Originally published in: Hawkins, Shane (Ed.): Access and Control in Digital Humanities. – Abingdon/New York: Routledge, 2021. Pp. 249-273. (Digital Research in the Arts and Humanities)
DOI: *https://doi.org/10.4324/9780429259616*

Paweł Kamocki

Trust is good, control is better?[1] The GDPR and control over personal data in digital humanities research

I. Introduction

The right to privacy and its younger relative—the right to data protection—are relatively new concepts. A brief overview of their history provides insight into the evolution of social norms, and the transition towards an information society. In 1890 Warren and Brandeis, in response to newspaper coverage of a high society Boston wedding, published a seminal paper defining the right to privacy as "the right to be left alone".[2] The 1948 Universal Declaration of Human Rights (UDHR) stated that "no one shall be subjected to arbitrary interference with his privacy" (Art. 12); the same right is also part of the 1950 European Convention on Human Rights, which, contrary to the UDHR, is legally binding. In 1974, a leading French journal *Le Monde* alarmingly announced that the French Ministry of Interior rented a supercomputer, Iris-80, with a whopping 3.2 billion bytes of storage space, or 32 GB—more than enough to store all the files of the French Police (Boucher 1974); the event led to the adoption of one of the first data protection laws in Europe.[3] In 1995, the Personal Data Directive was adopted with the ambition to harmonize data protection laws across the European Union.[4] Shortly after, the Charter of Fundamental Rights of the European Union listed not only the respect of private and family life (Art. 7), but also, as a separate freedom, the protection of personal data (Art. 8). The Charter was ratified in 2000 and came into full legal effect in 2009. In 2012, the proposal for what has become the General Data Protection Regulation (hereafter GDPR) was released (European Commission 2012). The Regulation was adopted in 2016 and entered into application on 25 May 2018, thus repealing the Personal Data Directive.[5] Meanwhile, in 2013, cross-referencing paparazzi photographs with data from anonymized public databases of New York taxi rides allowed researchers to reconstruct routes of some celebrities (such as actor Bradley Cooper), including street addresses as well

as information on whether or not they left a tip (Gayomali 2014). In 2015, a study revealed that 15 minutes' worth of data from brake pedal use allows researchers to identify the driver (out of 15 participants) with 87% of accuracy (Enev at al. 2016). In 2019, a paper in *Nature* described a model that allows for 99.98% of Americans to be correctly re-identified in any anonymized data set using 15 demographic attributes (Rocher et al. 2019). These examples clearly show that technological progress reduces the private sphere of individuals, which in turn may reinforce their desire for (an illusion of) privacy. One can hypothesize that the number of those who claim that they are not interested in protecting their privacy "because they have nothing to hide" is decreasing every year.[6]

In this context, the recent adoption and entry into application of the GDPR gained much attention from the general public, and even more from businesses, for which it caused a significant increase in costs (Zorz 2018). Academia in general, and Digital Humanities (hereafter DH) in particular, are also affected by this change and were forced to adjust their practices to this new legal framework.

This chapter is not intended to be a guide to GDPR compliance for DH researchers; Data Protection Officers at universities and research institutions have already developed and implemented their own strategies for achieving this. Rather, the author's ambition is to look at the use of personal data in DH projects from the data subject's perspective, and to see to what extent the GDPR kept its promise of enabling the data subject to "take control of his data".[7]

The chapter is structured as follows: Section II will briefly discuss the relation between the concept of data control and privacy and data protection law. Section III will then introduce the GDPR, and Section IV will explain its relevance for scientific research in general, and DH in particular. The main part of this article, Section V, will analyse two types of data control mechanisms (consent and data subject rights) and their impact on DH research. Finally, Section VI will provide a brief conclusion.

II. Privacy as control

As early as 1967 Alan Westin defined privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin 1967, p. 7). One year later, Charles Fried wrote: "Privacy is not simply an absence of information about us in the mind of others; rather it is the control we have over information about ourselves. To refer for instance to the privacy of a lonely man on a desert island would be to engage in irony. The person who enjoys privacy is able to grant or deny access to others" (Fried 1968, p. 482). These early philosophical works on privacy remain the foundation of today's data protection laws.

In 1983, this vision of privacy as control was embodied in a seminal ruling of the German Federal Constitutional Court (*Bundesverfassungsgericht*) concerning the population census.[8] The court ruled that "If someone cannot predict with sufficient certainty which information about himself in certain areas is known

to his social milieu and cannot estimate sufficiently the knowledge of parties to whom communication may possibly be made, he is crucially inhibited in his freedom to plan or to decide freely and without being subject to any pressure influence".[9] The ruling gave rise to the doctrine of informational self-determination, which shaped German and European data protection law.

For Paul M. Schwartz (2000), this privacy-as-control paradigm is characteristic of the liberal approach to privacy, which the author opposes to communitarism, in which individual privacy is seen as a threat to the common good.[10] Schwartz himself takes a median stance; data protection laws, as discussed below, are also striving to strike a good balance between the two interests.

One can easily imagine that giving individuals full control of their personal data is not always a workable solution, to say the least, and that the privacy-as-control paradigm has its limits even for the most liberal of us (Allen 2000). In fact, individuals often want more privacy than is socially desirable and acceptable—for example, many would likely choose not to disclose their income to the tax authorities, or their health data to an insurance company. On the other hand, some individuals may also opt for less privacy than is socially acceptable, for example, by choosing not to wear clothes in public. While both these extremes are to be avoided, individuals should still be given freedom to choose whether they want to disclose information about their health to their local baker, and to choose (within socially acceptable limits) how much of their body they want to cover with clothes. Additionally, in practice individuals may not be able to exercise meaningful control over all of their data due to, for example, information overload, information asymmetry, or lack of sufficient information literacy (Van Ooijen & Vrabec 2019). Probably there are few individuals who read all of the privacy policies that they accept, or adjust the privacy settings of all their applications and devices. For these reasons, full individual control is rarely the best solution, and legislators counterbalance control mechanisms with various safeguards and exceptions.

For a lawyer, the ultimate form of control is ownership, which is understood as the absolute right to use and dispose of property.[11] Ownership of intangible goods is the domain of Intellectual Property (IP), which is dynamically expanding to catch up with the development of new technologies.[12] In Europe, various forms of IP law grant ownership not only for such intangible assets as original works (copyright), inventions (patents), and trademarks or designs, but also, for example, for databases (*sui generis* right), and a new IP right in data is currently under discussion.[13] In this context, it may be tempting to imagine privacy as a form of IP, and indeed this was and still is discussed by some authors.[14] As of today, the form of control that individuals have over their personal information is in fact quite far removed from ownership.

Researchers, on the other hand, may and often do have ownership over the material that they gather and compile. Depending on the circumstances and the jurisdiction, this can be copyright, the *sui generis* database right, or another similar IP right. Apart from this ownership (which for researchers is often

quite secondary due to the—usually—low commercial value of research data), researchers have legitimate interest in exercising control over their research data. If these data include personal data, which is often the case (see below), a conflict arises between the researchers' and the data subjects' interests. The following sections analyse how this tension is resolved in the GDPR.

III. The GDPR and why it's bigger than you think

With its 173 recitals and 99 articles on 88 pages, the GDPR may not be the longest, but is probably the most ambitious piece of EU legislation.[15] As a regulation, the GDPR applies directly and (at least theoretically) uniformly in all the EU Member States, unlike directives, which are "binding as to the result to be achieved", but leave to the Member States "the choice of form and methods".[16] Under the GDPR's predecessor, the Personal Data Directive 1995, some significant differences regarding data protection arose between various EU Member States, and the GDPR's ambition is to put an end to this. However, the importance of the GDPR exceeds the European Union, and for a number of reasons.

Firstly, the GDPR applies not only in the 27 Member States of the European Union, but also, as part of the Agreement on the European Economic Area (EEA), in the whole EEA, which also includes Iceland, Lichtenstein, and Norway, totalling over 500 million people.

Secondly, the territorial scope of the GDPR is even larger than the EEA, and considerably larger than the territorial scope of the Personal Data Directive. It is defined (in Article 3 of the GDPR) on the basis of two alternative criteria: "establishment" and "targeting" (EDPB 2019). As a result, the GDPR applies to the processing of personal data by non-EU entities if they have an establishment (e.g. an office) in the EU (and the processing is linked to the activities of this establishment), or if they merely offer goods and services (also without payment) to individuals in the EU, or monitor the behaviour of individuals on the EU territory. In both scenarios ("establishment" and "targeting"), the GDPR applies even if the actual processing is carried out outside the European Union. Exceptionally, non-EU research organisations can meet these criteria, particularly the "targeting" one; it seems that according to the European Data Protection Board, this could be the case of surveys or behavioural research involving subjects from the EU (*idem*, p. 20). It is therefore not excluded that the GDPR applies to DH research, even in non-EU institutions.

Thirdly, due in part to the two previously mentioned factors, the GDPR has set global standards for data protection laws. The powerful "Brussels effect" (Bradford 2020) pushed some international companies to apply GDPR principles in their activities worldwide, and some legislators to adopt similar laws.[17] This does not come as a surprise, given that ensuring an appropriate level of data protection is in principle necessary for a non-EU institution or a non-EU business to be able to lawfully receive personal data from the EU, which in the globalised economy is almost synonymous with any form of international expansion or

cooperation.[18] Therefore, although this chapter only discusses the control of personal data under the GDPR, the author believes that non-EU readers will also find it interesting and useful.

IV. The importance of the GDPR for DH research

A. Scientific research in the GDPR—a preliminary remark

The GDPR provides for a number of derogations for processing carried out for scientific research purposes. These derogations, however, are only available where the processing is accompanied with "appropriate safeguards . . . for the rights and freedoms of the data subject" (Art. 89(1)). These safeguards are technical and organizational measures that ensure in particular the respect of data minimization, which is one of the main principles of GDPR, according to which data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed" (Art. 5(1), (c)). The GDPR only gives one example of such safeguards (Art. 89(1)): pseudonymization, that is, processing of personal data in such a manner that they can no longer be attributed to a specific person without the use of additional information provided that such additional information is kept separately (Art. 4(5)). The most basic application of pseudonymization consists of replacing names of data subjects with pseudonyms (e.g. "participant 1"), and then keeping a document that lists the names and the pseudonyms they were replaced with separately in a secure environment (e.g. making it available only to the project's primary investigator).

Some national legislators give examples of other appropriate safeguards, albeit in rather general terms.[19] The choice of safeguards should be adapted to the circumstances, but may include such measures as data encryption, access restrictions, or training sessions designed to increase awareness of concerned personnel. According to one commentator, the requirement of appropriate safeguards "instantiates into law what is already good scientific research practice" (Dove 2018, p. 1016). However, specifically in the case of DH, it seems that the GDPR requires researchers to take an extra step and carry out a documented analysis of existing privacy risks and appropriate ways to mitigate them for each project on a case-by-case basis, rather than relying on institutional solutions. Indeed, DH researchers work with a multitude of data types, ranging from interviews and surveys to newspaper articles, to user-generated content, and privacy risks associated with each of these data types vary from very high to almost non-existent.

B. Data, data everywhere, but . . . it's all personal?

The territorial scope of the GDPR was briefly presented in the previous section. It is now time to present its material scope, which prima facie is rather straightforward: the GDPR applies to the *processing* of *personal data*. Processing is easily defined: it is any operation performed on data, whether or not by automated

means, including collection, alteration, disclosure, combination, but also mere consultation, retrieval, storage or even deletion (Art. 4, (2)). Some types of personal data processing are expressly excluded from the scope of the GDPR: processing of personal data carried out by competent authorities for prevention, investigation, prosecution of criminal offences and execution of criminal penalties (which is governed by a directive called Police Justice), and processing by individuals in the course of purely personal or household activities with no connection to professional or commercial activities (e.g. private correspondence or social networking) (Rec. 18). In particular, the GDPR does apply to the processing of personal data for scientific research purposes (albeit with some alleviations explained below).

When it comes to personal data, the definition is more complex. According to Article 4(1) of the GDPR, personal data is "any information relating to an identified or identifiable natural person ('data subject')". This definition is not new, as it existed already in the Personal Data Directive 1995. In fact, the core of this concept was already present in German law in the 1970s, but its breadth keeps increasing with technological progress as more and more data can be traced back to an individual.[20]

The Article 29 Data Protection Working Party (hereafter WP29), a former advisory body made up of representatives of data protection authorities of all the EU Member States, provided an analysis of this definition in its classic Opinion 4/2007 on the concept of personal data (WP29 2007). The Opinion adopts a very broad approach to personal data, in line with the purpose of data protection laws, namely to protect individuals. Although it was published under the Personal Data Directive 1995, and WP29 has since been replaced by another body (the European Data Protection Board), the Opinion remains relevant and is still relied upon under the GDPR.

WP29 identified four elements of the definition: (1) any information, (2) related to, (3) identified or identifiable, and (4) natural person. Here, rather than following this linear order of presentation, the four elements will be discussed from the simplest to the most complex. Firstly, *any information* can be "personal data", regardless of its *nature* (i.e. whether the information is a fact or an opinion) and its *format* (digital or analogue, but also textual, graphic, audiovisual, etc.), but also of its *content*. In its opinion, WP29 uses the word "content" in quite a specific sense to signify an aspect of the definition that should not be overlooked: data can qualify as personal regardless of whether they concern the private or the public sphere of the life of an individual (WP29 2007, p. 6). Therefore, the processing of information about an individual's academic affiliation or the subject of his PhD thesis should abide by the same rules as the processing of his marital status or place of birth. As mentioned above, in the Charter of Fundamental Rights of the European Union data protection (Art. 8) is a distinct right from the right to private and family life (Art. 7), and indeed data protection does not necessarily have to do with one's privacy (and vice versa, one's privacy can be invaded without any processing of his or her personal

data). In this sense, referring to the GDPR as a "privacy law" or "data privacy legislation" does not do it justice.

That said, certain types of information are recognized as more sensitive and governed by stricter rules. These special categories of personal data are: information about one's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data (only when used with the purpose of uniquely identifying the person), as well as information about health, sex life, and sexual orientation (Art. 9(1)). The list may seem surprising to some; indeed, the information about someone being a rationalist (which arguably is a philosophical belief) is recognized as more sensitive than his income or his credit card number. In practice, however, the degree of control that one can exercise over his or her sensitive data is not substantially greater than for "ordinary" personal data. The practical impact of this distinction lies more on the side of the administrative burden placed on those who process sensitive data (which is manifested most notably in a limited catalogue of legal bases for processing), and as such is beyond the scope of this chapter (Art. 9(2)).

Secondly, personal data is data that relates to a *natural person*, that is, a living individual. Data related to legal entities (companies) or to the deceased are not directly covered by this definition. However, given WP29's broad interpretation of the phrase "related to" (see below), such data can still relate to living individuals (WP29 2007, pp. 21–24). For example, the information about the financial situation of a company may have an impact on rights and interests of its employees, who can be denied a loan or offered new professional opportunities on the basis of such information. This is even more obvious when it comes to information about a dead person (e.g. concerning her health or even political involvement), which may have an impact on her descendants. Moreover, the GDPR (unlike the Personal Data Directive) allows Member States to provide special rules regarding the processing of personal data of deceased persons (Rec. 27). Most notably, such rules have been adopted in France. Under French law, data subjects can give data controllers (e.g. their employers, or Facebook) specific or general instructions regarding retention, erasure, and communication of their personal data after their death.[21] In addition to that, French law allows the heirs to exercise certain rights of deceased data subjects (see below) for a very limited list of purposes. Given the amount of (potentially permanently stored) digital data that the current generation will leave behind, the fate of data of deceased persons should be seriously considered and efficiently regulated.

Thirdly, the person that data relate to needs to be *identified or identifiable*. A person is identified if she is singled out from a group. This singling out can be direct (typically by a name or name/surname combination, but also, e.g. with a social security number) or indirect (by a unique characteristic, e.g. the Prime Minister of Canada, or a unique combination of characteristics).

The notion of identifiability is particularly delicate and crucial for the definition of personal data. In determining whether a person is identifiable (possible to identify), account should be taken of "all the means reasonably likely

to be used" by the controller or by another person (Rec. 26). In assessing whether a mean of identification is likely to be used, the factors to be considered are the cost, the intended purpose, the way the processing is structured, the advantage expected, the interests at stake for the individuals, as well as the risk of organisational dysfunctions and technical failures.[22] It clearly appears that identifiability can change over time: information that is not identifiable at the moment of collection may become so, especially with exponential growth of publicly available online data which can be quickly and easily consulted and cross-referenced.

The Court of Justice of the European Union (CJEU) analysed the concept of identifiability in its 2016 ruling in the *Breyer* case.[23] The main question in the case was whether a dynamic IP address constitutes personal data. Since a dynamic IP changes with every connection, in itself it does not allow the website provider to identify the device (and therefore the person) behind the connection. However, if cross-referenced with information in possession of the Internet access provider, a dynamic IP address does allow one to identify the person. Under German law (the case was referred to the CJEU by a German court), the website provider does not normally have access to the information necessary to identify a dynamic IP, but he can, in the event of a cyber-attack, obtain the necessary information from the access provider (*via* a competent authority). Therefore, according to the Court, the website provider has the means reasonably likely to be used to identify the user on the basis of a dynamic IP. In ruling so, the Court confirmed WP29's position that the elements necessary to identify the person do not need to be in possession of one person; it is enough if there is a possibility for the necessary elements to be legally cross-referenced. In this approach, a username or an e-mail address, for example, even seemingly anonymous, will constitute personal data since they can be used by competent authorities to identify the person that uses them.

Fourthly, in order to qualify as personal data, the information should *relate to* a person. Once again, WP29 adopts a very broad interpretation of this aspect of the definition of personal data. According to the advisory body, information relates to a person not only if it says something about the person (relation via *content*) but also if it can be used to evaluate, influence or treat the person in a certain way (relation via *purpose*), or if it can have an impact (even minor) on the person's rights and interests (relation via *result*) (WP29 2007, pp. 9–12). Understood in this way, and taking into account technological progress, the notion of personal data may soon cover most if not all information, probably to the point of absurdity. One author argued provocatively that even information about weather can qualify as personal data, especially in a *smart city* (Purtova 2018). Indeed, in such a city everyone can be identified (using data from many Wi-Fi sensors), and information about weather conditions, even if it does not say anything about anyone, can still be used to predict (the *result* element) and to influence (the *purpose* element) their behaviour (e.g. via tailored messages sent and displayed to the person).

Between 2014 and 2017 it seemed that the CJEU would deviate from WP29's opinion and apply a somewhat narrower concept of personal data. The 2014 ruling in *YS and Others* concerned a "minute": an administrative document used in the Dutch immigration procedure (application for a residence permit), which contained information about the applicant and a "legal analysis", that is, an assessment of the applicant's situation in the light of the applicable legal provisions.[24] In the facts of the case, a group of applicants requested access to their data contained in "minutes"; however, the Dutch authorities only allowed them access to their "personal information", and not to the "legal analysis", claiming that this part does not in fact constitute personal data. The CJEU upheld this argument, despite the fact that the "legal analysis" could clearly be used to evaluate the applicants (so it related to them via *purpose* and possibly also via *result*). The decision, controversial as it was, was primarily motivated by the fact that granting applicants access to the "legal analysis" would not be compatible with the purpose of data protection, which is to protect privacy *inter alia* by granting the data subject the right of access, enabling him to check if the data about him is correct and processed lawfully.[25] In other words, the applicant for a residence permit is not in a position to assess whether the "legal analysis" concerning him is correct, and if its processing is lawful. One could argue that the arising problem was analysed at a wrong end—through the (re-)interpretation of the concept of personal data, rather than through the teleological interpretation of the right of access.

However, in its 2017 ruling in the *Nowak* case the Court returned to the broad interpretation of personal data.[26] This time, the facts concerned the candidate's right of access to an examination script containing both his answers and the examiner's comments. *A priori*, the comments are similar to the "legal analysis" in *YS and others*, and the solution should not differ. However, the Court ruled that the comments did constitute the candidate's personal data, precisely because, just like the answers themselves, their purpose is to evaluate the candidate (the *purpose* element) and they have an impact on the candidate's interests (the *result* element). Moreover, according to the Court, both the answers and the comments say something about the candidate, for example, about his level of knowledge (the *content* element). It is worth noting that in this approach, the comments constitute both the candidate's and the examiner's personal data, but for the Court it does not prevent the candidate from exercising his access right. In addition to this, the Court also found that the candidate's access to the script is in fact compatible with the purpose of data protection legislation, which is to safeguard the candidate's legitimate interest in the protection of his private life. *YS and others*, with its relatively narrow analysis, was therefore overruled.

The purpose of this chapter is not to debate about the justifiability of the very broad scope of personal data, nor to propose any changes in the definition. Instead, the author feels obliged to inform the readers that a very large proportion of data used in DH research is susceptible to qualifying as personal data, and therefore triggering a series of obligations and responsibilities on behalf of

their institutions, and a spectrum of rights on behalf of the data subjects, even if the data is not disclosed but simply consulted. This is the case of all sorts of interviews, audio- and video-recordings featuring research subjects, their writing samples, as well as meeting minutes, e-mail correspondence, the above-mentioned examination scripts, and so on.

Data anonymization, where it is possible, is a way out of this conundrum. Indeed, the GDPR does not apply to the processing of anonymized data (i.e. personal data that have been processed in such a manner that the data subject can no longer be identified by any means reasonably likely to be used). The standard of anonymization, however, is very high; most importantly, anonymization should be irreversible. In its 2014 opinion, WP29 have evaluated some common anonymization techniques, such as randomization, noise addition, k-anonymity and t-closeness, only to conclude that none of them produces fully satisfying results (WP29 2014a). Moreover, attempts at anonymization, especially concerning language data or any sort of audio or audio-visual material, often strip them of utility for research (which, truth be told, badly tolerates any irreversible alteration of the source material). Therefore, apart from being very demanding and costly (many operations have to be performed manually), anonymization of research data and the information loss that it entails are often too high a price to pay for GDPR compliance.

V. The data subject's control of personal data

This part of the chapter presents the various instruments that allow data subjects to exercise control over their personal data. In particular, this section will investigate to what extent these control mechanisms apply when the data are processed for research purposes.

A. The power (and the economics) of consent

The most fundamental prerogative of the data subject is to give—or to refuse—consent. In EU data protection law, consent embodies the theory of informational self-determination (see above). The GDPR employs this mechanism of control several times. Most importantly, consent is one of the available legal bases for data processing, both when it comes to ordinary personal data (Art. 6(1)(a)) and to sensitive data (Art. 9(2)(a), explicit consent). In addition to that, it can also exceptionally be used to legitimize data transfers outside of the EEA (Art. 49(1)(a)), as well as to allow the controller to make certain decisions concerning the data subject solely by automated means (which is normally prohibited) (Art. 22(2)(c), explicit consent).

Article 4(11) of the GDPR defines consent as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". Many guidelines on consent

are available, including very recent ones from the European Data Protection Board (hereafter EDPB; on consent see EDPB 2020). The obligation for consent to be *freely given* is meant to guarantee that the data subject can exercise real choice, in the spirit of informational self-determination. Valid consent cannot be given under duress or as a result of deception or intimidation. Ideally (in the somewhat idyllic vision of the EDPB), there should be no significant disadvantage for the data subject to refuse or withdraw consent (*idem*, p. 13). Therefore, consent should not bundled with other agreements (e.g. Terms of Service) into a 'take it or leave it' whole, but requested separately (*idem*, p. 10).

Consent also needs to be *specific*, that is, given for a specific purpose. According to the purpose limitation principle, personal data should only be collected for "specified, explicit and legitimate purposes", and not processed further for purposes which are incompatible with the initial purpose (Art. 5(1)(b)). If the same data are processed for several purposes (e.g. to be used in a research project, but also to create a mailing list to disseminate information about conferences, projects and publications) with consent as the legal basis, then two separate consents should be requested, which can be separately refused or withdrawn (this feature of consent is referred to as "granularity"; EDPB 2020, p. 12). Paradoxically, specificity of consent reduces the control that a data subject can exercise over his personal data—it is impossible to validly give blanket consent to one controller (e.g. university or a foundation) for any processing operation he may want to undertake now or in the future, regardless of its purpose. This derogation from the spirit of informational self-determination is clearly motivated by the intention to protect the data subject as the weak party.

Another element that aims at protecting the data subject rather than at enabling him to control the data is the requirement for consent to be *informed*. In order to validly consent, the data subject has to be provided at least with information about the identity of the controller, the purpose of the processing operation for which consent is sought, the categories of data that will be collected and used, and the existence of the right to withdraw consent (see below) (EDPB 2020, pp. 15–16). Without this information the data subject cannot—even if he genuinely wanted to—validly consent to the processing, which in fact weakens the degree of control that the data subject has over his data. Full control, one could theorize, would imply the freedom to make any decisions, including uninformed ones. It is worth noting that this requirement is distinct from the general obligation of transparency, whose scope is much broader (Art. 5(1)(a) and 12–14). It is therefore possible to obtain valid consent and still violate the transparency principle (and vice versa).

Consent does not have to be given in a written statement, it can also be an oral statement or any affirmative behaviour (like an oral confirmation, but also nodding, waving or clicking), as long as it is unambiguous (EDPB 2020, p. 18). However, it should be recorded, so that the controller can demonstrate that the data subject has indeed consented (Art. 7(1)). The freedom of form is an element empowering the data subject, rather than protecting him, as it would be the case

if writing were mandatory. Silence, understood broadly as absence of affirmative action, cannot amount to valid consent (EDPB 2020, p. 18). It is therefore impossible for the data controller to adopt a consent-by-default reasoning (e.g. online forms with pre-ticked boxes), even if the data subject is informed about it (e.g. "If you stay in this room, I understand that you consent to the use of your data").

Probably the most important characteristic of consent from the point of view of this chapter is that it can be withdrawn at any time, at no cost for the data subject (Art. 7(3)). It should be as easy for the data subject to withdraw consent as it was to give it in the first place. So, if consent was signified by ticking a box in an online form, withdrawal should also be signified with a single click (*ibid*.). Withdrawal is not retroactive, in a sense that it does not affect the lawfulness of prior processing, but after withdrawal, the controller shall delete the data if no other legal basis applies (EDPB 2020, p. 24). It should also be stressed here that it is not possible for the controller to silently switch to another legal basis (such as legitimate interest) without informing the data subject about it (*idem*, p. 25).

In the author's opinion, consent as it is shaped by the GDPR does empower the data subject to control his data, but only to a certain extent and quite far from the ideal of informational self-determination. In fact, it is a sort of 'controlled control'. The GDPR aims at guaranteeing the freedom of the data subject to refuse consent or to withdraw it at any time, but it only enables him to validly authorize the processing if his acceptance meets the conditions of specificity and informedness. One should not overlook, however, that consent is not always mandatory: it is only one of six legal bases listed in Article 6 of the GDPR—and they are all equally good. Nothing in the GDPR obliges the controller to seek consent first and opt for other grounds only if consent is not suitable or obtainable.

In the context of research there are, it seems, three possible legal bases for processing, and consent is rarely the most suitable one. The two conceivable alternatives are "legitimate interests" (Art. 6(1)(f)) and "public interest" (Art. 6(1)(e)). However, if public interest is to be a legal basis for data processing, then it should be laid down by national or EU law, which calls for a very narrow interpretation (Art. 6(3)). Moreover, some may argue that DH research is not in the public interest, as it does not directly contribute to the welfare and well-being of the general public the way that, for example, biomedical research does. Even regarding biomedical research, it would appear that, according to a preliminary opinion of the European Data Protection Supervisor, no laws that would clearly declare public interest in such research have yet been adopted (EDPS 2020, p. 23). Taking both these arguments into account, the author of this chapter is of the opinion that "public interest" in the current legal framework is not the best ground for processing personal data for DH research purposes.

On the other hand, legitimate interest in carrying out research can often be a very appropriate alternative to consent. However, this basis is not available for "public authorities in the performance of their tasks" (Art. 6(1), *in fine*). This led

some commentators to claim that public entities such as universities cannot base their processing activities on legitimate interest (European Parliament, Scientific Foresight Unit 2019, p. 22). The author of this chapter does not agree with this interpretation; in his opinion, universities and other public research institutions are public bodies, but not public authorities, and therefore nothing precludes them from relying on legitimate interest. Therefore, in the author's opinion, this basis is often the most suitable one as far as data processing for academic research purposes is concerned.

The "legitimate interest" ground is subject to a balancing test, that is, the controller's legitimate interest must not be overridden by the interests or fundamental rights and freedoms of the data subject, in particular taking into account the reasonable expectations of the latter.[27] Arguably, most DH research would meet the condition. However, in the author's opinion a more careful assessment is needed on a case-by-case basis if personal data were to be published. The outcome of the test would then really depend on the nature of the data and its direct relevance for the research community. Hypothetically, one could distinguish between publishing an interview in which the interviewee, when asked to share his memories of a famous writer, speaks about details of his private life (e.g. "I first met Philip Roth when I stayed at my aunt's place, her name was Linda Goldberg and she was one of the richest persons in New Jersey, and I immediately fell madly in love with him"), and an interview in which a speaker of an endangered language recounts the traditions of his people. The latter, unlike the former, would probably pass the balancing test.

Another factor that reduces the importance of consent in the context of research is the purpose extension mechanism.[28] As mentioned above, according to the purpose limitation principle, personal data should only be collected for specific, explicit, and legitimate purposes. However, the data can be further processed for purposes compatible with the initial purpose for which they were collected. By extension expressly provided for by the GDPR, scientific research (subject to "appropriate safeguards"—see above) is always regarded as a compatible purpose (Art. 5(1)(b)). This means that, in practice, personal data which had been lawfully collected (on the basis of consent or other ground) for any purpose (e.g. student essays collected for teaching purposes) can subsequently be re-used by the same controller for research purposes.

Finally, even if personal data are processed for research purposes on the basis of consent, another exception seems to exempt such consent (at least partially) from the specificity requirement. As per Recital 33 of the GDPR, "data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research". Even though the European Data Protection Board warns that the Recital "does not disapply the obligations with regard to the requirement of specific consent", and that "scientific research projects can only include personal data on the basis of consent if they have a well-described purpose" (EDPB 2020, p. 30), the Recital

still provides a significant relief for data-intensive research, where the exact purpose of each processing operation cannot always be specified.

In conclusion, although aimed at enabling the data subject to control the use of his personal data, the mechanism is effectively counterbalanced by numerous elements of the data protection framework and especially by the availability of alternative legal bases for processing or purpose extension. The impact of consent for DH research projects is particularly reduced. It seems that the data subject can only in very limited cases efficiently control the use of his or her data for research purposes through consent.

B. Is it all right to control?

In data protection law, consent is not the only mechanism introduced to empower data subjects, who are also granted a seemingly large variety of rights, many of them designed to compensate for shortcomings of consent. Chapter 3 of the GDPR provides for the rights of access (Art. 15), rectification (Art. 16), erasure (Art. 17), restriction (Art. 18), as well as the right to data portability (Art. 20) and the right to object (Art. 21).

Many details about these rights (perhaps apart from portability) are still unsettled. The European Data Protection Board is currently working on relevant guidelines which hopefully will alleviate some doubts.[29] In order to avoid any possible contradiction with the future guidelines, the following analysis will focus on essential elements of each right, at the risk of being somewhat superficial.

Access

The right of access to one's personal data is the mother of all rights of data subjects in the GDPR. It enables the data subject to obtain from the controller a confirmation that his data are being processed, access to the data, and a free copy thereof. Moreover, the data subject may request information about the purposes of processing, the categories of personal data concerned, the persons or entities (or categories thereof) to whom the data have been disclosed, the data retention period (or at least the criteria used to determine it), the source from which the data were obtained (if they were not obtained directly from the data subject), as well as about certain rights of data subjects (including the right to lodge a complaint with a supervisory authority). In principle, the right can be freely exercised regardless of the legal basis of processing (consent, legitimate interest or other). However, if requests from a data subject are manifestly excessive, in particular because of their repetitive character, the controller may either charge a reasonable fee or refuse to act on the request (Art. 12(5)).

The purpose of the right of access is to enable the data subject to know which information about him is in possession of the controller, and to check if the processing of his data complies with the requirements of the GDPR. If this is not

the case, the data subject can take appropriate further measures, ranging from rectification to erasure of the data.

Although conceptually linked, the rights presented below may also be exercised independently, that is, without a prior access request.

Rectification

If the data turn out to be inaccurate, the data subject can request their rectification from the controller. If the data are incomplete in view of the purpose of the processing, the data subject can have them completed, including by providing a supplementary statement (Art. 16). It seems that the right applies only to objective and factual, and therefore verifiable data (hard data), such as names of people and places, and dates and scores.[30] Therefore, a data subject cannot request rectification of his interview data, claiming that it does not accurately reflect his viewpoint, while in fact it is a faithful reproduction of his own words. Moreover, the controller is not obliged to blindly accept any requests for rectification; he may in fact contest the request, although it seems that in the light of the general principle of accountability (Art. 5(2)), it is always up to the controller to verify accuracy and completeness of the data, and not to the data subject to prove the contrary. The data subject may in turn request restriction of processing of the disputed data for the time necessary for the controller to make the verification (Art. 18). Restricted data can still be stored by the controller, but not further processed without the data subject's consent (see below).

Erasure (right to be forgotten)

At first glance, the most powerful right allowing the data subject to control the use of his personal data is the right of erasure, also referred to (somewhat pompously) as "the right to be forgotten". In short, this right enables the data subject to request from the controller that his personal data be deleted without undue delay. Moreover, if the data had been made public, the controller should take "reasonable steps" to inform anyone else who processes the data of the data subject's request for erasure (Art. 17(2)). Famously, this right can be used to request de-referencing from search engines such as Google Search.[31]

As a matter of fact, however, the right of erasure is rather limited. It can only be exercised in a restricted number of situations where the processing is actually illegal, that is, where it violates the GDPR. This is the case for example where the data subject has withdrawn his consent and no other legal ground for processing is available (i.e. the processing violates the principle of lawfulness), or where the data are no longer necessary to achieve the purpose of processing (i.e. the processing violates the principle of purpose limitation). The right of erasure, contrary to what some may believe, does not enable the data subject to arbitrarily request deletion of his personal data; rather, it provides the data subject with an efficient tool to control the compliance of the processing of his data

with the GDPR. It is particularly useful (albeit quite redundant in theory) as a complement for withdrawal of consent, especially for online uses (e.g. the data subject had consented to the use of his data in an online service, then withdraws consent—the right of erasure could then be exercised to obtain de-referencing from Google). Where personal data are processed in a way that complies with the GDPR, the right of erasure is quite powerless.

Restriction of processing (blocking)

The right to restriction of processing can be described as the right of erasure's little cousin. If the processing is unlawful, the data subject may choose to request blocking (restriction of processing) instead of erasure. Blocked data are not erased, they continue to be stored, but cannot be further processed (e.g. disclosed or deleted) without consent of the data subject. This is an interesting alternative in cases where the data subject wants to preserve the data, for example, as proof to support his legal claim. Blocking may also be requested if there is a dispute between the data subject and the controller regarding accuracy of the data (see above about the right of rectification).

Right to object

The right to object is in a way a substitute to withdrawal of consent in cases where processing is not based on consent, but instead on legitimate interest or on public interest. It enables the data subject to object to the processing "on grounds relating to his or her particular situation", which seems to mean for any reason including no reason at all (Art. 21(1)). The controller shall no longer process the data, unless he can demonstrate "compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject" (*ibid*.).

From the pragmatic point of view, by exercising his right to object the data subject forces the controller to run a second balancing test. The first balancing test was required (as per Art. 6(1)(f)) to assess whether the legitimate interests of the controller could be a legal basis for the processing, that is, if the interests of the controller are not overridden by the interest, rights, and freedoms of the data subject (see above). Upon receiving an objection from the data subject, the controller has to either demonstrate that his interest in the processing is not only legitimate, but also compelling, or if he fails to do so: delete the data. For now, little guidance is available about the standard for this second balancing test (for "compelling legitimate grounds"), but the question is likely to be discussed in the upcoming guidelines of the European Data Protection Board.

The only hypothesis where the right to object is absolute, that is, the controller has no possibility to respond and is obliged to delete the data, is when the data are processed for direct marketing purposes, such as sending targeted ads or e-mails, or text messages with commercial offers. Then,

the data subject may indeed arbitrarily cause the processing to stop (Art. 21(2)). In such cases it seems that the data subject can effectively control the processing of his data.

Right to portability

Right of portability was specifically designed to empower data subjects and affirm their control over their personal data. It enables a data subject to receive the personal data that he had provided to a controller (in a structured, commonly used and machine-readable format) and to transmit those data to another controller. Both data knowingly provided by the data subject (e.g. through online forms) and those gathered through observation of his behaviour (e.g. browsing history) are concerned (WP29 2017a, p. 10). In theory, this allows individuals to seamlessly switch between various IT service providers, for example, by requesting information about their viewing history and preferences to be transferred from one video streaming service to another. At first glance, the right of portability comes close to control, and even ownership of data, as it enables the data subject to take his data and go elsewhere. However, this is not really the case.[32] Firstly, the scope of the right of portability is significantly limited: the right only applies to cases where the processing is carried out by automated means and based on consent or on a contract. Secondly, a portability request is not automatically accompanied with a request for erasure, which means that in principle the data subject can only take a copy of his data, but the controller can still process them in a way compatible with the GDPR. Since, as discussed above, the scope of the right of erasure is quite restricted, it is not always possible for the data subject to effectively exercise this right every time he files a portability request. Even where a portability request is followed by withdrawal of the data subject's consent, the controller still may continue the processing if an alternative legal basis, such as legitimate interest, is available. Therefore, in practice, the right of portability only enables the data subjects—in some circumstances—to take a copy of their data, but usually without taking the data away from the controller. Coupled with practical difficulties related to data interoperability (despite the obligation to provide the data in a "commonly used" format, the data ported from one controller may be of little use for another controller) and distinguishing between the data "provided by the data subject" and "created by the controller" (e.g. despite the portability of the underlying data, a user profile generated by the controller on the basis of his browsing history is in and of itself not portable; WP29 2017a, p. 10), the right of portability only provides the data subjects with very limited control over their data at best.

This brief overview demonstrates that, for the most part, the rights of data subjects do not enable data subjects to exercise full control over their data. The following section will explore how these rights can be exercised in the context of DH research projects.

C. Rights of data subjects in DH research projects under the GDPR

The first right of data subjects affected by a research exception is information (Art. 12–14). Where data are not obtained directly from the data subject, but instead, for example, downloaded from social media, the controller is not obliged to provide the data subject with information about the processing if it is impossible, requires disproportionate effort, or is likely to render impossible or seriously impair the achievement of the research purposes (Art. 14(5)(b)). It should be noted that the exception does not apply where the data are collected directly from the data subject, for example, where the subject is directly interviewed by a researcher.[33] Moreover, it is not clear how data protection authorities will interpret the standard for serious impairment of the intended purposes, but it is not unlikely that their interpretation will be stricter than researchers would wish. Indeed, WP29 Guidelines on transparency suggest that the exception applies only if the controller is able to demonstrate that "the provision of the [required] information . . . alone would nullify the objectives of the processing" (WP29 2018b, p. 31). As an example of practical interpretation of this exception, the Guidelines quote a bank investigation into a money laundering scheme, which indeed seems quite distant from the reality of DH research (*idem*, pp. 31–32).

Similarly, in order to determine whether the provision of information would require disproportionate effort, a high standard should apply. According to Recital 62 of the GDPR, the factors to be taken into account include the number of data subjects (the higher the number, the greater the effort), the age of the data (the older the data, the greater the effort) and appropriate safeguards adopted (the more robust the safeguards, the smaller the required effort to qualify for the exception). Specifically, in the context of data obtained from social media it should be noted that social media services usually provide for an easy way to contact the author of each post, so impossibility or even disproportionate effort may be hard to demonstrate.

According to the authors of a recent study, the exception to the right of information has a chilling effect on all the remaining rights of data subjects (European Parliament, Scientific Foresight Unit (STOA) 2019, pp. 23–25). Indeed, if the data subject has no knowledge of the processing, he is extremely unlikely to exercise his rights of access, erasure, objection, or portability. However, it should not be forgotten that when the exception applies, the GDPR still requires the controller to "take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available" (Art. 14(5)(b)). Arguably, the safeguards that should accompany any research on personal data (see above) are enough to meet this requirement, and making the information publicly available is not always necessary.

Rights of access and rectification in the research context are not limited by the GDPR; however, as mentioned above, they are extremely unlikely to be exercised in cases where the data subject is not informed about the processing.

The right of erasure, like the right of information, is also limited when its exercise "is likely to render impossible or seriously impair the achievement of the objectives of that processing" (Art. 17(3)(d)). The right to object is only slightly modified in the research context: it cannot be exercised if the processing is "necessary for the performance of a task carried out for reasons of public interest" (Art. 21(6)). Arguably, this threshold is slightly lower than in the general framework, where the right to object has no effect if the controller "demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject" (Art. 21(1)), although the details of both these standards remain highly unclear. The forthcoming guidelines from the EDPB will hopefully shed more light on the question.

Finally, the right to portability can theoretically still be exercised in the research context. However, this right only applies where the processing is based on the data subject's consent, which is rarely an optimal solution for research, or on a contract, and not on legitimate interest, which seems to be the most suitable basis for data processing in a research context. Moreover, even where the processing is based on consent, in the context of DH research it is difficult to imagine the motivations of a data subject to have his data transmitted from one research institution to another. Probably the only possible scenario could involve moving data from one institutional archive to another. However, as explained above, the data subject would be unlikely to succeed in having his data removed from the first repository. Withdrawal of consent would not be enough, as an alternative ground (legitimate or public interest) would likely be available to resume the processing, and a request for erasure would likely be unsuccessful if it could seriously impair the achievement of research objectives.

D. Further limitations possible under national laws of the EU Member-States

As explained above, the GDPR, as a regulation, applies directly in all EU Member States and, unlike a directive, does not require transposition. It is intended to unify (and not, like a directive, merely harmonize) the data protection law across the EU. However, in some areas the European legislator decided to leave some leeway to the Member States. To an extent, research is one of those areas. Article 89(2) of the GDPR enables the Member States to provide for further derogations from the right of access, rectification, restriction, portability and the right to object in cases where processing is carried out for research purposes. These national derogations can only apply where the right is likely to render impossible or seriously impair the achievement of the specific purposes, and a derogation is necessary for the fulfilment of those purposes. In combination with limitations provided for in the GDPR itself (see above), these optional derogations have the potential of almost completely neutralizing the rights of data subjects in the context of research, and effectively depriving individuals of any control over research data.

It seems, however, that Member States used this prerogative with caution. For example, the German Federal Data Protection Act provides for exceptions from the rights of access, rectification, restriction and the right to object (but not to the right of portability).[34] The French Data Protection Act provides for an exception to the right of access, but only where the data are stored in a form that manifestly excludes any risks for privacy (a condition which, interpreted strictly, would be equal to anonymization, which deprives the exception of practical significance).[35] Some other countries chose to make the optional derogations only to research carried out in the public interest (European Parliament, Scientific Foresight Unit 2019, p. 25). Regretfully, this contributes to the fragmentation of the legal landscape applicable to research within the EU, which was one of the major problems of the Personal Data Directive that the GDPR was expected to overcome. These national specificities are not *per se* obstacles to data sharing in European research projects, but they need to be taken into account in cross-border projects.

VI. Conclusion

Admittedly, after more than four years since the GDPR's adoption and more than two years after its entry into force, it is still difficult to navigate through certain areas of the new EU data protection framework. Scientific research in general, and DH research in particular, is one of these areas. New studies and guidelines from various data protection stakeholders will hopefully be published in the following months and shed more light on the still blurry picture. Nevertheless, it can be said that the GDPR certainly contains mechanisms that strengthen the control that data subjects can exercise over their personal data, such as a high standard for valid consent, or the rights of erasure or portability. These mechanisms are counterbalanced with numerous safety valves, aimed at allowing the data controller to circumvent them where certain (usually strict, at least in theory) conditions are met, for example, where the processing serves legitimate or public interest. The author feels that, while it probably is still too early to evaluate the overall balance of this framework, the GDPR provides for a good compromise between the various interests at stake.

Specifically in the context of DH research, despite the fact that research material will often qualify as personal data, it seems that research projects respecting high ethical and academic standards (i.e. implementing appropriate safeguards for rights and freedoms of data subjects), especially where data processing is based on legitimate or public interest instead of consent, will rarely have to accommodate risks related to exercise of data subjects' control over the research data. This does not mean, however, that such projects can completely disregard the GDPR, as many obligations, such as maintaining a record of processing activities, ensuring data security, documenting and properly addressing data breaches or, where necessary, carrying out a Data Protection Impact Assessment, continue to apply.[36]

Notes

1 This quote is (commonly, but probably mistakenly) attributed to V. I. Ulyanov, known as Lenin.
2 Warren & Brandeis (1890). On the impact of this paper see Kramer (1990). Both Warren and Brandeis were relatively young when the paper was published (38 and 34 respectively); Brandeis later became associate justice of the US Supreme Court (1916–1939).
3 The first European privacy law was adopted in the German state of Hessen. In fact, when the French law was adopted, Germany already had a federal data protection law, which introduced the basic concepts around which European data protection law was later harmonized.
4 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
5 Formally, the GDPR does not apply retroactively. However, it applies to all processing operations after 25 May 2018, and this includes storage. Therefore, in order to be able to lawfully store the data after that date, the controller needs to meet all the GDPR requirements (unless the data subjects are now dead, in which case their data are in principle outside the scope of the GDPR).
6 The statement "Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say" is attributed to Edward Snowden (Reddit, 21 May 2015); https://en.wikiquote.org/wiki/Edward_Snowden [Viewed 15 July 2020].
7 See, e.g., European Commission (2018); cf. Recital 7, par. 2 of the GDPR ("Natural persons should have control of their own personal data").
8 Bundesverfassungsgericht, 15 December 1983 (Volkszählung), 1 BvR 269/ 83, BVerfGE 65, 1; see also Rouvroy & Poullet (2018).
9 Unofficial English translation available from: https://freiheitsfoo.de/files/2013/10/Census-Act.pdf (viewed 15 July 2020).
10 Another well-established approach to privacy—privacy as (restricted/limited) access—is driven by individuals' concern over their accessibility to others. On this approach, see esp. Gavison (1980, pp. 421–423).
11 Cf. Article 544 of the Napoleonic Code 1804.
12 Already in 1967 Harold Demsetz accurately predicted that "the emergence of new private or state-owned property rights will be in response to changes in technology and relative prices" (Demsetz 1967, p. 350).
13 See especially European Commission (2017) and most recently European Commission (2020). For a comprehensive analysis, see Stepanov (2020).
14 See especially Samuelson (2000) (critically), Liebenau (2016), and more recently Trakman et al. (2019).
15 This is the number of pages in the Official Journal of the European Union, see: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN
16 Article 288 of the Treaty on the Functioning of the European Union (TFUE). Available from: http://data.europa.eu/eli/treaty/tfeu_2012/oj
17 See esp. the California Consumer Privacy Act (CCPA) of 2018, or a bill to establish a Federal Data Protection Agency (S 3030) introduced on 13 February 2020 by US Senator Kirsten Gillibrand.
18 The framework for transfers is laid down in chapter 5 of the GDPR (Articles 44–50). One of the alternatives to either being located in a country that provides for appropriate level of data protection or ensuring such level of protection via contractual agreements (Standard Contractual Clauses, Code of Conduct or Binding Corporate Rules) is explicit consent given by the data subject after having been informed of the possible risks. This ground, however, can only be used in exceptional situations, and not in the regular course of action (cf. EDPB 2018).

19 E.g. § 22(2) of the German Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG; Available from: https://iapp.org/media/pdf/resource_center/Eng-trans-Germany-DPL.pdf) lists the following examples of "appropriate and specific measures":

1 technical organizational measures to ensure that processing complies with Regulation (EU) 2016/679;
2 measures to ensure that it is subsequently possible to verify and establish whether and by whom personal data were input, altered or removed;
3 measures to increase awareness of staff involved in processing operations;
4 designation of a data protection officer;
5 restrictions on access to personal data within the controller and by processors;
6 the pseudonymization of personal data;
7 the encryption of personal data;
8 measures to ensure the ability, confidentiality, integrity, availability and resilience of processing systems and services related to the processing of personal data, including the ability to rapidly restore availability and access in the event of a physical or technical incident;
9 a process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the processing;
10 specific rules of procedure to ensure compliance with this Act and with Regulation (EU) 2016/679 in the event of transfer or processing for other purposes.

20 The definition of personal data (personenbezogene Daten) in § 2(1) of the Bundesdatenschutzgesetz (in its initial version of 1977) reads: *(1) Im Sinne dieses Gesetzes sind personenbezogene Daten Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener)*.
21 Article 85 of the French Data Protection Act (*La loi no 78–17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*).
22 WP29 (2007, p. 15). Recital 26 of the GDPR lists two criteria to be taken into account in evaluating whether a person can be identified: costs and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.
23 Court of Justice of the European Union, Judgement of 19 October 2016 in Case C—582/14 (*Breyer*).
24 Court of Justice of the European Union, Judgement of 17 July 2014 in joined Cases C—141/12 and C—372/12 (*YS and others*).
25 *Idem*, pt. 45–46, 57.
26 Court of Justice of the European Union, Judgement of 20 December 2017 in Case C—434/16 (*Nowak*).
27 For more information on the balancing test, see WP29 (2014b, esp. pp. 33–43).
28 For more information on purpose extension, see WP29 (2013, pp. 19–36).
29 As announced in a LinkedIn article published by Greet Gysen (EDPB's Head of Activities) on 8 January 2020: www.linkedin.com/pulse/getting-data-subject-rights-right-greet-gysen/ (viewed on: 17 July 2020).
30 European Data Protection Supervisor (2014, p. 18). The document does not concern the GDPR (which it predates), but EU Regulation 45/2001 on the processing of personal data by European Union institutions and bodies (which has since been replaced by the Regulation 2018/1725). However, the right of erasure in this document seems to be essentially similar to the one in the GDPR.
31 See esp. Court of Justice of the European Union, Judgement of 13 May 2014 in Case C—131/12 (*Google Spain*), to which the origins of the right to be forgotten can be traced.
32 For an in-depth analysis of this question, see Graef et al. (2018, esp. pp. 1365–1375).
33 In such cases, the provision of information is gathered by Article 13 (and not 14) of the GDPR.
34 § 27(2) of the German Federal Data Protection Act (*Bundesdatenschutzgesetz*, BDSG).

35 Article 49 of the French Data Protection Act (*La loi no 78–17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*).

36 On maintaining a record of processing activities, see Article 30; on ensuring data security, Articles 5(1)(f) and 32; on documenting and properly addressing data breaches, Articles 33 (esp. 33(5) about documentation), 34, and WP29 (2018a); on carrying out a Data Protection Impact Assessment, Articles 35–36 and WP29 (2017b). On the idea of a code of conduct for language research, which could help harmonize practice and achieve greater legal security (through official approval mechanism) in the community, see Kamocki et al. (2018).

References

Allen, A. L., (2000). Privacy-as-data control: Conceptual, practical, and moral limits of the paradigm. *Connecticut Law Review*. 32, 861–875.

Article 29 Data Protection Working Party., (2007). *Opinion 4/2007 on the concept of personal data*. Adopted on 20 June (WP 136). [Viewed 16 July 2020]. Available from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

Article 29 Data Protection Working Party., (2013). *Opinion 03/2013 on purpose limitation*. Adopted on 2 April (WP 203). [Viewed 16 July 2020]. Available from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

Article 29 Data Protection Working Party., (2014a). *Opinion 05/2014 on anonymisation techniques*. Adopted on 10 April (WP 216). [Viewed 16 July 2020]. Available from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Article 29 Data Protection Working Party., (2014b). *Opinion 06/2014 on the notion of legitimate interests of the data controller under article 7 of Directive 95/46/EC*. Adopted on 9 April (WP 217). [Viewed 16 July 2020]. Available from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

Article 29 Data Protection Working Party., (2017a). *Guidelines on the right to data portability*. Adopted on 13 December 2016 as last Revised and adopted on 5 April (WP 242 rev. 01). [Viewed 17 July 2020]. Available from: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233

Article 29 Data Protection Working Party., (2017b). *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is 'likely to result in a high risk' for the purposes of regulation 2016/679*. Adopted on 4 April as last revised and adopted on 4 October (WP 248 rev. 01). [Viewed 17 July 2020]. Available from: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

Article 29 Data Protection Working Party., (2018a). *Guidelines on personal data breach notification under regulation 2016/679*. Adopted on 3 October as last revised and adopted on 6 February (WP 250 rev. 01). [Viewed 17 July 2020]. Available from: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612052

Article 29 Data Protection Working Party., (2018b). *Guidelines on transparency under regulation 2016/679*. Adopted on 29 November as last revised and adopted on 11 April (WP 260 rev. 01). [Viewed 17 July 2020]. Available from: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

Boucher, P., (1974). 'Safari' ou la chasse au Français. *Le Monde*. 21 March, p. 9. [Viewed 15 July 2020]. Available from: www.cnil.fr/sites/default/files/atoms/files/le_monde_0.pdf

Bradford, A., (2020). *The Brussels effect: How the European Union rules the world*. New York: Oxford University Press.

Demsetz, H., (1967). Toward a theory of property rights. *The American Economic Review*. 57(2), 347–359.

Dove, E. S., (2018). The EU General Data Protection Regulation: Implications for international scientific research in the digital era. *The Journal of Law, Medicine & Ethics*. 46(4), 1013–1030. https://doi.org/10.1177/1073110518822003

Enev, M., Takakuwa, A., Koscher, K. and Kohno, T., (2016). Automobile driver fingerprinting. *Proceedings on Privacy Enhancing Technologies*. 1, 34–51. https://doi.org/10.1515/popets-2015-0029

European Commission., (2012). *Proposal for a Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data*. 25 January, COM/2012/010 final. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012PC0010&from=en

European Commission., (2017). *Building a European data economy*. European Commission. 10 January, COM(2017)09final. [Viewed 21 July 2020]. Available from: https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy

European Commission., (2018). *It's your data: Take control: Data protection in the EU*. European Commission. [Viewed 15 July 2020]. Available from: https://ec.europa.eu/info/sites/info/files/data-protection-overview-citizens_en_0.pdf

European Commission., (2020). *A European strategy for data*. European Commission. 19 February, COM(2020) 66 final. Available from: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

European Data Protection Board (EDPB)., (2018). *Guidelines 2/2018 on derogations of article 49 under regulation 2016/679*. Adopted on 25 May. [Viewed 17 July 2020]. Available from: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_2_2018_derogations_en.pdf

European Data Protection Board (EDPB)., (2019). *Guidelines 3/2018 on the territorial scope of the GDPR (Article 3)*. Version 2.1. 12 November. [Viewed 16 July 2020]. Available from: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_3_2018_territorial_scope_after_public_consultation_en_1.pdf

European Data Protection Board (EDPB)., (2020). *Guidelines 05/2020 on consent under regulation 2016/679*. Version 1.1. Adopted on 4 May. [Viewed 16 July 2020]. Available from: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf

European Data Protection Supervisor (EDPS)., (2014). *Guidelines on the rights of individuals with regard to the processing of personal data*. European Data Protection Supervisor. 25 February. [Viewed 17 July 2020]. Available from: https://edps.europa.eu/sites/edp/files/publication/14-02-25_gl_ds_rights_en.pdf

European Data Protection Supervisor (EDPS)., (2020). *A preliminary opinion on data protection and scientific research*. European Data Protection Supervisor. 6 January. [Viewed 17 July 2020]. Available from: https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf

European Parliament Research Service, Scientific Foresight Unit., (2019). *How the General Data Protection Regulation changes the rules for scientific research*. European Parliament Research Service, Scientific Foresight Unit (STOA). Panel for the Future of Science and Technology. July (PE 634.447). [Viewed 16 July 2020]. Available from: www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf

Fried, C., (1968). Privacy. *The Yale Law Journal*. 77(3), 475–493.

Gavison, R., (1980). Privacy and the limits of law. *Yale Law Journal*. 89(3), 421–471.

Gayomali, C., (2014). NYC taxi data blunder reveals which celebs don't tip and who frequents strip clubs. *FastCompany.com* [online]. 2 October. [Viewed 15 July 2020].

Available from: www.fastcompany.com/3036573/nyc-taxi-data-blunder-reveals-which-celebs-dont-tip-and-who-frequents-strip-clubs

Graef, I., Husovec, M. and Purtova, N., (2018). Data portability and data control: Lessons for an emerging concept in EU law. *German Law Journal*. 19(6), 1359–1398.

Kamocki, P., Ketzan, E., Wildgans, J. and Witt, A., (2018). Toward a CLARIN data protection code of conduct. *Proceedings of the CLARIN Annual Conference 2018 Proceedings, 8–10 October 2018, Pisa*. pp. 49–52. [Viewed 17 July 2020)]. Available from: https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_ConferenceProceedings.pdf

Kramer, I. R., (1990). The birth of privacy law: A century since Warren and Brandeis. *Catholic University Law Review*. 39(3), 703–724.

Liebenau, D., (2016). What intellectual property can learn from information privacy, and vice versa. *Harvard Journal of Law & Technology*. 30(1), 285–307.

Purtova, N., (2018). The law of everything: Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*. 10(1), 40–81. https://doi.org/10.1080/17579961.2018.1452176

Rocher, L., Hendrickx, J. M. and de Montjoye, Y.-A., (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*. 10, article number 3069. https://doi.org/10.1038/s41467-019-10933-3

Rouvroy, A. and Poullet, Y., (2018). The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy. In: S. Gutwirth, Y. Poullet, P. De Hert, C. de Terwangne and S. Nouwt, eds. *Reinventing data protection?* Dordrecht: Springer. pp. 45–76.

Samuelson, P., (2000). Privacy as intellectual property? *Stanford Law Review*. 52(5), 1125–1174.

Schwartz, P. M., (2000). Internet privacy and the state. *Connecticut Law Review*. 32, 815–859.

Stepanov, I., (2020). Introducing a property right over data in the EU: The data producer's right: An evaluation. *International Review of Law, Computers & Technology*. 34(1), 65–86. https://doi.org/10.1080/13600869.2019.1631621

Trakman, L., Walters, R. and Zeller, B., (2019). Is privacy and personal data set to become the new intellectual property? *International Review of Intellectual Property and Competition Law*. 50, 937–970. https://doi.org/10.1007/s40319-019-00859-0

Van Ooijen, I. and Vrabec, H. U., (2019). Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. *Journal of Consumer Policy*. 42, 91–107. https://doi.org/10.1007/s10603-018-9399-7

Warren, S. D. and Brandeis, L. D., (1890). The right to privacy. *Harvard Law Review*. 4(5), 193–220.

Westin, A. F., (1967). *Privacy and freedom*. New York: Atheneum.

Zorz, Z., (2018). One in 10 C-level execs say GDPR will cost them over $1 million. *HelpNetSecurity.com* [online]. 13 April. [Viewed 15 July 2020]. Available from: www.helpnetsecurity.com/2018/04/13/gdpr-compliance-costs/