/ssrn.3328825

devoted to Artificial Intelligence (AI). If they are planning to use them for purposes related to biometrics, Internet of Things or Geospatial location, they should consult the parts of these Guidelines that are devoted to those issues. In order to avoid unnecessary repetitions, we are leaving such issues out of this analysis.

DISCLAIMER

This part of The Guidelines was written at a time when the ePrivacy Regulation had not been approved. It may happen that, at the time of using this tool, the Regulation is in force. If so, it will be necessary to take into account the possible changes that this may have produced in the regulatory framework. Until the ePrivacy Regulation enters into force, a fragmented situation will exist. Indeed, supervisory authorities face now a situation where the interplay between the ePrivacy Directive and the GDPR coexist and pose questions as regards the competences, tasks and powers of data protection authorities in those matters that trigger the application of both the GDPR and the national laws implementing the ePrivacy Directive.

1 Introduction to social networks and data protection issues

Some preliminary advice: It is absolutely necessary to keep in mind **that the fact that much of the data contained in a social network is easily apprehensible does not legitimize its processing.** This is a crucial aspect when it comes to the processing of data obtained from social networks: ICT researchers and innovators must carefully ensure that they have a legal basis that allows them access and storage of these data. Once they have already accessed them, they will have to make sure that the same and/or other bases of legitimacy allow them further processing of those data. In general, this means that they must have a profound knowledge of the Developer Policies imposed by the social networks (see "Lawfulness, Fairness and Transparency" in "Principles" within Part II of these Guidelines for further reading on this).

Furthermore, transparency implies that intended research subjects should be informed at some point about the research being performed, what sort of personal data controllers are collecting and how it will be used. Some services make it clear that this must be done before one starts harvesting. In the absence of a specific policy and where researchers/innovators are conducting observational research which the need to obtain consent up front could damage, they should let the individuals concerned know as soon as possible. The ICT researchers/innovators should always remove from their harvesting individuals who do not consent to being included.

1.1 Main concepts

The imprecise nature of the data gathered by social networks, along with personal data protection legal rules, suggests to social network managers, and those who use data gathered from these networks for research purposes, that, rather than the category of the social network, they should take into account **both the type of data being processed and the following main criteria:**

- the purpose for which they are using the data;

- the applicable regulation, and in particular the regulatory conflicts that may arise from their activity and the original purposes of personal data collection in the social networks.

A social network is an **information society service.** The concept of information society service is mentioned in Article 2.a and Recitals 17 and 18 of EC Directive 2000/31, as well as in Article 4 (25) GDPR. They all refer to Article 1.1.b of EU Directive 2015/1535. An information society service is any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services.

The operators of a social network have the dual status of social service provider and data controller, according to their privacy policy, which considers them as such.[3] As social service providers, they are subject to liability under Articles 12 to 15 of Directive 2000/31/EC. As data controllers, they are responsible both for ensuring that the data are processed in accordance with Article 5 GDPR, as well as for being able to demonstrate this (Art. 5.2 GDPR). Those who use the social network for purposes that go beyond being a mere user (e.g., using social networks for research) **shall also be regarded as data controller, and shall be liable accordingly. However, it is also true that, in case of joint controllership,** controllers may be involved at different stages of the processing of personal data and to different degrees. In such scenario, the level of responsibility of each of them must be assessed with regard to all the relevant circumstances of the particular case.[4]

1.2 Challenges

The use of data gathered by social networks raises, in itself, certain **challenges** related to data processing that should be taken into account. These challenges may be even more peculiar when the purpose of the processing is related to research. The main issues involved in the use of data gathered through social networks for research purposes are the following:

- Social networks favor and enhance the constant reuse of data, which poses risks related to
    - the application of principles such as purpose limitation (Art. 5.1.b), retention period limitation (Art. 5.1.e), integrity and confidentiality (Art. 5.1.f); etc.
    - or the legal status of personal profiles and other derived data, in particular whether they remain personal data and whether they are also works of intellectual property (IP) (the question whether or not inferred personal data are personal data or just the IP of their producers).
- The choice and correct use of a legal basis for the collection of data from social networks, which requires an adequate understanding and fulfilment of the requirements of their Developer Policies
- The choice of a legal basis for the re-use of data obtained through social networks and the adequate use of those data according to the basis selected:
    - Consent (and the possibility to obtain "altruistic consent" especially in light of the proposed Data Governance Act).

[3] In order to clarify the respective roles and responsibilities of social media providers and targeters, it is important to take account of the EDPB Guidelines (Guidelines 8/2020 on the targeting of social media users Version 2.0 Adopted on 13 April 2021, at: https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf, p. 11) and the relevant case law of the CJEU. The judgments in Wirtschaftsakademie (C-210/16), Jehovah's Witnesses (C-25/17) and Fashion ID (C-40/17) are particularly relevant here.

[4] See: 4 Judgment in Wirtschaftsakademie, C-210/16, paragraph 43; Judgment in Jehovah's Witnesses, C-25/17, paragraph 66 and Judgment in Fashion ID, C-40/17, paragraph 70.

- Legitimate interest

- Public interest

- Research exception

- The identification of risks arising from research with social media data, among which the following stand out:

  - harm to individual privacy through mass analysis of personal or non-personal data (group privacy), e.g. due to the identification (or re-identification) of data subjects through personal profiles (this clearly involves an extremely high risk due to the intention to promote massive analysis of data that could lead to profiling);

  - or damage to the honor, privacy or image of individuals or groups, for example, by publishing raw data without going through a correct aggregation or pseudonymization process.

- The expansive nature of personal data, which makes it advisable to assume by default that personal data are being processed, even though at first sight this may not appear to be the case.[5]

- Although on many occasions, and increasingly so, research through social networks is born as research, it is also frequent that the researcher's social network profiles do not have this initial purpose and only acquire it after some time.

-The common assumption that data made public through social media can be used freely. **This is clearly untrue unless data are actually published in fully public profiles ("manifestly made public by the data subject") and must be carefully avoided.**

- Finally, the opacity of data processing algorithms can have a negative impact on users and discourage research (see Part III on AI of these Guidelines)

1.3 Types of data that can be collected through social networks

Social networks can provide researchers with three different types of data: provided data, observed data, and inferred/derived data (or a combination of them all). These types of data could be defined in this way[6]:

[5] The Historic Graves project is a community focused grassroots heritage project. Local community groups are trained in low-cost high-tech field survey of historic graveyards and recording of their own oral histories. They build a multi-media online record of the historic graves in their own areas and unite to form a national resource. Since this is a project that collects data from graveyards, one might think that it is data of the deceased and therefore the GDPR does not apply (Recital 27). However, the data about graveyards and tombs are provided by the relatives of the deceased, who are obviously not deceased, and by providing the data of their deceased relatives they are also providing their own personal data.

[6] Guidelines 8/2020 on the targeting of social media users Version 2.0 Adopted on 13 April 2021, at: https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf

- "Provided data" refers to information actively supplied by the data subject to the social media provider and/or the controller. For example: social media users might indicate their age in the description of their profiles. In these Guidelines, we will not address the processing of such data, since this is not different to other data gathered by a service provider.
- "Observed data" refers to data provided by the data subject by virtue of using a service or device. These include:
    - data from a particular social media user might be gathered on the basis of the activity on the social media platform itself (for instance the content that the user has shared, consulted or liked);
    - data related to the use of devices where the social media's application is executed (for instance GPS coordinates, mobile telephone number);
    - data obtained by a third-party application developer by using the application programming interfaces (APIs) or software development kits (SDKs) offered by social media providers;
    - data collected through third-party websites that have incorporated social plugins or pixels;
    - data collected through third parties (e.g. parties with whom the data subject has interacted, purchased a product, subscribed to loyalty cards); or
    - data collected through services offered by companies owned or operated by the social media provider.
- "Inferred data" and "derived data" are those created by the data controller on the basis of the data provided by the data subject or as observed by the controller. These could be derived through deterministic computations or inferred probabilistically. For example, a social media provider might infer that individuals are likely to be interested in a certain activity or product on the basis of their web browsing behavior and/or network connections.

The way of obtaining the data is neither relevant when qualifying them as personal or non-personal data, nor when deciding whether they belong to special categories of data pursuant to Art. 9 GDPR. However, **it may have important consequences in other respects.** For example, when determining whether the data subjects could foresee, or not, a particular processing, or when determining the limits of their right of portability or the information to be provided to them. One must keep in mind that in the case of observed, inferred or derived data, users are usually unaware that data is being collected or generated.

Box 1: Inferring data. Examples

Example 1

"Company X has developed an application that, by analyzing raw data from electrocardiogram signals generated by commercial sensors commonly available for consumers, is able to detect drug addiction patterns. The application engine can extract specific features from ECG raw data that, according to previous investigative results, are linked to drugs consumption. The product, compatible with most of the sensors on the market, could be used as a standalone application or through a web interface requiring the upload of the data. Explicit consent of the user should be gathered to process the data for that purpose. Compliance with this consent requirement can be satisfied in the same conditions and at the time as when the consent is collected from the data subject under Article 7(a)."

Source: Art 29 Data Protection Working Party Opinion 8/2014 on the on Recent Developments on the Internet of Things (SEP 16, 2014) https://www.dataprotection.ro/servlet/ViewDocument?id=1088.

Example 2

Fitbit data could be relevant to prospective employers, who could make inferences about "impulsivity and the inability to delay gratification-both of which might be inferred from one's exercise habits-correlate with alcohol and drug abuse, disordered eating behavior, cigarette smoking, higher credit-card debt, and lower credit scores. Lack of sleep-which a Fitbit tracks-has been linked to poor psychological well-being, health problems, poor cognitive performance, and negative emotions such as anger, depression, sadness, and fear."

Source: Peppet, Scott R 'Regulating the Internet of Things: First Steps toward Managing Discrimination, Privacy, Security and Consent' (2014) 93 Tex. L. Rev. 85.

One must consider that inferring health data is a particularly sensitive processing since those data (no matter if they are inferred or not) are data of special categories.

1.4 Categories of data collected through social media

In principle, it is perfectly possible to gather **different types of data through social media. Indeed, these can be personal and non-personal data.** The understanding of data as non-personal is of great significance at a legal level and, obviously, for the preparation of these guidelines, insofar as the GDPR would not be applicable, but EU

Regulation 2018/1807 would. In practice, this division between these two types of data is blurring due to the increasing use of data analysis technology, allowing for greater data processing capacity and extrapolation of results (group privacy). This situation blurs the line between personal and non-personal data to the extent that, for example, profiles are becoming increasingly accurate even if they are not linked to any specific individual and are, therefore, not personal data.

The limit for considering data as personal lies in its capacity to directly or indirectly identify a person, and, in particular, if the costs and time involved in such identification are not excessive[7]. However, this sort of classification is not so easy to apply in practice. To begin with, some data that seem anonymous at first sight might be de-anonymized[8] (see subsections "Identification", "Pseudonymization" and "Anonymization" in the section "Main Concepts" within Part II of these Guidelines). Furthermore, personal data as a legal concept enjoys a sort of expansive nature insofar as the hyper-production of data and the capacity to process and analyze them is constantly growing, thereby reducing the costs and time needed to identify a person from any given set of data (personal or non-personal)[9].

Keeping all this in mind, one must conclude that, **in the case of social networks, the processing of personal data is generally the rule.** This is particularly true if we consider that in this context it is common for users to log in with a set of personal data. It is quite possible that (1) much of these data are not strictly necessary to log in and therefore does not comply with the data minimization principle (Art. 5.1.c GDPR) or that (2) the data are used for purposes that go beyond the mere login, in this case breaching the purpose limitation principle (Art. 5.1.b GDPR). Finally, personal profiling can reach a high level of accuracy irrespective of the type of data used for the production of such profiles. **This requires the following cautions to be taken into account:**

- The controllers **should assume by default that they are processing personal data** and act accordingly.

  - It is only advisable to avoid this assumption if the data to be used and the data inferred by the controller are entirely non-personal (e.g. weather data). In these cases, the controllers must document it in the records of processing.

  - If the data to be processed relates to deceased persons or legal entities, precautions must be taken to prevent these data from being linked to

[7] See Rec. 26 GDPR: "To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used".

[8] See Rec. 26 GDPR: "Personal data which have undergone pseudonymization, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person"

[9] See: in general G. Comandé (Editor) Encyclopedia of Data Science and Law Edwards Eldgar, 2021; forthcoming; G. Comandé - G. Malgieri, *"Sensitive-by-distance: quasi-health data in the algorithmic era" (2017)*, in *Information & Communications Technology Law*, Vol. 26, Iss. 3, p. 229-249; G. Comandé - G. Schneider, *"Regulatory Challenges of Data Mining Practices: The Case of the Never-ending Lifecycles of 'Health Data'" (2018),* in *European Journal of Health Law*, Volume 25, Issue 3, pages 284 – 307.

natural persons (e.g. relatives of deceased persons or natural persons linked to legal entities).

- If the data to be processed relates to deceased persons, national data processing rules must also be taken into account, since data from deceased persons are not personal data according to the GDPR.

- A level of granularity in profiling should be defined to sufficiently ensure the privacy of individuals who can potentially be linked to such profiling.

- Protocols should be developed to prevent or reduce the possibility of re-identification of data users whose data have been processed for profiling. They shall include a legally binding compromise not to seek for such re-identification and the adoption of measures devoted to avoid involuntary re-identification.

In addition to the initial distinction between personal and non-personal data, it should be taken into account, within personal data, **whether special categories of personal data are concerned**. This distinction is important insofar as the conditions for data processing vary depending on whether special categories of data (Article 9 GDPR) are concerned or not.

Finally, a note of consideration should be made on **derived or inferred data**. There has been some controversy as to whether or not derived data, and especially personal profiles, should be considered as intellectual property. Irrespective of this, it should be recalled that according to Article 4(1) GDPR **such data are personal data to the extent that they relate to an identified or identifiable person.** It should be added that it is possible to draw inferences related to what Article 9 GDPR considers special categories of data from ordinary personal data or even from non-personal data combined with other personal data (group privacy)[10]. To the extent that these inferences relate to an identified or identifiable person, they should be treated as special categories of data, irrespective of their understanding (or not) as objects of intellectual property.

2 Preliminary steps: the crucial issues to be considered

In this section, we provide some general advice on how to approach a research project in the **early stages of its production cycle,** i.e. when it is still not much more than an idea that has not yet been implemented. It is important to keep them in mind if one wants to ensure the implementation of data protection by design policies (see "Data Protection by design and by default" section on “Main Concepts” within Part II of these Guidelines).

The essential tips are:

[10] See in general Taylor, L., Floridi, L., van der Sloot, B. eds. (2017) *Group Privacy: new challenges of data technologies*, Dordrecht, Springer.

1. Make sure that your project is compatible with the data protection framework
2. Implement a training program in ethical and legal issues for ICT developers and other relevant stakeholders
3. Define the roles played by all agents involved in the processing
4. Promote end-users' engagement

2.1 Make sure that your project is compatible with the fundamental values of the EU

Before considering the use of data gathered from social networks for the project, the developers should have their primary objective clear in mind. It might happen that this use is not compatible with the ethical and legal standards of the EU included in the EU Charter of Fundamental Rights, for instance. **Were this the case, the project should not be endorsed.** On the other hand**, if an analysis shows that the processing needed will not be acceptable on the basis of the Developer Policy of the social network, the GDPR and/or the complimentary legal framework, the project should not be endorsed either.** Finally, developers shall assess whether the project is acceptable according to ethical standards, despite it being compliant with legal obligations (see "Privacy by Design and by Default" in the "Main Tools and Actions" within Part II of these Guidelines)

In addition, **a clear idea of the concrete use of data gathered through social networks will help controllers determine in the early stages of development some important legal issues regarding processing**, such as compliance with the Developer Policy of the social network, possible need of international transfers of data, existence of joint-controllers or processors -which need to be carefully selected-, or the security and organizational measures to minimize risks.

2.2 Implement a training program in ethical and legal issues for ICT developers and other relevant stakeholders

Implementing **basic training programs** for researchers/innovators involved in the processing might be extremely useful in order to avoid data protection issues while processing data obtained from social media. Some useful resources to this purpose are, for example, available by the Fundamental Rights Agency[11], IEEE and its ethics guidelines[12], and the European Commission[13]. **This training should also include a deep understanding of the Developer Policy of the social network from which the data will be gathered.**

[11] https://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law-2018-edition and
[12] https://ethicsinaction.ieee.org/
[13] https://ec.europa.eu/justice/smedataprotect/index_en.htm

If training is not possible, implementing advice from an **external expert** from the very beginning of the project could be an acceptable alternative. If the researchers/innovators are gathering data from a concrete social network, this training should include a careful analysis of its particular Developer Policy. An early involvement of DPOs from the participating institutions is highly advised.

Adopting appropriate measures in terms of ensuring confidentiality, integrity and availability of data is also strongly recommendable (see the 'Measures in support of confidentiality' subsection in the 'Integrity and confidentiality' section on 'Principles' within Part II of these Guidelines).

2.3 Define the roles played by all agents involved in the processing

The concepts of controller, joint controller and processor play a crucial role in the application of the GDPR, since they determine who shall be responsible for compliance with different data protection rules, and how data subjects can exercise their rights in practice[14] (see the "Main Actors" within Part II of these Guidelines, mainly the sections devoted to "Controller" or "Processor"). In the case of utilization of social networks for data processing, it is equally important to properly distinguish the data controller from the data processor, since the responsibilities of each are different.

Certain doubts may arise as to which of the parties involved in this framework plays the role of data controller, data processor or, as the case may be, whether there is a situation of joint controllership. **To dispel these doubts,** we must first turn to the list of definitions in the GDPR, interpreted in accordance with the EDPB Guidelines 7/2020 on the concepts of controller and processor in the GDPR and the EDPB Guidelines 8/2020 on the targeting of social media users [15] and the relevant case law of the CJEU[16].

In relation to the use of social networks for research, and without prejudice to the aforementioned casuistic caution, one might state that **there is no situation of joint controllership, insofar as the means and purposes of each processing operation are not determined jointly by the social network and the institution in charge of the ICT development, but rather the social network allows the developer use its environment.** The relationship between researchers and social networks is usually built on the so-called Developer Policies. Most social networks only allow researchers/innovators to collect data through their Application Programming Interfaces (APIs) if they follow the instructions settled in such policies. Thus, researchers/innovators shall ensure that they actually proceed to do so if they want to avoid taking responsibility for unlawful data processing. Of course, there is a possible exception to this general rule: if a developer hires the services of a social network to

[14] EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR, p. 3, at: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-072020-concepts-controller-and_en .

[15] EDPB Guidelines (Guidelines 8/2020 on the targeting of social media users Version 2.0 Adopted on 13 April 2021, at: https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf, p. 11).

[16] The judgments in Wirtschaftsakademie (C-210/16), Jehovah's Witnesses (C-25/17) and Fashion ID (C-40/17) are particularly relevant here.

process data on their behalf, this may involve joint controllership (it will depend on the concrete conditions of the contract and the way that responsibilities over the data are assigned to the partners). However, if such an exception does not apply:

- the social network shall be considered the controller in relation to the data processing it carries out in accordance with the purposes and objectives it pursues, and the ICT developer shall be data controller in respect of the data processing activities under its control;
- the relationship between the developer and the social network is as follows:
    - the social network plays the role of information society service provider, and
    - the research institution the role of information society service user.
- the activities carried out by the research institution from its research profile must be permitted by the social network as an information society service provider, but this does not imply that there is a situation of joint controllership nor that the licence to use the data guarantees a legal basis for personal data processing.

Thus, in most common scenarios, ICT researchers and innovators will play the role of a third party regarding social networks and data subjects. The network will provide them with data that belong to the data subjects. Once these data are already under the control of the researchers/innovators, they become controllers of those data and take the corresponding responsibilities.

Although a situation of joint controllership does not generally exist, it is not impossible for such a situation to arise at all. It is, therefore, worth recalling **the safeguards of Article 26 GDPR in the case of joint controllership** (see the "Main Actors" section within Part II of these Guidelines) **between the social network and the research institution:**

- Both the ICT developer and the social network shall, in a transparent manner, determine their respective responsibilities for compliance with the obligations under GDPR, in particular as regards to the exercising of the rights of the data subject and their respective duties to provide the information referred to in Articles 13 and 14, by means of **an arrangement between them.**
- The arrangement
    - shall be made available to the data subject;
    - may designate a contact point for data subjects;
    - shall duly reflect the respective roles and relationships of the joint controllers vis-à-vis the data subjects.
- Finally, all controllers, joint controllers and processors must remember that data subjects may exercise their rights under the GDPR (art. 26.3 GDPR).

2.4 Prepare the contracts with the social network and (in case) with the joint controllers, processors, etc. and document them

Gathering data from social networks often involves entering into some kind of agreement with their representatives. Indeed, access to their API, or similar tools, will probably not be provided if this agreement has not been documented. Sometimes, adherence to Developer Policies is not even part of this agreement, since it is crystal

clear that whoever receives data from the network has to follow them. The researcher/innovator should make sure, however, that this legal architecture is adequately fixed from the very beginning.

On the other hand, it is obvious that a controller will often entrust some of the technical tasks to a processor, who could even involve a sub-processor. In practice, however, there will be times when it will be difficult to ensure that the processor is not actually acting as a controller or joint-controller.

Researchers and innovators should do their best to avoid such issues, since the data protection regulation requires a clear answer to the question of "who is responsible for this processing?" to guarantee an "effective and complete" protection of the data subjects' rights and freedoms.[17] Thus, a key requirement of an adequate data protection by design policy is to **clarify from the very beginning who are the formal data controllers and processors, in order to ensure that the legal accountability is understood.**

In order to fulfil this goal**, written agreements between all agents involved in the development of the tools should be reached and documented, whenever possible (see art. 28 of the GDPR).** These should include clear specifications about the responsibilities taken by all participants. Promoting a continuous interaction between all DPOs involved might be an excellent option. Ad-hoc supervisory bodies and tools can be adopted to ensure a smooth oversight of the participants' processing.

2.5 Promote end-users' engagement

Since ICT involves the use of personal data from different types of data subjects, it is highly recommendable to hear the voices of the representatives of the collectives involved so as to ensure that the Data Protection by Design policies (see the "Data protection by design and by default" subsection in the "Main Concepts" within Part II of these Guidelines) are in line with their interest, rights and freedoms. Organizing some **preliminary discussions** with those representatives ensures the implementation of a bottom-up framework that could be very helpful to this purpose.

Checklist: Project Understanding

- The use of data gathered through social networks does not promote scenarios that are incompatible with the EU fundamental values.
- The ICT development does not involve a disproportionate use of personal data gathered through social networks

[17] See: EDPB Guidelines (Guidelines 8/2020 on the targeting of social media users Version 2.0 Adopted on 13 April 2021, at: https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf, p. 11)

- The controller ensured that the team members processing personal data have been\are adequately trained on the Developer Policy corresponding to the social network from which data will be extracted, and the key concepts on data protection issues
- Adequate assessment tools on data protection have been implemented from the very beginning of the project
- The roles played by all different agents involved in the data processing have been adequately clarified through the corresponding agreements and the controller can provide evidence on this.
- The ICT developer is well aware of the terms of use of the data gathered from the social networks
- The representatives of the key collectives involved in the data processing have been consulted on the impact of the use of the gathered data and the concrete social network selected.

3 Gaining access to data. Some essential tips

According to the GDPR, lawful processing requires a legal basis (see the "Lawfulness, fairness and transparency" subsection in the "Principles" within Part II of these Guidelines). If processing involves the type of activities that are included in the ePrivacy Regulation, the provisions made by this new tool will apply as soon as it is approved. At the present moment, Article 6 of the GDPR defines a total of six possible legal bases. In the case of processing data from social networks, it is essential to underline that **ICT researchers or innovators must be aware that they will certainly need different legal bases for data processing at the moment of accessing the data and at the moment they perform their research or innovation based on those data.** In the first case, what is needed is a legal basis to obtain the data from the social network. In the second case, it is a matter of finding a basis that allows the data, already legitimately acquired, to be used for research purposes. **It is essential to note that the mere fact that the data subjects have published their data in online public spaces does not allow for their processing.** These are still personal data, even if the data is publicly available. The publication might serve to avoid the ban included in Article 9.1 of the GDPR, if we are talking about data of special categories, but does not serve as a legal basis for processing. **As such, companies may not freely re-use the data, and may not further process it without the individuals' knowledge and without an adequate basis for lawful processing.**

3.1 Public domain does not mean public data!

T**he concept of "public domain" must be adequately analyzed in the context of social networks.** If the ICT researcher or innovator has had to register with a community of users in order to have access to specific data, these data are not public: they are data that the data subjects have wished to share exclusively with a community of users and under the terms and conditions determined by the social network in question, which are accepted at the moment the users create their profiles. If researchers are able to access a profile or other kinds of social media data on a site simply because they are registered users, this is not the same as that information being publicly available. It is therefore absolutely essential for the ICT researcher or innovator to have a precise knowledge of these terms and conditions, which may differ substantially from one social network to another.

Furthermore, even though the data are in the public domain, this does not at all mean that you could use them for purposes other than those for which they were made public. This is extremely important, since otherwise you could face legal responsibilities.

The Equifax case: using data from the public space does not necessarily legitimate processing

Equifax is a company that obtained data from the information portal used by public administrations to transmit information to citizens. From this data, it created a file that supposedly transmitted information on the solvency of citizens. All this, without informing the data subjects of these processing operations and using the legitimate interest of the company as a basis for legitimacy. On 26 April 2021, the Spanish Data Protection Agency (AEPD) fined Equifax €1 million for breach of data protection regulations, prohibited the continued use of this file, ordered the deletion of all the data of those affected and ordered Equifax to notify all companies that have consulted its file of the content of this Resolution so that they did the same and stopped using this data.

This ruling is of great importance for several reasons. The first is that it is the first major sanction arising from the change in criteria brought about by the GDPR and the national regulation (LOPDgdd) regarding the use of publicly accessible sources: the fact that data is accessible to the public does not mean that it can be used for any purpose and without further explanation. In the previous Spanish law, the 1999 LOPD, this criterion was not so clear and seemed to be the opposite.

In its Resolution, the AEPD recalled that (1) any secondary use of data must be compatible with the original purpose for which they were collected (principle of purpose limitation of data processing, article 5.1.b GDPR), (2) it must have its basis for legitimization (it is not sufficient to allege that the data are from publicly accessible sources), and that (3) the data subject must be notified of the secondary use of his or her data. The fine of €1 million was based on the breach of the purpose limitation principle.

3.2 Gaining access to data from a social network: some essential tips

These are some essential tips provided by the Ethics information for Linguistics and English Language[18] that you must follow if you are planning to gain access to data from a social network:

- If the data are in the public domain, you must abide by any requirements stated by the corpus provider, including with respect to anonymity, or any other conditions on use.
- Some corpora may require ethical approval, especially corpora that include physical or mental health data, or corpora that contain data that could be used to de-anonymize individuals (e.g. when free-text responses are allowed).
- If the data are not in the public domain, you must ensure that your use of the data conforms to any requirements stated by the corpus provider. For example, the data must not be shared in any unauthorized manner (e.g., posted online).
- In either case, if there is reason to suspect that the people who initially provided the data were not aware that it would be used for research purposes, you should carefully consider the ethical implications of your research, including whether you should obtain informed consent.

All these tips can be concreted in the following steps:

- First, always keep in mind the **reasonable expectations of the data subjects about the use of their data (Recital 47, GDPR).** This is essential in most social networks Developers Policies. For instance, Twitter Developer Policy states that "we prohibit the use of Twitter data in any way that would be inconsistent with people's reasonable expectations of privacy. By building on the Twitter API or accessing Twitter Content, you have a special role to play in safeguarding this commitment, most importantly by respecting people's privacy and providing them with transparency and control over how their data is used."[19]

- Second, obtaining approval to access the APIs and the Contents of a social network is never enough to ensure a lawful data processing. It is just the first step. Most of the social networks have developed detailed **Platform Usage Guidelines that researchers must strictly follow to ensure policy compliance for their planned use of the platforms and compliance with data protection ethical and legal requirements.**

- Third, most of the social networks have developed tools that **provide support to researchers** willing to use their Application Programming Interface (APIs). It is always recommendable that researchers use these services in case of doubt on data processing.

- Fourth, however, researchers and innovators should never forget that, as a controller, you are responsible for ensuring that the data protection framework is

[18] https://resource.ppls.ed.ac.uk/lelethics/index.php/frequently-asked-questions/corpus-research/
[19] https://developer.twitter.com/en/developer-terms/policy

adequately respected. Thus, you should check whether the statements on the legitimacy of the data processing carried out by the social networks correspond to reality. Reviewing their data collection policies to check the soundness of the consents granted from a GDPR perspective seems a necessary or, at least, prudent requirement.

- Fifth, researchers/innovators shall keep in mind that social networks **might change their policies** from time to time without notice. Since they usually introduce this caution in their own policies, researchers take responsibility for keeping themselves informed about these possible changes. Thus, periodic reviews of such policies are highly recommended.
- Sixth, since researchers will be processing data that have not been obtained from the data subject, they shall provide the data subject with the information requested by article 14 unless any of the circumstances quoted in its point 5 apply.
- Finally, in case of doubt, always consult your Data Protection Officer and, if necessary, the corresponding Data Protection Authority.

Box: Considering data subjects' expectations and concerns. The twitter case

Most researchers who use data sets of tweets do not gain consent from each Twitter user whose tweet is collected, nor are those users typically given notice by the researcher.

In 2017, Fiesler and Proferes developed an exploratory survey of Twitter users' perceptions of the use of tweets in research. At the time when this research was performed, Twitter's Privacy Policy mentioned that academics could use tweets as part of research. However, few users were previously aware of this fact, and the majority felt that researchers should not be able to use tweets without consent. However, these attitudes were highly contextual, and differed with respect to factors such as how the research was conducted or disseminated, who the researchers were, and what the study was about.

Source: Fiesler C., Proferes N. "Participant" Perceptions of Twitter Research Ethics. Social Media + Society. January 2018. doi:10.1177/2056305118763366

Researchers and innovators who use data obtained from social networks are responsible for complying with all policies settled by those networks. Thus, it is essential that they review and understand these policies before they access the social networks' APIs and contents. The time spent reviewing their policies may save researchers hours of further work down the road and may even help them avoid legal responsibilities.

Checklist. Gaining access to data

- If the data are in the public domain, the controllers have abided by any requirements stated by the corpus provider, including with respect to anonymity, or any other conditions on use.
- If the data are not in the public domain, the controllers have ensured that their use of the data conforms to any requirements stated by the corpus provider.
- The controllers are familiar with the Platform Usage Guidelines that they must strictly follow to ensure policy compliance for their planned use of the platforms and compliance with data protection ethical and legal requirements.
- The controllers have considered the reasonable expectations of the data subjects about the use of their data.
- The controllers have checked whether the statements on the legitimacy of the data processing carried out by the social networks correspond to reality
- The researchers/controllers are aware that they take responsibility for keeping themselves informed about possible changes in the platform policies. Periodic reviews of such policies are performed
- The controllers provide the data subjects with the information requested by article 14 unless any of the circumstances quoted in its point 5 apply

4 Choosing a legal basis for further processing

Once researchers/innovators become the controllers of the data gathered from social networks, they have to decide on the legal basis that will legitimate further processing of those data as soon as possible. However, and even before selecting the legal basis (or bases) for processing, the controller must consider whether the processing involves personal data of special categories. In that case, the controller should be aware of the fact that the processing is vetoed by Article 9.1 of the GDPR unless any of the circumstances described in Article 9.2 apply.

Once concluded that no data of special categories are involved or the veto posed has been adequately addressed, the controller shall select the appropriate legal basis for data processing. This must be done very carefully, since the legal basis cannot be changed during the processing. These are some criteria that should be kept in mind for this purpose:

- The necessity or usefulness of the use of the data obtained from the social networks for the achievement of the purpose or interest of the processing must be sufficiently justified in the lens of the legal basis selected.
- The data controller must carefully weigh up (1) the basis of entitlement used, against (2) the possible risks arising from the data processing.

- In addition, the controller should consider all adequate safeguards so as to ensure that the interests, rights and freedoms of the data subject are adequately preserved. This balancing must be particularly careful if the data subject's consent acts as the legal basis for processing.

The following tables provide brief overviews of the various alternative bases of legitimation under Articles 6 and the circumstances that circumvent the veto created by Article 9.1 of the GDPR and their relation to the processing of data from social media

Consent is the most traditional legal basis for data processing in the context of social networks. However, where a controller seeks to process personal data for research purposes, public interest might be an excellent option. Unfortunately, it requires that certain conditions apply (see "Data protection and scientific research" subsection in "Main Concepts" within Part II of these Guidelines). Legitimate interest, on the other hand, is an alternative suitable legal basis for processing in this context, but one cannot assume it will always be appropriate[20]. It is likely to be most appropriate where controllers use people's data in ways they would reasonably expect and which have the least possible relevant impact on data protection or privacy issues, or where there is a compelling justification for the processing.[21]

Possible Legal bases (Art. 6 GDPR)

Legal bases for processing	**Use in the context of social networks**
6.1.a –consent	Probably, the most popular legal basis for data processing, although its widespread use is increasingly being questioned[22] (see following section)
6.1.e - 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	It may be applicable, but the following cautions should be observed: - The public interest purpose must be clearly identified as well as the connection to the research, - Reasons must be given as to why the use

[20] Ad ex., Public authorities can only rely on legitimate interests if they are processing for a legitimate reason other than performing their tasks as a public authority, so "public task" is a better legal basis in these situations (ICO: Legitimate interests, at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/).

[21] ICO: Legitimate interests, at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/

[22] See, on data processing for health purposes in the American privacy system, Charlotte A. Tschider, ´The consent myth: improving choice for patients of the future´ (2019) 96 Washington University Law Review 1506.

	of data from social media is necessary or highly desirable for the objectives pursued. -The basis for the processing has been laid down by Union law; or a Member State law to which the controller is subject.
6.1.f - processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child	It may be applicable, and indeed is the best alternative to consent as a basis for legitimacy. The following cautions should be observed: - the data controller must carry out and give reasons for an appropriate balancing of (1) the legitimate interest pursued and (2) the impact on the fundamental rights and freedoms of the data subject; this balancing must be carried out with particular care if data from minors are involved.

Special categories of personal data (Art. 9 GDPR)

Basis for legitimacy	**Use in the context of social networks**
9.1.a –consent	It is widely used
9.2.e - processing relates to personal data which are manifestly made public by the data subject	It may be applicable, but particular caution should be taken with regard to the following safeguards: - respect for the purpose limitation principle (Art. 5.1.b GDPR), taking into account the expectations of the data subject and the context (social network and impact of the profile) in which the data has been published[23]; - measures of aggregation in order to lower possibilities of re-identification.
9.2.g - processing is necessary for	It may be applicable, provided that the data

[23] Recently, the Spanish Data Protection Board fined Equifax for using creditworthiness data published by official sources to feed its own files, for breach of the purpose limitation principle insofar as it is an incompatible use of the data despite being publicly accessible data. The criterion of this Resolution may also be applicable if data published by the data subject itself is used, insofar as the uses derived from such data are incompatible.

reasons of substantial public interest, on the basis of 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	controller observes the following precautions: - the public interest pursued must be clearly identified, as well as the applicable regulations; - it must be sufficiently justified that the research via social networks is necessary or highly suitable for this purpose; - special care must be taken to develop measures to protect against undue impacts on fundamental rights of data subjects.
9.2.j - 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	It is fully applicable. It has the advantage that the purpose limitation principle is less strict (cf. Art. 5.1.b GDPR) and that it allows the processing of data independently of the consent of the data subjects, provided that the data controller observes the following safeguards: - it must clearly identify its purpose (archiving, scientific research, historical research or statistical purposes); - it must justify the proportionality of the data processing in relation to the intended purpose; - it should justify the usefulness of the use of social networks in the research; - must develop measures to avoid undue impacts on fundamental rights of data subjects, focusing on (1) sufficient level of aggregation, and (2) other safeguards to avoid re-identification - must strictly follow the prescriptions of art. 89 GDPR

4.1 Consent

Consent is the first of six bases for lawful processing of personal data listed in Article 6. According to Article 6, paragraph. 1(a)[24], such processing is lawful if data subjects have given consent to the processing of their personal data for one or more specific purposes. Thus, if data is used for multiple purposes, consent shall be given for each purpose separately. Specific consent is key to avoiding invalid consent. Indeed, "if a data processing has multiple purposes, then consent must be sought for each of them.

[24] EDPB: Guidelines 05/2020 on consent under Regulation 2016/679, https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

Specificity of consent promotes transparency as the data subjects know about each purpose of data processing, increases their control over these purposes and safeguards against function creep."[25]

The specificity requirement is particularly important in the case of the re-use of data from social networks. End users of social networks are often unaware of the fact that their data are used for purposes other than those that they pursue when they provide those data. However, most social networks ensure that the data subjects provide consent to this further processing and their Developer Policies will surely cover this issue. Researchers and developers willing to process the data obtained from social networks for research purposes might obtain a new consent from the data subjects. This, of course, is hard and not always necessary. They could rely on the original consent provided by the data subject to the social network. However, **the researchers/innovator should, however, ensure that the processing they are willing to perform is allowed by the consent originally provided by the data subject or find an alternative legal basis (by asking for a new consent or using legitimate interest or public interest as an alternative, for instance).** Consulting the terms of use of the social network and the consent gathered originally is an excellent way to check if the secondary use of data could be considered compatible with the purposes for with data were originally collected (see the "Purpose limitation principle" subsection in the "Principles" within Part II of these Guidelines).

If the research involves using **data gathered from different social networks**, researchers should focus on **designing intra-provider and eventually inter-provider privacy risk evaluation mechanisms that take into account personal data revealed for all data processing activities for a concrete social network and for all OSNs that a data subject uses, respectively.**

Last, but not least, since researchers will be processing data that have not been obtained from the data subject, they shall provide the data subject with the information requested by article 14 unless any of the circumstances quoted in its point 5 apply (see the "Right to information" subsection in the "Data Subject Rights" section within Part II of these Guidelines).

[25] Joyee De S., Imine A. (2019) On Consent in Online Social Networks: Privacy Impacts and Research Directions (Short Paper). In: Zemmari A., Mosbah M., Cuppens-Boulahia N., Cuppens F. (eds) Risks and Security of Internet and Systems. CRiSIS 2018. Lecture Notes in Computer Science, vol 11391. Springer, Cham. https://doi.org/10.1007/978-3-030-12143-3_11

Box: the case of deleted data

Some social networks users post data to their platforms and subsequently delete it. If that data has been retrieved by a researcher before deletion, it is not clear whether the user's initial consent for their data to be used remains intact. Depending on the sensitivity of the data and analysis researchers should agree up-front how to manage this issue. For example, it may not be necessary to delete the count of a post from a time series, but it may be unethical to quote an individual post which has since been deleted. However, this is as yet an unclear issue. Therefore, researchers should still be cautious about the use of deleted data.

See: Social Media Research Group, Using social media for social research: An introduction May 2016, p. 17 at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524750/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf

Checklist: consent

 Controllers are able to demonstrate that, after balancing the circumstances of the processing, they have concluded that consent is the most appropriate legal basis for processing.

 Controllers have made sure that the consent provided by the data subject to the social network covers the type of processing they are willing to perform.

□ If this is not the case, controllers must ask data subjects for a renewed consent.

4.2 Legitimate interest

Legitimate interest constitutes an alternative basis for lawful processing that might be applicable to the use of data gathered from social networks, even though public authorities ***cannot*** rely upon this basis when acting. For those who can use this legal basis, three cumulative conditions should be met[26]:

(i) the pursuit of a legitimate interest by the data controller or by the third party, or parties, to whom the data are disclosed,

(ii) the need to process personal data for the purposes of the legitimate interests pursued, and

[26] 9 CJEU, Judgment in Fashion ID, 29 July 2019, C-40/17, para. 95 - ECLI:EU:C:2019:629.

(iii) the condition that the fundamental rights and freedoms of the data subject whose data require protection do not take precedence.

Thus, in principle legitimate interest could be the perfect legal basis for processing in this context. However, there are some good reasons to consider that this basis will not always apply to the use of data for scientific research:

- First, legitimate interest should apply to all joint controllers, in case that joint controllership applies to the processing. In the Fashion ID case, the CJEU specified that in such circumstances "it is necessary that each of those controllers should pursue a legitimate interest […] through those processing operations in order for those operations to be justified in respect of each of them".
- Second, controllers should be able to demonstrate that the balancing test has been adequately performed (see "Legitimate Interest and Balancing Test" section in the "Main Tools and Actions" section within Part II of these Guidelines). This means that joint controllers are able to establish that the processing is necessary to achieve those legitimate interests. This is hard to reach, since "necessary" requires a connection between the processing and the interests pursued. This means that it should be considered whether other less invasive means are available to serve the same end. Similarly, processors should be able to demonstrate that their legitimate interests at stake are not overridden by the data subject's interests or fundamental rights and freedoms. This is all hard to demonstrate, especially if minors are involved in the processing. [27]
- Third, legitimate interest could hardly apply as a legal basis for lawful processing if such processing involves intrusive profiling and tracking practices, for example those that involve tracking individuals across multiple websites, locations, devices, services or data-brokering.[28]
- Fourth, instead, if we are considering data pertaining to data subjects who have already had a previous relationship with the ICT researcher and innovator through the social network, using legitimate interest as a legal basis seems pretty reasonable. However, controllers should take into consideration if the previous relationship was similar to the one that is about to be built.

If legitimate interest is finally chosen as the legal basis for processing, controllers shall keep in mind that the duties of transparency and the **right to object** require careful

[27] See Article 29 Working Party Opinion 06/2014 on the concept of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 9 April 2014 https://ec.europa.eu/justice/Article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

[28] Article 29 Working Party, Opinion on profiling and automated decision-making, WP 251, rev. 01, p. 15, see also Article 29 WP, Opinion on legitimate interest, p. 32 and 48: « Overall, there is an imbalance between the company's legitimate interest and the protection of users' fundamental rights and Article 7(f) should not be relied on as a legal ground for processing. Article 7(a) would be a more appropriate ground to be used, provided that the conditions for a valid consent are met ».

consideration. **Data subjects should be given the opportunity to object to the processing of their data for targeted purposes before the processing is initiated.** Users of social media should not only be provided with the possibility to object to the processing when accessing the platform, but should also be provided with controls that ensure the underlying processing for specific purposes of their personal data no longer takes place after they have objected to the processing.[29]

Checklist: legitimate interest

 The controllers have checked that legitimate interest is the most appropriate basis for processing.

 The controllers have checked that the processing is necessary and there is no less intrusive way to achieve the same result.

 The controllers have done a balancing test, and are confident that the individual's interests do not override those legitimate interests.

 The controllers are not using people's data in ways they would find intrusive or which could cause them harm, unless there is a very good reason.

 If the controllers foresee the processing of children's data, they have taken extra care to make sure that legitimate interest is a suitable database.

 The controllers have considered safeguards to reduce the impact where possible.

 The controllers have introduced adequate tools to ensure that the right to object is easy to implement by the data subjects.

 If the controllers have identified a significant personal data protection impact, they have considered whether they also need to conduct a DPIA.

 The controllers include information about their legitimate interests in their privacy information.

4.3 Public interest and the scientific research framework

According to Article 6 (e) of the GDPR, processing is lawful if it is necessary for the performance of a task carried out in the public interest. Here, one must keep in mind

[29] Guidelines 8/2020 on the targeting of social media users Version 2.0 Adopted on 13 April 2021, at: https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf, p. 11)

that "scientific research" is an overly broad term that generally refers to the search for knowledge, through a certain methodology, in any area of human knowledge. Thus, it is quite probably that if controllers are using a scientific methodology and, somehow, searching for knowledge thought the use of data, such processing could be lawful on the basis of the public interest legal ground.

Furthermore, public interest could serve to skip the veto included in Article 9.1 of the GDPR if they are using special categories of data when other legal bases (such as research for instance) are not applicable to the case. However, in this case, the processing shall be based on the law of the EU or a Member State and shall be proportionate to the aim pursued, respect the essence of the right to data protection, and shall provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject[30] (see the "Data protection and scientific research" subsection in the "Main Concepts" section within Part II of these Guidelines**).**

On the other hand, one must remember that Article 5 (b) GDPR establishes the purpose limitation principle, under which data cannot be processed for purposes other than the specific initial ones. Interestingly, this article provides that certain purposes, which include scientific research, are deemed compatible with the initial purpose, rendering its subsequent processing presumptively lawful. Therefore, where the controller can argue and document that the purpose of the processing is scientific research, **secondary uses of personal data are in principle considered compatible with the original purpose of the personal data processing** (see the "Purpose limitation principle" subsection in the "Principles" section within Part II of these Guidelines).

Moreover, it is quite probable that the social network that originally gathered the data included a clause in the consent by the data subject that allowed it or a third party further processing for research purposes or, at least, informed the data subject that such processing would be considered compatible with its initial consent. If this were the case, processing for research purposes would be legitimate on the same lawful basis that permitted the social network gathering the data.

This assessment, however, needs to be carried out prior to the subsequent processing for secondary purposes and must be based on objective criteria. The legal framework on this issue might change considerably between the EU Member States. Thus, controllers should be aware of the applicable concrete normative framework. Consultation with their DPOs is highly recommended for this purpose[31] as well as the inclusion of an ethical-legal advisor/unit within the given project.

Checklist: scientific research

- The controllers have checked that their project fits well with the concept of scientific

[30] See EOSC-Pillar Guidelines ´D4.1: Legal and Policy Framework and Federation Blueprint´ (2021), p. 76-77. At: https://repository.eosc-pillar.eu/index.php/s/tbqe6B7rDycdFCJ#pdfviewer

[31] See some practical questions and answers on this here: https://www.ru.nl/rdm/gdpr-research/faq-gdpr-research/

research.

 The controllers have consulted their DPOs about the use of this exception to the ban on the processing of data of special categories.

 The controllers have consulted the national legal framework about this topic.

 The controllers have implemented the safeguards and organizational measures devoted to align with article 89 of the GDPR and corresponding national regulation.
 The controllers have documented all the information regarding this issue in the DPIA

5 Fairness and Transparency issues

Fairness is an essential principle in the GDPR. Arguably, all of data protection and thus the GDPR is about fairness towards data subjects. The GDPR can be seen in spelling out what *fair* actually means. In the case of data gathered through the use of social networks, it is particularly important to avoid biases related to gender, race, age, sexual orientation, national origin, religion, health and disability, etc. This might be problematic since it is possible that some of the data gathered via social networks do not correspond to real users, or their sensitive data are not at all accurate. This might create hidden biases (see the "Lawfulness, fairness and Transparency" subsection of the "Main Concepts" section within Part II of these Guidelines).

Transparency, on the other hand, is a main strategy to balance power between controller and data subject. It works by pulling everything into the light and thus opens it up to scrutiny. The main focus of transparency is to inform **data subjects** up-front of the existence of the processing and its main characteristics. Other information (such as the data about the data subject) is available on request. Data subjects also have to be informed of certain events, most notably data breaches (in the case where the data subject is exposed to high risk). Evidently, transparency is a pre-requisite for detecting and intervening in case of non-compliance see the "Lawfulness, fairness and Transparency" subsection of the "Main Concepts" section within Part II of these Guidelines).

In the case of using data from social networks, transparency means, in our opinion, that "intended research subjects should be informed at some point about the research being performed, what sort of personal data controllers are collecting and how it will be used. Some services make it clear this must be done before you start harvesting. For others without a specific policy and where researchers/innovators are conducting observational research which obtaining consent up front could damage, they should let the individuals concerned know as soon as possible. The ICT researchers/innovators should always remove individuals from their harvesting who do not consent to being included."[32]

[32] https://info.lse.ac.uk/staff/divisions/Secretarys-Division/Assets/Documents/Information-Records-Management/Social-media-personal-data-and-research-guidance-v.1.pdf

In the case of using data from social networks, it is necessary to point out that, in general, Article 14 of the GDPR will be applicable at some point. Thus, data subjects should be made fully aware that their data is being shared with third parties (see the "Right to information" subsection in the "Data Subject Rights" section within Part II of these Guidelines). This could be done in different ways. For instance, the CNIL advised that data controllers could either include all third-parties in an exhaustive privacy notice, but periodically updated, or insert a link in this notice and redirect individuals to the list with the third-parties and their own privacy policies.[33]

Controllers shall guarantee transparency not only by providing adequate information, but also by using a number of **complementary tools.** Appointing a DPO, who then serves as a single point of contact for queries from data subjects, is an excellent option. Preparing adequate records of processing for the supervisory authorities, or performing DPIAs, are also highly recommended measures to promote transparency. Likewise, undertaking analysis that evaluate the effectiveness and accessibility of the information provided to the data subjects helps to ensure the efficient implementation of this principle[34].

Last but not least, implementing the so-called Transparency Enhancing Tools (TETs)[35] might be an excellent option to guarantee that the Transparency principle rules, especially when massive or automated data processing is expected.

5.1 Biases

Biases create prejudice and discrimination against certain groups or people. Harm can also result from the intentional exploitation of (consumer) biases, or by engaging in unfair competition, such as the homogenization of prices by means of collusion or a non-transparent market. Using data gathered through social networks could contribute to exacerbate such a situation mainly by building biased datasets. This might happen, for instance, due to an inadequate collection of the data produced by the data subjects. "Social media data can be difficult to verify – users may lie about their age, location, job, or any number of other characteristics. **Researchers must be aware of this issue and address this difficulty where relevant.** It is not advisable to understand users as the 'general public', due to inequalities in access to the internet, and researchers should consider how to foster diversity (where relevant) in their sample."[36] It might also happen that inferred or derived data create such biases due to their own technical issues.

[33] https://www.cnil.fr/fr/transmission-des-donnees-des-partenaires-des-fins-de-prospection-electronique-quels-sont-les

[34] See EOSC-Pillar Guidelines ´D4.1: Legal and Policy Framework and Federation Blueprint´ (2021), pp. 44 et seq. At: https://repository.eosc-pillar.eu/index.php/s/tbqe6B7rDycdFCJ#pdfviewer

[35] TETs can be subdivided into 'ex ante' and 'ex post'- TETs. Ex ante-TETs guide the user's decision-making process before she makes her choice pertaining to disclosing any personal data to a data controller. Conversely, ex post-TETs visualize disclosed personal data in such a way as to make transparent the processes that have taken place since the user has disclosed her data (see P. Murmann; S. Fischer-Hübner, `Usable Transparency Enhancing Tools – A Literature Review´ (2017), working paper. At: http://www.diva-portal.org/smash/get/diva2:1119515/FULLTEXT02.pdf).

[36] University of York, **Guidelines for the Use of Social Media Data in Research**, at: https://www.york.ac.uk/staff/research/governance/research-policies/social-media-data-use-research/

If these biased data fuel profiling or automated decision-making, this could bring unacceptable social consequences. Of course, if the research involves the use of AI, this will probably increase the risk related to biases (see the "Lawfulness, fairness and Transparency" subsection of the "Main Concepts" section within Part II of these Guidelines).

In order to avoid such a scenario, **critical assessment of the provenance of the data is required**. To this purpose, organizational measures should be implemented to guarantee the accuracy and reliability of the gathered data, while still ultimately deferring to the right of users to withhold private information (e.g. confirming whether or not a record is accurate). Furthermore, performing an audit devoted to detecting biases in raw data or in the inferred or derived datasets is required especially when controllers are using datasets produced via social networks.

5.2 Transparency

Research bases on data gathered via social networks often involves processing a lot of personal data. This creates a complex scenario. Controllers must be aware that, even though it might be hard to achieve, data subjects must be able to understand how, and for what purpose, their personal data is being used. In general, this means that **the researchers should use tools able to provide such knowledge in the easiest possible way.** Explainability is particularly important in the case of automatic processing of data or profiling. "The methods for giving information, offering a right to refuse or requesting consent **should be made as user-friendly as possible.** Therefore, information policies must focus on information which is understandable by the user and should not be confined to a general privacy policy on the controllers' website".[37]

If the controller "plans" to carry out a processing for purposes other than those for which the data were collected from the social network, they must inform users or data subjects beforehand of such further processing, providing information and complying with all other requirements, such as having a legal basis for this new purpose or carrying out a compatibility assessment (see the "Purpose limitation principle" subsection of the "Main Concepts" section within Part II of these Guidelines). Of course, the requirements of transparency are clearly related to the fairness principle, since the harder it is for the user to understand data processing, the greater the difference between different types of users. In general, "the larger the amount of data, the harder is a clear, intelligible overview in text form. Symbols offer a way to represent personal data categories in a lean and recognizable way. This requires meaningful and self-explanatory graphical representations of the data."[38]

According to the GDPR, the information that a controller must provide to the data subjects varies depending on whether this information has been obtained from them or

[37] Art 29 Data Protection Working Party (2014) Opinion 8/2014 on the on Recent Developments on the Internet of Things (SEP 16, 2014) https://www.dataprotection.ro/servlet/ViewDocument?id=1088

[38] Bier C., Kühne K., Beyerer J. (2016) PrivacyInsight: The Next Generation Privacy Dashboard. In: Schiffner S., Serna J., Ikonomou D., Rannenberg K. (eds) Privacy Technologies and Policy. APF 2016. Lecture Notes in Computer Science, vol 9857. Springer, Cham. https://doi.org/10.1007/978-3-319-44760-5_9

not. If the personal data is not obtained from the user (Art. 14 GDPR), such as in the case of receiving the data from a social network, the controller must be particularly attentive to providing the data subject with adequate information, especially since massive data gathering is being performed. Thus, controllers shall inform the user of the provisions of Art. 14 of the GDPR[39].

It is necessary, however, to mention that sometimes it might be extremely difficult for controllers who have gathered the data from a social network to inform data subjects about the processing. If this is the case, they might recall article 14.5 (b), which states that "the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 89(1) or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing" (see "Data protection and scientific research" subsection of the "Main Concepts" section within Part II of these Guidelines).

In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available (see the "Right to information" subsection of the "Data Subject Rights" section within Part II of these Guidelines). Thus, in principle controllers could avoid providing information about the processing to the data subjects if this is rendered impossible, but only if they *take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.*

Note with caution however, that disproportionate effort may in some jurisdictions be interpreted narrowly. For example, there was a recent decision (March 2019) by the Polish Data Protection Authority (Polish DPA) when it fined a data scraping company €220k for its failure to provide privacy notices to 5.7 million individuals whose data was scraped from a public register. The Polish DPA rejected the argument that placing a privacy notice on the data scraping business' website was enough to notify individuals, particularly where individuals were not aware that their data had been scraped and was being processed.[40]

Checklist: fairness and transparency
Fairness
 The controllers perform audits aimed at detecting biases in the datasets built and/or the conclusions of the analysis

[39] See: CNIL, La réutilisation des données publiquement accessibles en ligne à des fins de démarchage comercial, at: https://www.cnil.fr/fr/la-reutilisation-des-donnees-publiquement-accessibles-en-ligne-des-fins-de-demarchage-commercial

[40] Campbell, Fiona, Data Scraping – Considering the Privacy Issues, at: https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/data-scraping-considering-the-privacy-issues

 The controllers have implemented adequate measures to avoid biases provoked by the use of AI tools.

Transparency

 The controller provides

- a panoramic *overview of what personal data have been disclosed to what data controller under which policies;*
- *online access to the personal data and how they have been processed;*
- *counter profiling capabilities helping the user to anticipate how their data match relevant group profiles, which may affect future opportunities or risks.*

 Since the personal data were not provided by the data subject, the controllers provided all the information listed in Article 14.1 GDPR;

 Since the personal data is not provided by the data subject, the information is provided:

- within a reasonable period after obtaining the personal data, but at the latest within one month;
- if the personal data are to be used for communication with the data subject, at the latest at the time of the first communication to that data subject;
- if a disclosure to someone else is envisaged, at the latest when the personal data are first disclosed.

 The information is provided concisely, transparently, intelligibly, and in an easily accessible way. It is clear and redacted in plain language.

 If providing the information is rendered impossible, the controllers take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.
 The controllers have documented all the information regarding these issues.

6 Data governance: minimization, purpose limitation and storage limitation principles

The minimization principle (see the "Minimization principle" subsection of the "Main Concepts" section within Part II of these Guidelines) states that personal data shall be **adequate**, **relevant and limited to what is necessary in relation to the purposes** for which they are processed. On the other hand, according to Article 5(1) (e) of the GDPR, personal data should be "kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed". Finally, purpose limitation means that personal data cannot be processed for purposes other than the ones stipulated in the privacy policy when the data were collected, unless these further purposes are compatible with the original purposes and pursuant to appropriate safeguards (art. 6.4 GDPR). For instance, the further processing corresponds to archiving activities of public interest, purposes of scientific and

historical research or statistical purposes (see the “Data processing and scientific research” subsection of the “Main Concepts” within Part II of these Guidelines).

The combination of these three principles creates a combined normative tool that must be strictly followed by controllers using data gathered through social networks. In general, controllers[41] must make the purposes of the processing explicit: "disclosed, explained or expressed in an intelligible form". In line with the principle of data minimization, they should also identify the minimum amount of personal data needed to achieve their objectives. In addition, in respect of the accountability principle, data controllers should be able to demonstrate that they only collect and hold the personal data needed, and that it is used solely for the specific purposes that have been informed under an adequate legal basis.

Summarising, setting clear objectives for the processing will help ensure that the personal data to process are:

- adequate: sufficient to fulfil the stated purpose;
- relevant: they should have a rational link to the purpose;
- limited to what is necessary: they should not hold more data than those needed for the stated purpose.

6.1 Minimization principle

The minimization principle states that personal data shall be **adequate**, **relevant and limited to what is necessary in relation to the purposes** for which they are processed. (see the “Data minimization” section in the “Principles” section within Part II of these Guidelines). According to this principle, controllers should be aware of the goal that is to be reached through the processing, so as to avoid using more data than need be. Furthermore, controllers should also try to avoid using special categories of personal data if they are not strictly necessary.

When researchers/innovators gather data from social networks, they might end up processing far more personal and sensitive data than they really need for the specific purposes of the research. There are some ways through which such a scenario might be avoided. In principle, controllers **should promote the use of anonymized data** (see the see “Identification, Pseudonymization and Anonymization” within Part II section “Main concepts” of these Guidelines). Indeed, avoiding the identification of specific individuals from big data analytics, or the re-identification of data users whose data has been pseudonymized, is a fundamental safeguard to prevent undue impact on data subjects caused by data processing[42]. If they do not need personal data, they could ask

[41] it is important to identify who the "data controller" is; developers are rarely the "data controller", since they are not responsible to take care of the business objective, this is a task for the management of the company.

[42] WP29 Guidelines 3/2013 on purpose limitation (p. 3) highlight the adoption of safeguards to prevent undue impacts on data subjects as a key factor to take into account when evaluating the compatible further uses of data.

the social network to provide them with anonymized data. Of course, they could also anonymize the data once gathered, but, in this case, they should not forget **that anonymization involves data processing and, thus, they would need to have a legal basis that legitimates it** (see the "Identification, Pseudonymization, and Anonymization" subsection in the "Main Concepts" section within Part II of these Guidelines).

Furthermore, researchers/innovators should keep in mind that anonymization might be hard to reach. Quite often, aggregation and inferring data practices can easily de-anonymize datasets. Thus, **controllers should not presume that their anonymization processes will serve well to preserve data subjects' privacy. Indeed, they should perform DPIAs and risk assessments to ensure such a belief** (see accountability in this part of the Guidelines)

An alternative to anonymization as such is the use of **aggregated data**. In the context of data protection, two kinds of aggregation have to be distinguished (see the "Data minimization" section in the "Principles" section within Part II of these Guidelines):

- **Single Person**: Aggregation of data elements pertaining to a **single person**: Taking for example a person's average monthly income over a year reduces the information content pertaining to that person.
- **Multiple Persons**: Aggregation of data elements pertaining to a **multitude of persons**: Taking for example the average yearly income over group of persons also reduces the overall information content (data minimization). In addition, it also weakens the degree of association between a data element and a given person. This kind of aggregation is therefore also pertinent to storage limitation

When the purpose of the processing can be achieved using aggregated data, this is recommendable (see the "Data minimization principle" subsection of the "Principles" section within Part II of these Guidelines). Under such circumstances, no one but the data subject should access the raw data (obtained or observed data), unless an extremely relevant reason applies (for example, national security issues interpreted restrictively). Indeed, sometimes a specific research only needs aggregated data and has no need of the raw data collected in the social networks. Therefore, **controllers must delete raw data as soon as they have extracted the data required for their data processing**. As a principle, deletion should take place at the nearest point of data collection of raw data (e.g. on the same device after processing).

6.2 Purpose limitation

The purpose limitation principle (see the "Purpose limitation" subsection "Principles" section within Part II of these Guidelines) requires that personal data collected are processed only for the purpose for which they were gathered. Purpose limitation is a key concept when processing data obtained from social networks and most platforms include it in their Developer Policies. Researchers and innovators shall strictly follow such policies. On the other hand, it is often true that data subjects are not truly aware of the permissions they provide the social networks for the processing. This is a

particularly important reason why controllers using those data should not process the data for purposes that could be considered as incompatible with the initial consent.

Thus, **controllers should implement tools able to ensure that processing does not take place if the data subjects do not provide their consent, unless an alternative legal basis allows for the processing (**see the "Lawfulness, fairness and Transparency" subsection of the "Principles" within Part II of these Guidelines). The utility of stored data for the intended purpose of research will need to be periodically reassessed to avoid unlawful data processing.

It should be noted that when data are used for reasons of public interest or for research, archiving or statistical purposes, these derived uses will not be considered incompatible with the initial purposes, as long as they are properly pseudonymized, whenever that further processing of such data does not allow re-identification of the data subjects (Art. 5.1.b & 89.1 GDPR) (see the "Consider if the regulatory framework regarding scientific research applies to the activity" section in this part of the Guidelines).

6.3 Storage limitation

The principle of storage limitation obliges data controllers not to store personal data for 'longer than is necessary for the purposes for which the personal data are processed' and to introduce pseudonymization and anonymization measures that reduce/eliminate the identifiability of data subjects as soon as possible for such purposes. The problem here is that usually social networks might use the stored data for different purposes. Furthermore, sometimes data are collected and stored "just in case" they might serve for some future use. Controllers should be aware that even though the GDPR allows storage for longer periods, **there should be a good and real reason to opt for such an extended period (**see the "Storage limitation principle" subsection in "Principles" section within Part II of these Guidelines). That is, a controller should not be tempted to keep the data longer than strictly needed, with the aim of having it available in case novel purposes or projects arise in the future, different to those lawfully permitted.

In order to avoid unlawful storage, a necessity test must be carried out by each and every stakeholder in the provision of a specific service in the social network, as the purposes of their respective processing can in fact be different. For instance, personal data communicated by users when they subscribe to a specific service in the social network should be deleted as soon as they put an end to the subscription. Similarly, information deleted from their account by the users should not be retained. When users do not use the social network for a defined period of time, the user profile should be set as inactive. After another period of time the data should be deleted. The users should be notified before these steps are taken, with whatever means the relevant stakeholder has at its disposal.[43]

To sum up, if controllers do not need the data, and there are no compulsory legal reasons that oblige them to conserve the data, they should fully anonymize or delete

[43] Art 29 Data Protection Working Party Opinion 8/2014 on the on Recent Developments on the Internet of Things (SEP 16, 2014) https://www.dataprotection.ro/servlet/ViewDocument?id=1088

them. Researchers should consult their DPOs if they wish to store data for a longer period of time and be aware of the applicable national regulation.

This could also be an excellent moment to **envisage time limits for erasure of the different categories of data, and document these decisions clearly** (see the "Accountability principle" subsection in the "Principles" section within Part II of these Guidelines). In this regard, the appropriate balance between sustainability of research, reproducibility, open data, open science and principle of minimization under GDPR must be preserved, considering also that the processing of pseudo/anonymized datasets could generate pseudo/identifiable datasets. For this purpose, the criteria set out in Rec. 156 GDPR should be followed:

(1) the processing of personal data for scientific research purposes must be subject to appropriate safeguards for the rights and freedoms of the data subject where it is ensured, in particular that technical and organizational measures are implemented to respect the principle of data minimization;

(2) further processing of personal data should take place where the controller has assessed the feasibility of fulfilling those purposes by means of data processing which does not allow identification of data subjects or which provides sufficient guarantees of pseudonymization;

(3) the conditions and safeguards in question may include specific procedures for data subjects to exercise their rights, as well as technical and organizational measures to minimize the processing of personal data in accordance with the principles of proportionality and necessity.

Checklist: data governance
Minimization

 The controller only processes anonymized or pseudonymized data whenever possible.

 The controller processes the minimal amount of data necessary to reach the pursued goals.

 The controller only processes data of special categories if it is strictly necessary

Purpose limitation

 The controllers only use the data for the purposes they were collected, unless a legal basis allows their lawful processing.

Storage limitation

 Controllers do not store personal data for 'longer than is necessary for the purposes for which the personal data are processed'.

 Controllers check the utility of the stored data for the intended purpose of the

research.

- Data are stored in a way that hinders personal data processing as much as possible.
- The controllers have documented all the information regarding these issues.

7 Accountability and oversight

The accountability principle in the GDPR is risk-based: the higher the risk of data processing to the fundamental rights and freedoms of data subjects, the greater the measures needed to mitigate those risks (See the "Accountability Principle" subsection in the "Principles" section within Part II of these Guidelines)[44]. Since the processing of personal data gathered from social networks might be considered as high risk,[45] the researchers/innovators shall also appoint a DPO and perform a DPIA. Also, controllers should create a Data Protection Policy that allows **the traceability of information** (See the "Accountability Principle" subsection in the "Principles" section within Part II of these Guidelines).

7.1 Data Protection Officer

In most cases, ICT research based on data from social networks involves operations that, by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of data subjects on a large scale. Therefore, the appointment of a DPO is compulsory according to the conditions settled by Article 37(1) apply. Even if this is not the case, **it is always recommendable to proceed to do so, at least in terms of transparency** (see the "Lawfulness, fairness and transparency Principle" subsection in the "Principles" section within Part II of these Guidelines)**.**

7.2 Data protection Impact Assessment

Performing a DPIA is often compulsory in the case of social networks since it involves a systematic monitoring of a publicly accessible area on a large scale (Article 35(3) of the GDPR). Even if this were not the case, some other circumstances could make it compulsory or, at least, highly recommendable (see the "Data Protection Impact Assessment" subsection of the "Main Tools and Actions" section within Part II of these Guidelines).

---

[44] See Articles 24, 25 and 32 of the GDPR, which require controllers to take into account the "risks of varying likelihood and severity for the rights and freedoms of natural persons" when adopting specific data protection measures.

[45] See, in particular, Article 35(3)(a), according to which data processing is considered as high risk in cases of, inter alia, "a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person".

Checklist

- The controller has conducted a DPIA for the processing activity. The controller has ensured that it:
  - Has started as early as possible (following the principle of Data Protection by Design).
  - Has provided a clear overview of what a DPIA is.
  - Has used the guidance and templates provided by the competent Data Protection Supervisory Authority (DPA) where possible. If not (for example, if the DPA does not provide such material or has to cater to many areas of competence of different DPAs), the DPA has followed the guidance provided by the Article 29 Working party in wp248rev.01.
  - Has assembled the team necessary to conduct the DPIA.
  - Has considered ways of facilitating your work.

7.3 Design your Privacy Policy and prepare the documentation of processing

The Privacy Policy is the public document that explains how a research project processes personal data and how it applies data protection principles, according to articles 12-14 of the GDPR. All data subjects must have access to this Privacy Policy. It should be documented. A non-official, but recommendable template can be found here: https://gdpr.eu/wp-content/uploads/2019/01/Our-Company-Privacy-Policy.pdf

Controllers must always keep in mind that, in the case of data gathered from social networks, they might end up mixing different datasets or create inferred or derived data. The traceability of the processing, the information about possible re-use of data, and the use of data pertaining to different datasets in either the same or different stages of the life cycle must be ensured by the records. Whoever processes personal data (including both controllers and processors) needs to document their activities primarily for the use of qualified/relevant Supervisory Authorities. This must be done through records of processing activities that are maintained centrally by the organization across all its processing activities, and additional documentation that pertains to an individual data processing activity (see the "Documentation of processing" subsection in the "Main Tools and Actions" section within Part II of these Guidelines).

The first stages of the project development are the perfect moment to set up a systematic way of collecting the necessary documentation, since it will be the time when the organization conceives and plans the processing activity[46].

[46] Article 25(1) of the GDPR calls this "the time of the determination of the means for processing".

Last but not least, controllers must keep in mind that ethics committees will probably play a key role in personal data processing. However, this might change considerably between sectors and countries. Controllers shall ask their DPO about this topic.

Finally, controllers shall not forget that there might be ethical implications beyond legal compliance. Consultation with an expert in the ethics of social networks is always recommended.

Checklist. Privacy Policy

- The controller has contacted the office/person who is keeping the processing records for the organization.
  - If necessary, the Data Protection Officer can help establish this contact.
- The controller has informed the above office/person early on of the intention to process personal data.
  - This processing activity needs to be entered in the records before processing starts.
- The controller has followed the instructions on:
  - what information is needed to provide for the processing records,
  - when the controller needs to send updates of this information.

Additional documentation pertaining to a single processing activity).

The following items must be documented:

- Assessment of whether the processing activity results in a high risk to the rights and freedoms of natural persons.
- A Data Protection Impact Assessment where the above assessment yields an affirmative result.
- Potential consultation of the competent supervisory authority prior to processing.
- Requirements and acceptance tests for the purchase and/or development of the employed software, hardware, and infrastructure.
- Implemented technical and organizational measures.
- Regular testing, assessing and evaluating the effectiveness of technical and organizational measures
- Requirements and acceptance tests for the selection of processors.
- Contracts stipulated with processors.
- Possible inspections and audits of the processor.

- Method to collect consent.
- Demonstrations of individual expressions of consent.
- Information provided to data subjects.
- Implementation of data subject rights.
- Actual handling of data subject rights.
- Possible breach notifications to the competent supervisory authority.
- Possible communication of data breaches to concerned data subject.
- Any other communication with the competent supervisory authority.

8 Integrity and confidentiality

According to the GDPR, personal data shall be processed in a manner that **ensures appropriate security** of the personal data, including protection against **unauthorized** or **unlawful processing** and against **accidental loss, destruction** or **damage**, using appropriate technical or organizational measures ('*integrity and confidentiality*'). (See the "Integrity and confidentiality" subsection in the "Principles" section of the General Part of these Guidelines).

This principle involves three main issues: integrity, confidentiality and availability. Availability and integrity are somehow linked, since only data that are adequately preserved can be made available to the data subject. Confidentiality, instead, is a more complex issue that deserves complex measures due to the kind of processes involved and the risks inherent to such processes.

8.1 Availability and integrity

Research fuelled by the use of data from social networks sometimes involves gathering an impressive amount of data. The processing of these data usually takes place in remote locations in the cloud and, to be able to reach them, it is necessary to use shared networks, public networks, etc. Under such circumstances**, it is usually extremely hard to make all data available for the data subjects.** On the other hand, it is worth noting that the integrity of the data might be compromised by the way in which they are shared and stored. It might happen that one of the processors or joint controllers deletes or damages the data at some point. In order to prevent such scenarios, backup copies are highly recommended. Their creation should be foreseen from the first stages of the research.

8.2 Perform a security risk analysis

According to the confidentiality principle, controllers should minimize the risks to data subjects' rights, interests, and freedoms. For this purpose, they should work on a risk-based approach (See the "Integrity and confidentiality" subsection in the "Principles"

section within Part II of these Guidelines). In all cases, controllers need to ensure that they comply with data protection requirements and are able to show how they comply e.g. through documentation (see the "Accountability" subsection in the "Principles" within Part II of these Guidelines).

To manage the risks to individuals that arise from the processing of personal data gathered from social networks, it is important that controllers develop a mature understanding and articulation of fundamental rights, risks, and how to balance these and other interests. Ultimately, it is necessary for controllers to assess the risks to individuals' rights that the use of the data poses, and determine how they need to address these and establish the impact this has on their use for research purposes. For this purpose, there are two key factors that must be considered:[47]

- Risks arising from the processing itself, such as the emergence of biases associated with profiling or automated decision-making systems.
- Risks arising from the processing in relation to the social context, and the side effects indirectly related to the object of processing that may occur.

In order to minimize such risks, controllers must ensure that appropriate technical and organizational measures are implemented to eliminate, or at least mitigate, the security risk, reducing the probability that the identified threats will materialize, or reducing their impact. It is necessary to take into account the security standards that already exist in the market, as well as the compliance standards in relation to data protection that will apply to the processing. Furthermore, developers should always remember that Article 32(4) GDPR clarifies that an important element of security is to ensure that "any natural person acting under the authority of the controller or the processor who has access to personal data does not process them except on instructions from the controller, unless he or she is required to do so by Union or Member State law" (See the "Integrity and confidentiality" subsection in the "Principles" section within Part II of these Guidelines).

The general description of the technical and organizational security measures must become a part of the processing records, where possible (Article 30(1) (g) for controllers, and 30(2) (d) for processors) and all implemented measures must form part of the DPIA, as supporting remediation measures to limit risk. Finally, once the selected measures are implemented, the remaining residual risk should be assessed and kept under control. Both the risk analysis and the DPIA are the tools that apply. The risk evaluation and the decisions taken "have to be documented in order to comply with the requirement of data protection by design" (of Article 25 of the GDPR) (see the "Data Protection by Design and by Default (**DPbDD**)" subsection in the "Main Concepts" section within Part II of these Guidelines").

Finally, the controllers should always be aware that, according to Article 32(1) (d) of the GDPR, data protection is a process. Therefore, **they should test, assess, and evaluate the effectiveness of technical and organizational measures regularly**.

[47] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.30. Available at: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (accessed 15 May 2020).

Procedures that help controllers to identify changes that would trigger a revisit of the DPIA should be created at this moment. Whenever possible, controllers should try to impose a dynamic model of monitoring the measures at stake (See the "Integrity and confidentiality" subsection in the "Principles" section within Part II of these Guidelines).

Checklist: integrity and confidentiality

 The controllers have introduced the necessary procedures to ensure that the data subject rights are adequately satisfied, no matter if the data subjects are the end-users or third parties.

 The controllers have introduced the necessary procedures to ensure that the data subject rights are satisfied in time (maximum one month after request).

 The controllers have introduced efficient tools to ensure that data subjects are able to exercise their rights in a practical manner, for instance by introducing data interoperability standards.

 Data subjects are in a position to have access to all their personal data, including the raw data that are gathered from the social networks.

 The controllers have implemented tools to locally read, edit and modify the data before they are transferred to any data controller. Furthermore, personal data processed by a device is stored in a format allowing data portability.

 The controllers have introduced tools able to communicate rectified data to each recipient to whom the personal data has been disclosed, unless this proves impossible or involves disproportionate effort.

 The controllers have introduced tools able to ensure that all data are efficiently deleted at the data subjects' request if there are no lawful reasons to oppose that request.

 The controllers have ensured that withdrawal schemes should be fine grained and should cover:

(1) any data collected by a specific means;

(2) a specific type of data collected by any means;

(3) a specific data processing.

 The controllers have documented all the information regarding these issues.

9 Data subjects' rights

Chapter III of the GDPR provides for a set of rights that the data subjects can exercise to safeguard their personal data. Although each right has specific details and issues that could affect and be affected by ICT research and development (see the "Data protection and scientific research" subsection in "Main Concepts" within Part II of these Guidelines) they all share some general features concerning their transparent information, communication and modalities of exercise (Article 12 GDPR). In this section, we analyze each specific right in the light of a processing that uses data gathered from social networks. However, since we have already analyzed the right to information (see the "Transparency" section of this part of the Guidelines), and the right not to be Subject to Automated Decision-Making has been extensively addressed in the "Human Agency" section of this part of the Guidelines, we will focus now on the remaining rights.

9.1 Right of access

Article 12(a) provides that data subjects have the right to obtain from the controller confirmation as to whether or not personal data concerning them are being processed, and, where that is the case, access to the personal data (see the "Right of access" subsection in the " Data Subject Rights" section within Part II of these Guidelines). In a nutshell, the data subject has the right to obtain from the controller information on (1) the personal data stored as well as their categories, (2) the source and the recipients of personal data to whom the data is disclosed, (3) knowledge of the logic involved in automatic processing of data concerning the data subject, and (4) the purpose of processing personal data. The whole requirement is included in article 15 GDPR. **This right is particularly important in the case of data gathered from social networks since data subjects are usually unaware of the existence of such data. Furthermore, inferred data might be created by the controller and these data might be of particular interest for the data subject.** Thus, controllers shall ensure that they have implemented adequate tools to satisfy the data subjects' requirement according to the precisions included by the GDPR.

9.2 Right to rectification

As laid down in Article 16 GDPR, data subjects hold the right to have their personal data rectified (see the "Right to Rectification" subsection in the "Data Subjects' Rights" section of Part II of these Guidelines). This is particularly relevant in the case of data gathered from social networks, since data subjects can provide false or inaccurate information due to a lack of understanding of the implications that it might have. Controllers are obliged to communicate the rectified data to each recipient to whom the personal data has been disclosed, unless this proves impossible or involves disproportionate effort. Controllers cannot argue that managing large datasets is too complex to ensure rectification in order to avoid this requirement.

9.3 Right to erasure

Data subjects have a right to ask controllers for the deletion of their personal data (see the "Right to Erasure" subsection in the "Data Subjects' Rights" section of Part II of these Guidelines). However, the use of cloud computing, the existence of diverse servers and repositories, the possibility that the data are processed by different processors and controllers, makes it hard to ensure that all backup copies and the personal data –and not only their encryption keys- are deleted. To avoid such results, controllers should monitor procedures carefully.

Finally, controllers shall keep in mind that this right does not cover processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or when it will 'adversely affect the rights and freedoms of others'. If deleting some data might cause severe damage to the rights and freedoms of others, erasure should not be allowed. Needless to say, this involves the need to balance the different interests involved.

9.4 Right to restrict the processing

According to article 18 of the GDPR, the data subject shall have the right to obtain from the controller restriction of processing where one of the circumstances described in this article applies (namely: the accuracy of their data is contested; the processing is unlawful and the data subject opposed erasure of their personal data; the controller no longer needs the personal data, but is required to store it; or the data subject otherwise objects to the processing).

Since a controller other than the social network who originally gathered the data is involved in the processing, it might be good to keep in mind that this right shall be exercised through any of the actors involved, who should inform the rest about the requirement and proceed accordingly. In this context, it can be very useful to develop data sharing agreements that help to clarify the responsibilities attributed to each of these roles in the performance of the specific data processing activities to be carried out, if the developer policies do not clarify this issue.

9.5 Right to object

Data subjects must have a possibility to revoke any prior consent given to a specific data processing and to object to the processing of data relating to them (see the "Right to Object" subsection in the "Data Subjects' Rights" section of Part II of these Guidelines). The exercise of such right must be possible without any technical or organizational constraints and the tools provided to register this withdrawal should be accessible, visible and efficient. Thus, researchers/innovators should make this option available for data subjects as soon as they start processing the data gathered from social networks.

9.6 Right to data portability

According to the GDPR, data subjects have a right to portability (see the "Right to Portability" subsection in the "Data Subjects' Rights" section of Part II of these Guidelines). In order to face this requirement, controllers should store the data in standardized formats that allows the data subjects to transmit those data, which they have provided, from one automated application, such as a social network, to another one.[48]

Anyway, it is necessary to highlight that the right to data portability only applies to data 'concerning' the data subject and data they 'provided to' the controller. As a consequence, both anonymized and inferred or derived data are not included in the right to portability, since anonymized data do not concern the data subject, and inferred or derived data have not been provided by the data subject.

Checklist: data subjects' rights

- The controllers have introduced the necessary procedures to ensure that the data subject rights are adequately satisfied, no matter if they are the end-users or third parties.
- The controllers have introduced the necessary procedures to ensure that the data subject rights are satisfied in time (maximum one month after request, extendable by two additional months with regard to the complexity of the task and the number of requests).
- The controllers have introduced efficient tools to ensure that data subjects are able to exercise their rights in a practical manner, for instance by introducing data interoperability standards.
- Data subjects are in a position to have access to all their personal data, including observed, obtained, derived and inferred data.
- The controllers have provided the data subjects with remote access to their personal data. Particularly, controllers which provide online services based on personal data have provided an online tool for this purpose.
- The controllers have introduced tools able to communicate rectified data to each recipient to whom the personal data has been disclosed, unless this proves impossible or involves disproportionate effort.
- The controllers have introduced tools able to ensure that all data are efficiently deleted at the data subjects' request if there are no lawful reasons to oppose that request.

[48] See I. GRAEF, Mandating Portability and Interoperability in Online Social Networks: Regulatory and Competition Law Issues in the European Union (22 July 2013). Telecommunications Policy 2015, Vol. 39, No. 6, p. 502–514.

- Controllers have introduced user-friendly interfaces for users who want to obtain both aggregated data and/or raw data that they still store. These tools enable data subjects to easily export their data in a structured and commonly-used format.

- The controllers have documented all the information regarding these issues.