What does it mean for a data subject to make their personal data 'manifestly public'? An analysis of GDPR Article 9(2)(e)

What does it mean for a data subject to make their personal data 'manifestly public'? An analysis of GDPR Article 9(2)(e)

Edward S. Dove* and Jiahong Chen†

Key points

- This article investigates an under-discussed and potentially significant provision in the EU General Data Protection Regulation (GDPR), namely Article 9(2)(e), which permits processing of special category personal data if the 'processing relates to personal data which are manifestly made public by the data subject'.
- This provision may be of increasing interest to data controllers in a variety of cloud-based, internet-related, and/or social media contexts. We specifically consider the application of this provision in the context of genetic data and open data sharing (ie data that can be freely used, re-used, and redistributed by anyone), illustrating this by way of several cases of initiatives that seek to share genetic data. We query whether by uploading one's own genetic data onto the internet, a person has made their data 'manifestly public' within the meaning of the GDPR.
- Our response to this query is that in general, the answer should be no, but it remains possible. We argue that Article 9(2)(e) must be construed narrowly; outside of clearly defined contexts, it would be legally inappropriate to invoke and rely upon this manifestly public self-disclosure exception in data protection law. Our narrow interpretation of the provision aligns with the limited guidance made available from data protection authorities. As part of this argument, we propose a legal test that must be satisfied before Article 9(2)(e) may be lawfully invoked, and which is grounded in the intent of the data subject.

Introduction

Under European data protection law, certain kinds of personal data are considered 'special'—in other words, sensitive—and therefore deserving of even greater legal protection than that afforded to non-special (regular) personal data. As Recital 51 of the EU's General Data Protection Regulation (GDPR)[1] states, a high level of protection for sensitive data is regarded as necessary since they 'are, by their nature, particularly sensitive in relation to fundamental rights and freedoms [and thus] merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms'.[2] Special categories of personal data are those revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership; it also covers data concerning health and data concerning a natural person's sex life or sexual orientation; and it covers the processing of genetic data or biometric data for the purpose of uniquely identifying a natural person.[3] Whereas with non-special

* Edward S. Dove, School of Law, University of Edinburgh, Edinburgh, UK. Email: edward.dove@ed.ac.uk.

† Jiahong Chen, Horizon Digital Economy Research, University of Nottingham, Nottingham, UK

The authors would like to thank Regina Becker and Mark Taylor for their comments on a previous draft. Jiahong Chen's work was supported by the Engineering and Physical Sciences Research Council (grant numbers EP/M02315X/1, EP/T022493/1).

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ 2016 L 119/1 (General Data Protection Regulation) [hereinafter, GDPR].

2 For a thorough analysis of the concepts of the sensitive data and health data in EU data protection law, see Paul Quinn and Gianclaudio Malgieri, 'The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework' (2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3713134> accessed 4 February 2021.

3 Art 9(1) GDPR.

(regular) personal data, processing is lawful only where there is a legal basis, with special categories of data, processing is generally prohibited; it will only be permitted if the person processing the data has a legal basis *and* meets a special category condition (ie an exception to the rule against processing such data). In other words, processing 'special categories' of personal data requires two conditions:

1. the processing must have a legal basis, ie one of the six legal bases outlined in Article 6(1) of the EU's General Data Protection Regulation (GDPR), and
2. it must fall within at least one of the 10 exceptions specified in Article 9(2) of the GDPR.

Since the days of the (now-superseded) EU's Data Protection Directive 95/46/EC (DPD),[4] a provision has been available to data controllers wanting to process special categories of data. This provision was found in the special category data section of the DPD—Article 8(2)(e)—and finds itself in the equivalent section of the GDPR—at Article 9(2)(e). It has been phrased identically in both statutes: 'Paragraph 1 [the prohibition of processing special category personal data] shall not apply if one of the following applies: [...] processing relates to personal data which are manifestly made public by the data subject'. In both statutes, the meaning of the phrase 'manifestly made public by the data subject' is not defined. And, interestingly, whereas several of the exceptions that are provided in Article 9(2)—such as the possibility to process special category data for reasons of 'scientific research', 'public health', and 'substantial public interest'—are defined in terms of a 'purpose', Article 9(2)(e) makes no reference to the purposes of the data controller. What makes this provision even more special is the fact that EU data protection law does not generally make a substantial distinction between personal data in a private space and in a public one.[5] Unlike in jurisdictions where the 'third-party doctrine' applies,[6] disclosing one's own personal data publicly in the EU does not automatically render the data no longer protected. The exemption therefore raises several immediate questions regarding an interpretative approach that would ensure consistency with the overall objectives of EU data protection law.

Looking to guidance from European regulatory authorities as to the meaning of this phrase, one is struck by the relative paucity of information. There is little guidance from national data protection authorities (DPAs) or the European Data Protection Board (EDPB) on Article 9(2)(e); nor does there seem to be much precedent for its invocation. Consequently, it may be seen as an untested pathway for controllers to process special category personal data without having to fulfil onerous obligations associated with other exceptions under Article 9(2). Moreover, though a large part of the limited guidance available recommends that the phrase 'manifestly made public by the data subject' should be construed narrowly—and indeed, given that Article 9(1) prohibits processing special category data unless an exemption applies under Article 9(2), a narrow interpretation should be applied to all of the listed exemptions[7]—because it is largely untested in law, its scope remains uncharted and thus this position of narrow interpretation may not be strictly followed by data controllers. And, given the challenges associated with securing other exceptions under Article 9(2), such as explicit consent in some data processing activities (eg resources for obtaining data subject consent for large-scale processing activities, concern about implications of consent withdrawal by the data subject), 'manifestly made public by the data subject' may be seen as a lighter-touch exception for data controllers to pursue, in contexts such as the data subject publishing their data in the mass media or putting them on online social networks, the latter of which is a vastly growing area of communication and commerce—not to mention data processing.[8] This may be especially apparent in the context of large-scale genomic research activities, whereby research organizations seek to collect biological samples from thousands of participants, sequence them for genetic analysis, 'interpret' them, and in turn generate hypotheses, research breakthroughs, effective treatments, and ideally, innovative treatments. Moreover, Article

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, OJ 1995 L 281/31 [hereinafter, Data Protection Directive].

5 Lilian Edwards, 'Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective' (2016) 2(1) European Data Protection Law Review 28.

6 Eunice Park, 'Objects, Places and Cyber-Spaces Post-Carpenter: Extending the Third-Party Doctrine beyond CSLI: A Consideration of IoT and DNA' (2019) 21 Yale Journal of Law and Technology 1.

7 Ludmila Georgieva and Christopher Kuner, 'Article 9. Processing of Special Categories of Personal Data', in Christopher Kuner, Lee Bygrave and Christopher Docksey (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 375 ('The list of exceptions is exhaustive and all of them are to be interpreted restrictively').

8 Video surveillance alone is unlikely to fulfil the conditions of art 9(2)(e), as we elaborate later in this article. We note that the European Data Protection Board has stated that '...data controllers processing [special categories of personal] data in the context of video surveillance cannot rely on Article 9 (2) (e), which allows processing that relates to personal data that are manifestly made public by the data subject. The mere fact of entering into the range of the camera does not imply that the data subject intends to make public special categories of data relating to him or her'. See European Data Protection Board, *Guidelines 3/2019 on Processing of Personal Data through Video Devices* (2019) para 69.

9(2)(e) may be seen as an especially attractive option in the context of open data sharing, ie for those research endeavours that seek to make data, personal or otherwise, openly available to people around the world without restriction. If data subjects upload their genetic data onto the internet, it might be the case that researchers (and others) could make use of such data without the data subject's explicit consent as per Article 9(2)(a). In our view, the heightened sensitivity of uploading *genetic* data, given that these data are particularly prone to re-identification and contain identifying information about genetic relatives,[9] makes the applicability of Article 9(2)(e) of more acute current concern and, as we will argue in this article, explains why Article 9(2)(e) should be read restrictively in such novel contexts.

Genomic research organizations may be reluctant to rely on Article 9(2)(e) to process genetic data and data concerning health, though, in the absence of any clear regulatory guidance or precedent. And given the paucity of such guidance and precedent, we lack a good understanding of what this provision means. We do not know what the permitted role of intermediaries (e.g. internet service providers, search engines, social media platforms) is, if any, in the process of making an individual's data manifestly public; nor do we know the conditions under which it is necessary to satisfy Article 9(2)(e) as distinct from those needed to satisfy explicit consent under Article 9(2)(a) or scientific research under Article 9(2)(j)—other exceptions that are relevant in the context of large-scale genomic research activities.

Thus, the aim of this article is to chart the contours of Article 9(2)(e) and specifically to consider its application in the context of genetic data and open data sharing.[10] We argue that, even by the strict standards of Article 9(2) interpretation, Article 9(2)(e) must be construed narrowly; outside of clearly defined contexts, it would be legally inappropriate to invoke this exception. Our narrow interpretation aligns with the limited—and rather generic—guidance made available from DPAs. As part of this argument, we propose a legal test that must be satisfied before Article 9(2)(e) may be lawfully invoked, grounded in the intent of the data subject. Recognizing that grounding the test in the intent of the data subject may create some risk in undermining the certainty or workability of the exemption for the data controller, we consider that this test does not amount to an 'absolute' approach. Instead, what is required is objective evidence based on that which the data controller may reasonably establish to be such an intent.

In what follows, in 'The case of genetic data: which special category exception?' section, we discuss several case studies of international large-scale genomic research activities. Using these case studies for subsequent analysis, we query whether by uploading one's own genetic data onto the internet, a person has made their data 'manifestly public' within the meaning of Article 9(2)(e). In the 'Explicit consent and public self-disclosure as voluntary grounds: why does it matter?' section, we elaborate in depth the significance of ascertaining the appropriate exemption under Article 9(2), explaining why relying on different conditions may lead to different levels of protection for data subjects. As part of this, we highlight the risks of using public availability (Article 9(2)(e)) as a way around certain data controller duties. With such a risk in mind, in 'The curious case of 'manifest' self-disclosure in data protection law' section, we further consider various interpretations of Article 9(2)(e) and apply it to the context of the voluntarily permissive mechanisms under the GDPR, arguing that the lines between (explicit) consent and public disclosure are sometimes unclear; though they both represent voluntary permissions of using (sensitive) personal data, the legal and practical consequences are starkly different. In 'A legal test for Article 9(2)(e)' section, we consider the role of intermediaries in the process of making an individual's data manifestly public, and the conditions under which it is necessary to satisfy Article 9(2)(e) as distinct from those needed to satisfy explicit consent under Article 9(2)(a). We also propose a legal test that must be satisfied before Article 9(2)(e) may be lawfully invoked, grounded in the intent of the data subject; as part of this test, we unpack the individual elements of the provisions, including the meaning of the phrase 'by the data subject'. Finally, we conclude with parting thoughts about the scope and suitability of Article 9(2)(e) as a condition generally in data protection law, and specifically in the context of genetic research.

We begin our assessment by looking at several genetic research and data-sharing initiatives, posing the question: which special category exception under Article

9 Thomas Finnegan and Alison Hall, *Identification and Genomic Data* (PHG Foundation 2017) <https://www.phgfoundation.org/documents/PHGF-Identification-and-genomic-data.pdf> accessed 4 February 2021.

10 Another emerging scenario in the context of art 9(2)(e) is carrying out research on publicly available social media data, such as those on Twitter, TikTok, Instagram, or Facebook. While we do not explore this context in the article, our analysis is intended to be applicable across domains.

9(2) may be appropriate to process participants' genetic data?

The case of genetic data: which special category exception?

As mentioned in the previous section, Article 9(2)(e) may be seen to apply in different contexts, such as publishing data in the mass media or putting them on online social media platforms; in our view, one of the more interesting and novel (that is, largely untested) contexts where this provision might apply is large-scale genomic research activities that promote open sharing of such data. In this section, we introduce several case studies of initiatives (ie platforms) that seek to share genetic data in different ways, on a scale of semi-open to fully open: GenomeConnect, Personal Genome Project, and openSNP.

GenomeConnect

GenomeConnect is an American-based not-for-profit online registry designed by the Clinical Genome Resource (ClinGen, a National Institutes of Health-funded resource) for people who are interested in sharing 'de-identified' genetic and health information to improve understanding of genetics and health.[11] Participants who register are also able to connect with other individuals and families through GenomeConnect's participant matching feature. Participants can learn about other research opportunities and receive updates about their genetic testing results. The registry is open to anyone who has had genetic testing, is considering testing, or has a family member that has had testing regardless of test results or diagnosis. Interestingly, it appears that participants can only consent to all uses of the data outlined by GenomeConnect; they cannot consent only to some aspects of the service.

The process works whereby participants register online, consent to participate and share their health history by completing an online survey, and then upload a copy of their genetic resting report(s) to their GenomeConnect account. GenomeConnect welcomes participants from any country in the world, provided they are able to read and understand English; however, they do not process data from participants in the EU/EEA or UK due to the GDPR.[12]

In the GenomeConnect Participant Matching feature, the uploaded genetic testing reports are used to create a searchable gene list. Once a participant uploads their report, GenomeConnect's team of genetic counsellors review the report and any genes listed are added to the searchable gene list.

GenomeConnect shares the de-identified health and genomic data with approved users and databases. In other words, the personal data on the GenomeConnect database itself remain private and only ClinGen staff can access it. From this, we can gather that while GenomeConnect does not include participants in the EU/EEA or UK as it seeks to avoid the long-arm reach of the GDPR, even if, hypothetically speaking, the GDPR were to apply, GenomeConnect would likely argue that the legal basis for processing personal data of participants is consent, albeit arguably not to the GDPR standard (it would have to demonstrate that this consent is freely given), and the permitted exception to process their genetic and health data is explicit consent. Given data are uploaded into a semi-open (or semi-closed, depending on one's perspective) database and made available only to approved users in a de-identified format, as we will see, it is difficult to see how Article 9(2)(e) could be confidently relied upon.

Personal Genome Project

The Personal Genome Project (PGP) is a global project initiated in 2005, with the aim of creating and sharing public genome, health, and phenotypic trait data.[13] The PGP recruits people who are expressly willing to publicly share their personal data. Thus, in contrast to GenomeConnect, which shares only de-identified data with approved external researchers (and others), PGP aims to have *all* personal data be freely available to everyone, in line with the ethos of open data and open science. To date, PGP has projects in the USA, Canada, UK, Austria, and China.

Uniquely, the risks of participant re-identification are addressed up front, as an integral part of the consent and enrolment process. Participants are informed that neither anonymity nor confidentiality of their identities or their data can be guaranteed. To qualify for enrolment, prospective participants must complete an online exam to demonstrate their comprehension of the risks

11 GenomeConnect, available at <https://www.genomeconnect.org/> accessed 4 February 2021.

12 GenomeConnect, 'Questions from Participants' <https://www.clinicalgenome.org/genomeconnect/for-patients-genomeconnect/faq/questions-from-participants/> accessed 4 February 2021.

13 Personal Genome Project <https://www.personalgenomes.org/gb> accessed 4 February 2021.

and protocols associated with being a member of the PGP. In the UK project, for example, participants must also be a UK citizen or permanent resident, at least 21 years of age, and capable of giving consent.[14]

The information sheet and consent form for PGP-UK[15] specifies that University College London (UCL) is the sponsor of PGP-UK and will keep identifiable information about participants for five years until 2028, after the study finishes in 2023. Though not explicitly stated, it would seem that the GDPR legal basis and special category exception for processing health and genetic data being relied upon is (explicit) consent, as indicated in the information sheet, though certain phrasing of language gives some doubt as to the 'freely given' nature of the consent:

> By signing this consent form, you authorize the PGP-UK to publish your specimen analysis data and other personal information you have submitted to the PGP-UK. This means that the PGP-UK may publish this data and information without legal restriction and without your being asked to provide any additional consent. The PGP-UK will publish the data and information on a publicly accessible website and database. It may also publish the data and information in other formats and/or media. Your ability to withdraw your consent once the PGP-UK has published all or some of this data and information is limited and is described in Article X of this consent form. There may be risks to you associated with the publication of this data and information. Those risks are described in Article X of this consent form.
> [. . .]
> By signing this consent form, you authorize the PGP-UK to distribute to others, in a manner consistent with the conditions described above, the cell lines created from your tissues without further notice to you and without the PGP-UK obtaining any additional consent from you. Your ability to withdraw from the portion of this study that involves the distribution of your cell lines is limited once your cells lines have been created and distributed and is described in Article X of this consent form. There may be risks to you associated with the creation and distribution of your cell lines. Those risks are described in Article VI of this consent form.[16]

The form itself does not explicitly state the legal basis for processing personal data and such information could not be located on the website in a privacy notice (informational requirements under Article 13 GDPR). This said, there is some ambiguity with the phrase, 'By signing this consent form, you authorize the PGP-UK to publish your specimen analysis data and other personal information you have submitted to the PGP-UK', potentially signalling potential manifest self-disclosure by the participant and reliance on Article 9(2)(e) as the permitted exception to lawfully process genetic and health data.

openSNP

openSNP is a crowd-funded initiative that enables customers of direct-to-customer (DTC) genetic tests (such as those from the company 23andMe) to publish online their test results, find other people with similar genetic variations, learn more about their results by getting the latest primary literature on their variations, and help scientists find new associations.[17] This is done by having participants upload their raw genotype data that have been downloaded from their DTC test provider. As with GenomeConnect and PGP, participants sign up to participate by registering online. Unlike the other two projects, however, openSNP offers a relatively light-touch consent form; moreover, they appear to rely on a form of manifest self-disclosure akin to Article 9(2)(e) for processing special category personal data. Indeed, because the purpose of the processing is not made clear, we have some doubt that consent would be the legal basis upon which they rely. openSNP asks participants to recognize '. . .that you have understood the possible risks and side-effects that can occur by making your genetical [sic] and medical information available on this platform', that '. . .all data you upload to openSNP will be freely available online (well, except your mail-address and password) under a Creative Commons Zero license', and that 'There is zero privacy anyway, get over it'.[18] Likewise, in the FAQ section of the website, in response to the question, 'How open is my data?', the reply is: 'Completely open: Everyone can see everything you enter or upload (except your private messages and your password, of course). We warn every user twice about this: Once during the user-creation and once before the genotyping-file-upload'.[19]

14 PGP-UK Enrolment Process and Requirements <https://www.personalgenomes.org.uk/volunteer/> accessed 4 February 2021.

15 Informed Consent for Enrolment in the PGP-UK, UCL Research Ethics Project Number 4700/100 <https://www.personalgenomes.org.uk/assets/docs/uk/PGP-UK_FullConsent_v20_October2018.pdf> accessed 4 February 2021.

16 Ibid 8.

17 openSNP <https://www.opensnp.org/> accessed 4 February 2021.

18 openSNP, 'Signup' <https://www.opensnp.org/signup> accessed 4 February 2021.

19 openSNP, 'FAQ' <https://www.opensnp.org/faq> accessed 4 February 2021.

Summarizing the three platforms

In sum, then, though it is not made explicitly clear, at least some of these three platforms appear to rely on a form of Article 9(2)(e) as the permitted exception to process health and genetic data. For PGP and openSNP, the data are made public. Downstream researchers (and others) are unable to contact participants for consent. These platforms might rely on the scientific research exception, Article 9(2)(j) (discussed below), but they could, in principle, equally rely on Article 9(2)(e): participant–data subjects have agreed to make their data public and have been informed that their data are being made public. However, this is not satisfyingly clear to us, and thus, the legal question remains: by uploading one's own genomic data onto the internet, on what basis can someone process their data? In other words, what is the specific exception under Article 9(2) to process the data? And, once the data are published on the platform, what is the specific exception under Article 9(2) for downstream users to perform subsequent data processing for their own purposes? Must the platform and others (ie researchers from other organizations), in order to utilize such data, obtain the data subject's explicit consent as per Article 9(2)(a), including every time they use the data, or can they rely on another exception, such as that the processing is for scientific research purposes as per Article 9(2)(j) and in accordance with Article 89(1), or indeed that the data have been manifestly made public as per Article 9(2)(e) and thus the data can be used repeatedly and in some sense, indefinitely and by anyone?

These questions are all the more important because were consent to be the lawful basis under Article 6(1) *and* the permitted exception under Article 9(2), if a participant subsequently decided to withdraw their consent for data processing, in principle (though not without some concern about fairness), a data controller may then seek to 'switch' to an alternative legal basis and permitted exception,[20] such as legitimate interests and scientific research, or indeed rely on the basis that the data subject, by uploading their genetic data to the internet, has manifestly made it public. The choice of such a legal basis and permitted exception will determine the legal duties imposed on the data controller as well as the safeguards available to the data subject, which we now proceed to analyse in detail in the next section.

Explicit consent and public self-disclosure as voluntary grounds: why does it matter?

To understand why it is of practical significance for an open genome platform to choose the appropriate exemption under Article 9(2) and to make it explicit and clear to both the data subject and downstream data users, one first needs to analyse the legal consequences resulting from such a choice. Specifically, this concerns the legal effects of a data subject's action when they sign up for an open genome programme and upload their genomic data. As will be shown below, relying on the data subject's manifestly public self-disclosure does not mean the data controller can avoid their ongoing legal and ethical responsibilities. Yet, the selection between explicit consent, scientific research, and manifestly public self-disclosure would still make a significant difference on a range of matters. We focus in particular on contrasting Article 9(2)(a) with 9(2)(e).

First, consent as a valid legal basis under Article 6(1)—and as a permitted exception under Article 9(2)—is subject to a wider range of further conditions, mainly set out in Articles 7 and 8. One obvious difference is that while the data subject has 'the right to withdraw his or her consent at any time',[21] such a right does not apply when a legal basis other than consent is invoked under Article 6(1) and another permitted exception is invoked under Article 9(2). Consequently, the data controller would not have to inform the data subject of this right. This means that if the data subject is deemed to have manifestly made their data public, they will not be able to restrict downstream uses of such data as one would by withdrawing their consent. This does not mean that once data are in the public domain, the data subject no longer has any ability to control their use. Conceivably, a data subject on certain platforms can 'take down' the data by removing them from the platform; this, of course, is easier to do in a digital context than an analogue one, but it does not guarantee that further processing will no longer take place. It is arguable that once a data subject has removed their

20 See art 17(1)(b) GDPR. However, it should be noted that such a switch to an alternative legal basis must be communicated to the data subject and cannot be exploited to retrospectively remedy flawed consent. See Article 29 Data Protection Working Party, *Guidelines on Consent under Regulation 2016/679* (WP259 rev.01, 2018), 22–23 <https://ec.europa.eu/newsroom/article29/document.cfm?doc_id=51030> accessed 4 February 2021. There are also concerns that it may be inherently unfair to tell people they have a choice as to whether to consent to have their data processed, but then continue the data processing after they withdraw their consent. On this point, see Information Commissioner's Office (UK), 'How should we obtain, record and manage consent?' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/how-should-we-obtain-record-and-manage-consent/> accessed 4 February 2021.

21 Art 7(3) GDPR.

genomic data from the open platform—in other words, by making them no longer publicly available—any future use of such data based on Article 9(2)(e) should cease to take place, assuming users are put 'on notice' about this, in accordance with the lawfulness, fairness, and transparency principle under Article 5(1)(a). However, this is not clearly provided for by the GDPR.

Secondly, certain data subject rights may or may not be relevant depending on the chosen Article 9(2) exception. Closely related to the withdrawal of consent as discussed above is the right to erasure, or sometimes more commonly known as the right to be forgotten. Once the data subject has withdrawn their consent for the processing of data, Article 15(1) would be triggered, whereby the data subject may request 'the erasure of personal data concerning him or her without undue delay'.[22] In addition, where the personal data concerned are shared with others, the data controller should at the same time 'take reasonable steps, including technical measures, to inform controllers which are processing the personal data'. In the case of an open genomic data-sharing platform, this means that if the data subject withdraws their consent and asks their data to be removed, the platform must not only delete the data from their platform, but also endeavour to communicate that request to those who have downloaded the dataset as long as this is technically and financially feasible. It should, however, be noted that, for scientific research purposes, the data controller may refuse the right to erasure request if the right 'is likely to render impossible or seriously impair the achievement of the objectives of that processing'.[23]

In contrast, if the genomic data are being processed on the basis of Article 9(2)(e), it would be much more difficult for the data subject to establish and exercise the right to erasure, although not entirely impossible. Apart from withdrawal of consent, another ground on which the right to erasure can be exercised is the data subject's objection to the processing.[24] Resorting to the right to object, however, faces two major legal hurdles in the case of open genomic data sharing. First, the data subject's objection does not guarantee the erasure of the personal data concerned. Rather, the data controller would be required to perform a balancing test, and may refuse to stop processing or to delete the data if 'the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights, and freedoms of the data subject'.[25] This point has been made clear in the Court of Justice of the European Union (CJEU)'s *GC & others v CNIL and Google* decision in a search engine setting.[26] It is certainly arguable that where the data subject has taken down their genomic data from the platform, there would be no 'compelling legitimate grounds' for further processing, and ultimately this would be subject to judicial review in the case of legal dispute, but allowing the data controller to make a decision in the first instance makes it a cumbersome process for the data subject to exercise their right to object/erasure. Secondly, the general right to object[27] is applicable only when the legal basis for processing is Article 6(1)(e) (public interest) or 6(1)(f) (legitimate interests). To sum up, if manifestly public self-disclosure were chosen as the permitted exception for processing a data subject's genomic data, for that data subject to exercise their right to erasure, it would have to be established that either public interest or legitimate interests is relied upon at the same time as the Article 6 legal basis, and such an interest is overridden by the data subject's own interests in having the data erased.[28]

Moreover, restrictions that can be introduced by Member States may also vary depending on the chosen permitted exception for processing sensitive data. Under the Article 9(2)(a) explicit consent exception, national laws may 'provide that the prohibition [on processing sensitive data] may not be lifted by the data subject' through their explicit consent. This effectively empowers Member States to draw up a national 'blacklist' for certain types of processing performed against sensitive data. A comparable proviso does not exist under Article 9(2)(e), and under the CJEU's jurisprudence, this would mean Member States cannot introduce further restrictions for cases where the sensitive data are manifestly made public by the data subject.[29] It should be noted, however, that Article 9(4) provides that 'Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data

22 Art 15(1) GDPR.

23 Art 17(3)(d) GDPR.

24 Art 17(1)(c) GDPR.

25 Art 21(1) GDPR.

26 *GC and Others v Commission Nationale de l'Informatique et des Libertés (CNIL)*, Case C-136/17, [2019] (ECLI:EU:C:2019:773), paras 49–69 <http://curia.europa.eu/juris/liste.jsf?num=C-136/17> accessed 4 February 2021.

27 As opposed to the special right to object (art 21(2) and (3)), which applies to data processing for direct marketing purposes and prevents further processing in an absolute manner (ie no need for a balancing test).

28 We also note, in passing, that the data subject's right to data portability under art 20 may be affected if the processing is justified on the basis of art 9(2)(e), since this portability right is relevant only when consent or contract forms a legal basis for such processing and the processing is carried out by automated means.

29 See joined cases ASNEF and FECEMD, Case C-468/10 and Case C-469/10, [2011] OJ C 25/18 (ECLI:EU:C:2011:777), paras 30–32; see also Breyer, Case C-582/14, [2016] OJ C 475/3 (ECLI:EU:C:2016:779), para 57.

concerning health'. While Member States may invoke this provision to legislate restrictions on the use of such data, regardless of the chosen legitimizing ground under Article 9(2), it is unclear if this amounts to the power to prohibit such processing altogether, considering the different wording from Article 9(2)(a).

The question remains, though, whether explicit consent is valid if the purposes of subsequent, downstream data processing are (relatively) unclear. If DPAs were to contest the validity of a consent for 'research use' in general, we query how much more they should refuse to accept that explicit consent can be the basis for 'anyone can do anything', as would be the case with a truly open access platform, reflected in the language in PGP and openSNP discussed above. This would, of course, differ in the context of a controlled access or semi-open platform that restricts to some degree the purposes of data use (eg bona fide research) and those who can access the data (eg bona fide researchers); yet this kind of limitation would not seem to fall under manifest public self-disclosure, but rather consent to make data available for (more) defined purposes. Though even here, it is difficult to argue in favour of the validity of an explicit consent if research use is broader than a certain disease or, at most, something like 'health research'.[30]

Certain Member States have made use of Article 9(4) to prohibit certain types of processing of genetic data (Greece, eg prohibits the processing of genetic data for health and life insurance purposes[31]). However, while to our knowledge, no Member State has so far prohibited the publication of genetic data, some have put in place restrictions, in particular when the publishing means that data can be used subsequently outside the GDPR jurisdiction.[32] What is also relevant, though, is that some Member States require consent[33]—and consent is often seen as not valid if subsequent use and users cannot be defined precisely. As such, it might be seen as an indirect prohibition.

With all these different legal consequences resulting from the choice between Article 9(2)(a) and Article 9(2)(e), it can be concluded that recognizing explicit consent as the permitted exception would lead to a wider range of safeguards afforded to the data subject and as a result, a wider range of duties imposed on the data controller. This could be seen as creating a compliance incentive for downstream data users to opt for the data subject's manifest public self-disclosure as the permitted exception, or, if possible, scientific research under Article 9(2)(j).

In practice, then, this would mean that the operators of open genome platforms may feel more inclined to draft their terms and conditions in such a way that a data subject's uploading of their data amounts to making such data publicly available, or to design their platform in such a way that the involvement of the platform is kept to a minimum and their role is closer to a file sharing service. Third-party projects making use of the genomic data may also have a stronger tendency to choose a platform where data have been shared this way. Of course, there will be other considerations such as commercial reputation, ethical restrictions, data integrity, and confidentiality, and the impact on the recruitment of participants. Generally speaking, however, if one could broadly argue for using genomic data based on Article 9(2)(e), this could potentially undermine the interests of the data subject by removing the safeguards that would otherwise apply, were an alternative legal basis chosen. In this regard, from the viewpoint of protecting the data subject, it is important that the scope of Article 9(2)(e) be construed narrowly. The validity of relying on this exception depends on whether the arrangement genuinely reflects the technical configurations of the platform, as well as the actual legal relationship between the data subject and the controller.

A cautious approach to Article 9(2)(e), however, does not necessarily lead to disproportionate compliance burdens on downstream users. If they are bona fide researchers, they can rely on the legal basis for the primary collection of the genetic data—whether explicit consent or the research exemption—as long as the secondary use of such data is compatible with the original purpose (which is presumed to be the case under Article 5(1)(b)). They would

30 See eg Recital 33 GDPR, which states that '. . .data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognized ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose'.

31 Law 4624/2019 (Greece), art 23.

32 See eg Danish Data Protection Act (Law No 502 of 23/05/2018), art 10.

33 Spain, for example, in its Seventeenth Additional Provision of the Organic Law on Data Protection and Guarantee of Digital Rights, requires consent as an additional safeguard regardless of the actual exemption chosen to justify the use of sensitive data. See Iciar Alfonso Farnós, Guillermo Alcalde Bezhold, and Miriam Méndez García, 'Cuestionario/guía para la evaluación de proyectos de investigación con datos por un CEI / Questionnaire/guide for the evaluation of research projects with personal data by a Research Ethics Committee' (2019) Extra 1 Revista de derecho y genoma humano (Law and the Human Genome Review) 25; Asociación Nacional de Comités de Ética de la Investigación (ANCEI), 'Guías y Recomendaciones' (2020) <https://ancei.es/guias-y-recomendaciones/> accessed 4 February 2021. See also Jiahong Chen, Edward S Dove and Himani Bhakuni, 'Explicit Consent and Alternative Data Protection Processing Grounds for Health Research' in Eleni Kosta and Ronald Leenes (eds), *Research Handbook on EU Data Protection* (Edward Elgar forthcoming).

not have to ask the data subject for their explicit consent again,[34] but the safeguards associated with Articles 9(2)(a) and (j) would remain available to the data subject. We acknowledge that there will be challenges for downstream users to rely on explicit consent (eg the 'specific' requirement and the possibility of withdrawal of consent) or the research exemption (eg the need for national legislation and cross-border issues, as discussed below in the next section), but the data subject's public self-disclosure should be a last resort only when neither of those options proves possible and all conditions, set out later in this article, are satisfied. This would be a more balanced approach between protecting data subjects and promoting scientific research.

The differences mirrored in these options, again, highlight the importance of identifying the appropriate permitted exception under Article 9(2). It is therefore especially crucial for the operator of an open genome platform to explicitly specify the chosen legal basis for the primary collection of data. Uncertainty in such a choice by the platform (as in the case of PGP-UK and openSNP) does not help downstream users navigate such different compliance options. For the operator of a genome sharing platform and the downstream users alike, the next question would be under what conditions the uploading of a user's genetic data can be considered an act of making such data manifestly public. In the next section, we outline the scope of Article 9(2)(e) by reviewing guidelines issued by DPAs and the relevant case law.

The curious case of 'manifest' self-disclosure in data protection law

In practice, choosing the lawfulness ground and permitted exception for using sensitive data in genomic research is perhaps not as simple as it seems and is dependent on the function of the data-sharing platform, but a general discussion with the three platforms discussed above would nevertheless be helpful as a typical operational model. In this section, we first examine the general prohibition on processing sensitive data under Article 9 GDPR and the applicability of a few common exemptions, and then provide a more specific analysis of the different approaches to the scope of public self-disclosure, before advocating our case for a more restrictive interpretation of this provision.

Applicability and the exemptions of Article 9 GDPR

There should be little doubt that genomic data constitute one form of genetic data—in both scientific and legal terms—to the extent that the former is a genome-wide type of the latter.[35] Genetic data is defined by the GDPR as 'personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question',[36] including 'chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis, or from the analysis of another element enabling equivalent information to be obtained'.[37] For all three platforms mentioned above, the data uploaded by a user would fall squarely within this definition.

What is sometimes less clear, however, is whether such data would be covered by data protection law as personal data in the first place. While GenomeConnect claims that the shared data is 'de-identified', it is questionable whether the degree of de-identification is sufficient to pass the anonymity test under the GDPR, namely 'the data subject is not or no longer identifiable', taking into account all the means reasonably likely to be used to achieve identifiability, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly, and considering objective factors, such as the costs of and the amount of time required for identification, as well as the available technology at the time of the processing and technological developments.[38] Depending on the actual arrangements of a particular scheme, the operator of the sharing platform may or may not have kept a record of the participants' identities. The enrolment process of PGP-UK, for example, includes a stage of identity verification, which would likely allow the operating research team to attribute a set of genomic data to an identifiable individual. Other platforms may decide not to collect personal or demographic details about the volunteers and simply function like a file storage and sharing service. Either way, it is highly unlikely that the downstream users of such platforms would be given access to the personal details (eg names or other identifiers) of the data subject uploaders. Yet, the fact that

34 See Recital 50 GDPR.

35 Clare Bycroft and others, 'The UK Biobank Resource with Deep Phenotyping and Genomic Data' (2018) 562(7726) Nature 203–09; see also World Health Organization, 'WHO Definitions of Genetics and Genomics' <https://www.who.int/genomics/geneticsVSgenomics/en/> accessed 4 February 2021.

36 Art 4(13) GDPR.

37 Recital 34 GDPR.

38 Recital 26 GDPR.

the platform operators and/or the downstream data users do not have access to the identities of the data subject does not automatically render the genomic data in question 'anonymous information'.[39] Existing research has shown how re-association can be possible based on phenotypic[40] or otherwise available genetic information.[41] In a hypothetical scenario, police may compare a piece of genetic information of a suspect against genomic datasets, which could immediately make a matched record identifiable. As such, given the uniqueness of one's genomic sequence, we consider genomic data, even not linked to other forms of identifiers, personal and sensitive data under the GDPR.[42]

Once it is established that a set of data is brought within the scope of Article 9, as noted above, the processing of such data would be subject to a general prohibition unless one of the exemptions provided for by Article 9(2) applies. Among those ten listed exemptions, several stand out as potentially relevant to the processing of genomic data: explicit consent (point (a)); manifest public availability (point (e)); substantial public interest (point (g)); health or social care (point (h)); public health (point (i)); and archiving, scientific research or statistics (point (j)). The latter two provisions are relevant where the data are used by researchers, and some of these provisions may come across as particularly pertinent in a time of public health crisis, such as points (g) and (i).[43] The rest of this article will, however, only focus on explicit consent and manifest public availability for a number of reasons. First, some of the exemptions are clearly applicable only to emergency-type scenarios (eg vital interests of the data subject), professional care services (eg health or social care), or situations in which processing is necessary for reasons of 'substantial public interest', including those for public health, and not to the day-to-day operation of an open genome platform, either for making the data available in the first instance or for using the data for subsequent purposes. Second, explicit consent and public self-disclosure are the only two exemptions requiring the data subject's voluntariness in the first place, as opposed to the mandatory nature of the other exemptions. Third, and as a related point, all other grounds mentioned above—including the seemingly promising scientific research provision under Article 9(2)(j)—explicitly require Member State laws to provide a legal basis (which, to the disappointment of many scientists, makes cross-border processing of genomic and health-related data exceedingly difficult in Europe). Fourth, and consequently, relying on these other exemptions may potentially subject an open genome initiative—especially as an international collaboration—to the uncertainty of applicable data protection law and the fragmentation of national rules across the EU.[44] For these reasons, Article 9(2)(a) explicit consent (drafted somewhat broadly to cover future, downstream purposes) and Article 9(2)(e) manifest public self-disclosure seem to be the most practical bases—both for the establishment and operation of an international open genomic data platform as well as the subsequent, downstream uses of such data for research. This said, we recognize that genomic projects keeping data within individual Member States (ie not making the data available across jurisdictions) may find that the scientific research provision is the most suitable provision.

What remains uncertain is when a researcher, for example, intends to gain access to the data via one of the platforms mentioned above, can they lawfully do so on the basis that the volunteer has given their explicit consent? Or that such data have been made manifestly available to the public by the volunteer? While there has been a significant amount of theoretical and practical discussions surrounding explicit consent as a permitted exception, the alternative option of manifest public availability remains largely unexplored. The next two sub-sections will thus focus on the scope of Article 9(2)(e).

39 Yaniv Erlich and others, 'Redefining Genomic Privacy: Trust and Empowerment' (2014) 12(11) PLoS Biology e1001983; Mahsa Shabani and Luca Marelli, 'Re-identifiability of Genomic Data and the GDPR' (2019) 20(6) EMBO Reports e48316.

40 Christoph Lippert and others, 'Identification of Individuals by Trait Prediction using Whole-Genome Sequencing Data' (2017) 114(38) PNAS 10166–71.

41 Yaniv Erlich and others, 'Identity Inference of Genomic Data Using Long-Range Familial Searches' (2018) 362(6415) Science 690–94. See also Xinghua Shi and Xintao Wu, 'An Overview of Human Genetic Privacy' (2017) 1387(1) Annals of the New York Academy of Sciences 61–72.

42 See also Murat Sariyar, Stephanie Suhr, and Irene Schlünder, 'How Sensitive Is Genetic Data?' (2017) 15(6) Biopreservation and Biobanking 494–501.

43 See European Data Protection Board, *Guidelines 03/2020 on the Processing of Data Concerning Health for the Purpose of Scientific Research in the Context of the COVID-19 Outbreak* (2020), paras 25–27; European Data Protection Board, *Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR)* (2019) paras 10–14; European Data Protection Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research* (EDPS 2020), s 6.8.

44 We note, however, that fragmentation across the EU may exist all the same in this particular context, as art 9(4) allows Member States to maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health. See also David Townend, 'Conclusion: Harmonisation in Genomic and Health Data Sharing for Research: An Impossible Dream?' (2018) 137(8) Human Genetics 657–64.

Contradictory interpretations of Article 9(2)(e)

Article 9(2)(e) was perhaps one of the least discussed provisions in the course of legislating the GDPR. Since the introduction to the Commission's first legislative proposal in 2012,[45] this clause has been retained in all major subsequent versions[46] without any question or debate. The phrase 'processing relates to personal data which are manifestly made public by the data subject' is a word-by-word repetition of the first half of Article 8(2)(e) DPD, which first appeared in the Council's Common Position on the draft DPD in 1995.[47] However, the Statement of Reasons by the Council shed no light on the rationale behind this addition, nor was it challenged in later stages by the Parliament[48] or the Commission.[49] The lack of clarification in the recitals of the DPD and the GDPR, exacerbated by the lack of documentation of the legislative deliberation, makes it difficult to decipher the lawmakers' considerations when formulating this exemption. Straightforward as this provision seems at first glance, it may, as discussed below, be subject to different interpretations in practice, not least in the case of open genomic data.

Perhaps because of the lack of any legislative guidance, be it during the debate stages or in the recitals or legislative text itself, DPAs have adopted a restrictive interpretation of both elements of the self-disclosure exemption, i.e. 'manifestly made public' and 'by the data subject'. Regarding the first element, the UK's Information Commissioner's Office (ICO), for example, is of the view that the 'manifestly made public' condition amounts to a test of accessibility by anyone. As explained in their guidance:

> To be manifestly made public, the data must also be realistically accessible to a member of the general public. The question is not whether it is theoretically in the public domain (eg in a publication in a specialist library, or mentioned in court), but whether it is *actually publicly available in practice*. Disclosures to a limited audience are not necessarily 'manifestly public' for these purposes. In particular, information is not necessarily public just because you have access to it. The question is whether any hypothetical interested member of the public could access this information.[50]

As regards the second element, the 'by the data subject' condition, the ICO has emphasized the importance of the awareness and voluntariness by the data subject, something that we saw in at least two of the platforms discussed above:

> You need to be confident that it was the individual themselves who actively chose to make their special category data public and that this was unmistakably a deliberate act on their part. There is a difference between assenting to or being aware of publication, and an individual actively making information available. For example, by blogging about their health condition or political views. You might also find it hard to show that someone has manifestly made information public if, for example, they made a social media post for family and friends but default audience settings made this public. You should therefore be very cautious about using this condition to justify your use of special category data obtained from social media posts.[51]

In other words, the data subject's misunderstanding of who would have actual access to their data may render the disclosure involuntary, and as a result, the processing invalid altogether. Given that a data controller may have no ability to determine whether the data subject had a misunderstanding of 'default audience settings' and who could access their data, this would seem to set

45 Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (2012) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2012:0011:FIN> accessed 4 February 2021.

46 European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 — C7-0025/2012 — 2012/0011(COD)) (Ordinary legislative procedure: first reading) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52014AP0212> accessed 4 February 2021. See also Position (EU) No 6/2016 of the Council at first reading with a view to the adoption of a Regulation of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) Adopted by the Council on 8 April 2016, OJ 2016 C 159/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52016AG0006(01)> accessed 4 February 2021.

47 Common Position (EC) No 1/95 adopted by the Council on 20 February 1995 with a view to adopting Directive 95/. . ./EC of the European Parliament and of the Council of . . . on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ 1995 C 93/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:51995AG0413(01)&from=EN> accessed 4 February 2021.

48 Minutes of Proceedings of the Sitting on Thursday, 15 June 1995 (95/C 166/04) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:JOC_1995_166_R_0077_01&from=EN> accessed 4 February 2021.

49 Opinion of the Commission pursuant to art 189 b (2) (d) of the EC Treaty, on the European Parliament's amendments to the Council's common position regarding the proposal for a European Parliament and Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data (COM(95) 375 final-COD287) (1995) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:51995DC0375&from=EN> accessed 4 February 2021.

50 Information Commissioner's Office (UK), 'What are the Conditions for Processing?' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/what-are-the-conditions-for-processing/> (emphasis added) accessed 4 February 2021.

51 Ibid.

a high bar for Article 9(2)(e) to be met. The need for confidence in an 'unmistakably deliberate act' by the data subject themselves would permit the use of only a narrow range of personal data self-disclosures.

This view is shared by data protection legal scholars Ludmila Georgieva and Christopher Kuner, who write:

> Sensitive data may be processed when the data subject has manifestly made them public. In this context, 'making public' should be construed to include publishing the data in the mass media, putting them on online social network platforms or similar actions. However, the data must have been 'manifestly' made public, which requires an affirmative act by the data subject, and that he or she realised that this would be the result. The EDPB has stated that 'data controllers processing those data in the context of video surveillance cannot rely on Article 9(2)(e), which allows processing that relates to personal data that are manifestly made public by the data subject. The mere fact of entering into the range of the camera does not imply that the data subject intends to make public special categories of data relating to him or her'. Data processing will not fall within this exception if the data have been made public illegally.[52]

While Georgieva and Kuner do not go so far as to cast doubt on social media posts per se, unlike the ICO's guidance, they do also suggest that controllers must have confidence that the data subject themselves made an affirmative act to disclose their data, and that they were aware this disclosure would make the data public (ie neither private nor semi-private).

DPAs at EU level also seem to share a similar view with these statements. The *Handbook on European Data Protection Law*, which is published jointly by the Council of Europe and the European Union, provides high-level guidance on this exception:

> Article 9(2)(e) of the GDPR provides that processing is not prohibited if it relates to data which are manifestly made public by the data subject. Even though the meaning of 'manifestly made public by the data subject' is not defined in the regulation, since it is an exception to prohibiting sensitive data processing, *it must be construed strictly and as requiring the data subject to deliberately make his or her personal data public*. Thus, where the television broadcasts a video taken from a video surveillance camera, showing, among other things, a firefighter getting injured trying to evacuate a building, it cannot be considered that the firefighter has manifestly made public the data. On the other hand, if the firefighter decides to describe the incident and publish the video and photos on a public internet page, he or she would have made a *deliberate, affirmative act to make the personal data public*. It is important to note that *making one's data public does not constitute consent, but it is another permission for processing special categories of data*.
> The fact that the data subject had made public the processed personal data does not exempt controllers from their obligations under data protection law. For instance, the principle of purpose limitation continues to apply to personal data even if such data have been made publicly available.[53]

The Article 29 Working Party (now the EDPB) has conducted an analysis of the equivalent provision in the Law Enforcement Directive (LE Directive)[54] concerning self-disclosure. While the GDPR and the LE Directive are mutually exclusive in terms of their scope, the two instruments share the fundamental objective of protecting personal data, and almost the same wording can be found in Article 9(2)(e) GDPR and Article 10(c) of the LE Directive. Hence, the examination of the latter should be highly relevant, if not fully applicable, to the interpretation of the former. In fact, the current European Data Protection Supervisor has cited the Working Party's analysis of the LE Directive to support his interpretation of Article 9(2)(e) GDPR in the context of scientific research—a matter governed by the GDPR:

> Special categories of data may be processed if the data subject has manifestly made them public. EU data protection authorities have argued that this provision has to be 'interpreted to imply that the data subject was aware that the respective data will be publicly available which means to everyone' including, in this case, researchers, and that, 'In case of doubt, a narrow interpretation should be applied, as the assumption is that the data subject has voluntarily given up the special protection for sensitive data by making them available to the public including authorities'. Publishing personal data in a biography or an article in the press is not the same as posting a message on a social media page.[55]

In explaining when competent law enforcement authorities may use sensitive data based on public self-disclosure by the data subject, the Working Party has opined that 'this has to be interpreted to imply that the data subject was aware that the respective data will be

52 Georgieva and Kuner (n 7) 378.

53 European Union Agency for Fundamental Rights, European Court of Human Rights, European Data Protection Supervisor, *Handbook on European Data Protection Law* (Publications Office of the European Union 2018), 162–63 (emphasis added).

54 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ 2016 L 119/89.

55 European Data Protection Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research* (EDPS, 2020), s 6.3 (p 19).

publicly available which means to everyone including authorities'.[56] Further examples are provided:

> In cases such as the publishing of personal data in a biography, in the press or on a public website the intention is clear. In other cases, this is more difficult to decide. For example, registering for a social network might include the acceptance of certain data protection rules which provide that all the partners of the provider (including national police authorities) have access to personal data. In such cases most of the users probably do not actively take notice of this and are in fact not aware that their data are available to police authorities.[57]

In practice, however, courts seem to have embraced broader interpretations than those of DPAs and legal commentators, thereby casting a wider scope. For the 'manifestly made public' condition, while the Court of Justice of the European Union (CJEU) has not yet had an opportunity to expound on that requirement under the GDPR, there is case law at EU level dealing with the same issue under Regulation 45/2001[58] (now replaced by Regulation 2018/1725[59])—which governs the use of personal data by EU institutions. As with the LE Directive, Regulation 45/2001 (as well as its successor) contains the same self-disclosure exemption, with identical wording, to the general prohibition on processing of sensitive data.[60]

In *Esch-Leonhardt and Others v ECB*,[61] the CJEU's Court of First Instance (now the General Court) had to decide whether the applicants could request the removal of an email from their personnel profiles on the basis that it contained sensitive data. The applicants were employees of the European Central Bank (ECB) and were union members of the International and European Public Services Organization (IPSO). They made requests that the ECB remove from their personnel profiles an email sent by them to all staff and signed on behalf of IPSO, on the grounds that the email contained sensitive data about their union membership and there was no justification to store such data in their profiles. The Court rejected their claim and reasoned that, even if the reference to IPSO revealed the applicants' union membership, the processing of such data was justified by the fact that the applicants made this information manifestly public *within the ECB* by signing off their email with the IPSO signature line.[62] Clearly, the Court did not take the view that the 'manifestly made public' condition equals 'accessibility by anyone' or universally public; rather, the criterion of 'public' here, following the Court's logic, should be appropriate to the context in which the data are processed. This could mean 'public' within a defined community setting, such as an organization or network. In other words, the court's broader notion of 'public' renders it possible to argue that making data public in a confined space, and keeping them within that confined space, provides sufficient legitimation to process the data.

As for the 'by the data subject' condition, the voluntariness test has also been interpreted in a broader manner by the courts. In *NT1 & NT2 v Google LLC*,[63] for example, the High Court of Justice for England and Wales considered whether the claimants could exercise their 'right to be forgotten' against Google in relation to de-referencing the search results about their past criminal convictions. When assessing whether Google could process such sensitive data on the basis that the data had been made manifestly public by the data subject, the Court analysed the wording of Article 8(2)(e) DPD as well as Schedule 3 of the Data Protection Act 1998 (as the UK's transposition of the DPD), and concluded that:

> It does not require a deliberate decision or 'step' by the data subject 'to make' the information public, but rather (a) the taking by him of a deliberate step or steps, as a result of which (b) the information is 'made public'. A person who deliberately conducts himself in a criminal fashion runs the risk of apprehension, prosecution, trial, conviction, and sentence. Publicity for what happens at a trial is the ordinary consequence of the open justice principle [...] The same must be true of the details of the offending, and other information disclosed in open court, including information about himself which a criminal reveals at a trial or in the course of an application. [...] I attach weight to the fact that a narrow interpretation would tend towards the unacceptable conclusion discussed above. I see a good deal of force in the point made by [Google] that if this condition were not available in respect of processing of this nature it is hard to see how ordinary members of the public could

56 Article 29 Data Protection Working Party, *Opinion on Some Key Issues of the Law Enforcement Directive* (EU 2016/680) at 10 <http://ec.europa.eu/newsroom/just/document.cfm?doc_id=48804> accessed 4 February 2021.

57 Ibid.

58 Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data, OJ 2001 L 8/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32001R0045> accessed 4 February 2021.

59 Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ 2018 L 295/39 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R1725> accessed 4 February 2021.

60 Art 10(2)(d) reg 45/2001; art 10(2)(e) reg 2018/1725.

61 Case T-320/02 (ECLI:EU:T:2004:45).

62 Ibid para 57.

63 [2018] EWHC 799 (QB).

lawfully discuss online the convictions (whether recent or historic) of those appearing before the courts.[64]

According to the High Court, the causal connection between the data subject's action and the consequence that the information has been made public does not have to be as strong as the ICO has suggested. This of course does not mean that *any* publicly available sensitive data can be processed based on the Article 9(2)(e) self-disclosure exemption, not least because some of them might have been made public unlawfully.[65] Equally though, under the Court's interpretation, the disclosure does not have to be an action directly triggering the dissemination of the sensitive data.

The courts' relatively broad interpretation of these two conditions presents a stark contrast to the clearly more cautious approach taken by the regulators. It is perhaps conceivable why, from the regulators' point of view, limiting the scope of Article 9(2)(e) is desirable for maintaining a general balance between the data subject and the data controller: if 'manifestly made public' can be stretched to mean a group of people, this may lead to an arbitrary definition of 'public', even in cases such as private group messages; if 'by the data subject' were also to cover unintentional acts, data subjects may find themselves deprived of the protection when the sensitive data are, for example, leaked to the public by accident. The courts, however, are understandably more mindful of the specifics of the cases at hand, focusing on whether the data subject can fairly claim that the information remains in the private domain.

It should be pointed out that the 'manifestly made public' and 'by the data subject' elements are not necessarily met or unmet at the same time. It can be argued, for example, that the data concerned in *ECB* are indeed made available 'by the data subject' with a voluntary intention, but only to a specific group of people rather than virtually anyone, and thus fail the ICO's 'manifestly made public' criterion, but not the Court's. Likewise, fulfilling the 'manifestly made public' condition does not guarantee the fulfilment of the 'by the data subject' condition. We expound on this point further in the next section.

A strict interpretation for genomic research?

These somewhat differing interpretive approaches to the meaning of 'manifestly made public by the data subject' in data protection law signify the complexity of balancing sometimes conflicting interests in the use of sensitive data, which is not helped by the lack of legislative guidance. Specifically, in the case of using genomic data submitted by the data subject to an open platform, the interests involved are both the informational self-determination of the data subject and the (putative) societal benefits of scientific research activities, though of course, given the open nature of such a platform, the purposes may not be scientific research, much less something of societal benefit. Moreover, these values do not necessarily have to come into conflict and in fact, they are aligned on a fundamental level: one of the foundational assumptions underlying data protection and confidentiality law is that the value of information sharing can be optimized only when people voluntarily disclose such information based on the trust that the information will be protected.[66] Likewise, open genome initiatives would not achieve their objective if donation of genetic data is deterred by the lack of protection and control given to the donating volunteers.

The challenging part is where exactly the line should be drawn so these private and public interests can be best coordinated. With regard to what should be the appropriate permitted exception under Article 9(2) for using genomic data contributed by the data subject themselves, the differences between explicit consent and manifest self-disclosure can be subtle but not insignificant. As explained in detail in the previous section, relying on explicit consent would give the data subject a more straightforward corollary right to withdraw their consent and to subsequently request the erasure of data, but subject to the condition that this would not render the objective of research unachievable. This said, relying on manifestly public self-disclosure would still allow the data subject to object to the processing of their data, but whether that would lead to the erasure of data depends on a balancing assessment carried out by the data controller. As such, neither option would guarantee exclusive control by one side and leave the other side with no control at all. Yet, when it comes to who has greater control in the two scenarios, explicit consent tilts more toward the data subject, whereas manifestly public self-disclosure tilts more toward the data controller. The different legal consequences essentially mirror various degrees to which the data subject is considered to have given up some control of their data under different circumstances.

To the extent that the selection of the legal basis under Article 6(1) and permitted exception under Article 9(2) may determine the range of legal safeguards

64 Ibid paras 111–13.

65 See Georgieva and Kuner (n 7) 378.

66 For a classic theory of the value of privacy, see Richard Posner, 'The Economics of Privacy' (1981) 71(2) The American Economic Review 405–09. For a more up-to-date overview of different theories, see Alessandro Acquisti, Curtis Taylor and Liad Wagman, 'The Economics of Privacy' (2016) 54(2) Journal of Economic Literature 442–92.

available to the data subject, this should not always be a matter of unilateral choice by the data controller—and sometimes not even a bilateral choice by both the data controller and the data subject, when such a choice proves manifestly unfair. This is perhaps something better discussed in the literature regarding the relationship between consent and contract as two separate legal grounds for processing of non-sensitive data.[67] The key question here concerns whether, for example, an intermediary such as a website may include the permission of using personal data for advertising purposes as an inseparable part of the contractual obligations on the user. The Working Party has discussed this issue in various Opinions under both the DPD and the GDPR,[68] and they consistently maintain that:

> The provision [Article 7(b), performance of contract as a legitimising basis] must be interpreted strictly and does not cover situations where the processing is not genuinely *necessary* for the performance of a contract, but rather unilaterally imposed on the data subject by the controller. Also the fact that some data processing is covered by a contract does not automatically mean that the processing is necessary for its performance. For example, Article 7(b) is not a suitable legal ground for building a profile of the user's tastes and lifestyle choices based on his clickstream on a website and the items purchased. This is because the data controller has not been contracted to carry out profiling, but rather to deliver particular goods and services, for example. Even if these processing activities are specifically mentioned in the small print of the contract, this fact alone does not make them 'necessary' for the performance of the contract.[69]

It is therefore clear that a data controller cannot force the legal basis for their use of personal data in a merely formulaic manner by explicitly stating their choice in an agreement. They are sometimes bound by the actual circumstances that define the nature of their relationship with the data subject. By the same token, just because an open genome sharing platform has made it clear to the data subjects-participants that 'the uploaded data will be accessible by anyone', it does not mean that the platform or data users can *always* lawfully rely on the exception of manifestly public self-disclosure by the data subject. Instead, and as we elaborate later in the next section below, one must examine whether the data subject's permission is meant for a specific group of persons to process the data for a specific purpose (in which case explicit consent or scientific research should be the appropriate exception) or simply to make certain information known to unspecified persons (in which case manifestly public self-disclosure should be the appropriate exception). In short, 'manifestly made public' is not related to a specific use only. It can include specific use and it can be a driving force, but it would still have to be communicated that the use is general.

A legal test for Article 9(2)(e)

In this section, we weave the preceding analysis together to propose a legal test for controllers to ascertain whether they can rely upon Article 9(2)(e) as a permitted exception to process special category personal data. Our specific focus remains on genomic data in the open genomic data-sharing platform context, but we intend the test to be universally applicable across all special category data contexts.

Our test breaks down the provision 'relates to personal data which are manifestly made public by the data subject' into its constituent elements. The constituent elements are threefold: (i) processing [that] relates to personal data; (ii) personal data which are manifestly made public; and (iii) [personal data] manifestly made public by the data subject. All three elements, we argue, are grounded in the intent and reasonable expectations of the data subject. Unlike other exceptions under Article 9(2), this exception rests on a voluntary, deliberate action undertaken by the data subject. For that reason alone, a narrow interpretation is warranted, arguably even more so than the other listed exceptions under Article 9(2).

The first arm of the test must consider whether the processing 'relates to personal data'. This arm of the test has a relatively low threshold to meet. In our view, this means that the data processing activity must have a direct connection with the special category of personal data under consideration. Thus, the first arm would be met where a controller wished to process any special category of personal data concerning an individual, and that personal data emanates from the data subject themself.

The second arm of the test must consider 'personal data which are manifestly made public'. Here, focusing on the word 'manifestly', we argue that there must be clear evidence of a deliberate, affirmative act by the data subject themself to make their data available, such as

67 Frederik Zuiderveen Borgesius, 'Personal Data Processing for Behavioural Targeting: Which Legal Basis?' (2015) 5(3) International Data Privacy Law 163–76; Philipp Hacker, 'Personal Data, Exploitative Contracts, and Algorithmic Fairness: Autonomous Vehicles Meet the Internet of Things' (2017) 7(4) International Data Privacy Law 266–86.

68 Article 29 Data Protection Working Party, *Opinion 15/2011 on the Definition of Consent* (01197/11/EN, WP187) 7–8; Article 29 Data Protection Working Party, *Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC* (844/14/EN, WP 217) 16–17; Article 29 Data Protection Working Party, *Guidelines on Consent under Regulation 2016/679* (17/EN, WP259 rev.01) 8–10.

69 Article 29 Data Protection Working Party, *Opinion 06/2014* (n 68) 16–17 (emphasis in original).

inclusion of their signature or a social media post from their account or, in our genetic context of an open access platform, they themselves voluntarily and knowingly uploading a file of their genetic data (or other special category data). This focuses on the intent of the data subject and requires evidence of an intentional, voluntary, positive act made by them to manifestly make their personal data public. Explicit manifestation of the intention predominates in this assessment. We therefore reject a standard of implied intention, or what might be called the 'reasonable data subject' standard, ie that it is acceptable for the controller to establish that, in the circumstances of the particular case, a reasonable person in the data subject's position would be likely to consider their personal data as having been manifestly made public. We ascribe to the narrow interpretation of this permitted exception; adopting an explicit, reasonable data subject standard opens the door too widely for uses of sensitive data that the data subject would not necessarily be aware of, and had never considered they had made 'public' in the first place. It should, however, be clarified that this criterion does not amount to an 'absolute' standard that inquires only about the true intention of the data subject, regardless of any evidence. What is required is objective evidence (eg a record of signature) of the explicit subjective intention (e.g. a statement of making the uploaded file accessible by anyone)—an intention that would also pass the 'public' test, as explained below. The intentional element we advance here aligns more closely with the DPA's narrow interpretation (although not identically), and points towards ways in which this can be evidenced in practice.

This second arm of the test also must consider the meaning of the word 'public'. As with the UK's ICO, we interpret this phrase literally and as meaning actually publicly available in practice. In this sense, then, we agree with the ICO's interpretation that this part of the test must consider 'whether any hypothetical interested member of the public could access this information'. Applying that test to the facts of *Esch-Leonhardt and Others v ECB* would result in a very different outcome: the Court would have had to rule that the emails were not manifestly public because not any hypothetical interested member of the public could access their emails—only those receiving them (or otherwise having access to them) could. In our view, 'public' must mean available to everyone. Similar to the language in Recital 26 of the GDPR, for everyone to 'hypothetically have access' to the personal data, account should be taken of all the means reasonably likely to be used to access the data, and to ascertain this, account should be taken of all objective factors, such as the costs of and the amount of time required for accessing the data, taking into consideration the available technology at the time of the processing and technological developments. In other words, in addition to what has been set out in the DPA's guidance, if a disproportionate, resource-intensive amount of effort is needed to access the data, it is less likely to be considered 'public'.

Finally, the third arm of the test must consider 'by the data subject'. Here, we also adopt a literal interpretation of the phrase: it must be the data subject who makes their personal data manifestly public, or at least initiates the action to make their data public. Applying this element of test to the question we posed in the introduction about the role of intermediaries in the process of making an individual's data manifestly public, there would need to be a clear indication made by the data subject that they were relying upon the intermediary to make their data public. For example, in the social media setting, this would be clear reliance on an ISP and Twitter, by logging into one's Twitter account (with one's personally identifying information available to everyone) and then tweeting a post containing sensitive information. Here, the act of tweeting alone would signal clear reliance on the intermediary. In the open genomic data platform scenario, there would likely need to be a stronger signal of reliance on the platform given that the data are much more extensive and 'raw' than a short-form tweet and less common than something like a blog post or Facebook update. Here, it may be advisable for open genomic platforms to require data subjects to explicitly acknowledge that they are relying on them to make their genetic data available on their platform. The role of an intermediary and the data subject's reliance on such a tool are not fully explored in the current regulatory guidance, something we think worthy of further discussion.

To summarize, our three-arm legal test is as follows: for special category personal data to be lawfully processed under GDPR Article 9(2)(e), it must be established by the person seeking to process the data that:

1. The processing activity is directly connected to those personal data that have been manifestly made public by the data subject;
2. There is evidence of a deliberate, affirmative act by the data subject themself to make their data available, *and* the data are public such that any hypothetical interested member of the public could access them; and
3. The data have been made manifestly public by the data subject themself (ie directly made public by them), or the data subject has given a clear indication to an intermediary to make their data public (ie indirectly made public by them).

The foregoing analysis leads us to conclude that while open genomic platforms can in principle rely on Article 9(2)(e) as a permitted exception, it will need to be carefully circumscribed. The threshold to meet each constituent element of the test is high in comparison to explicit consent under Article 9(2)(a) or scientific research under Article 9(2)(j); this is partly because there are no protective mechanisms found elsewhere in the GDPR for this provision, as there are for consent under Article 7 and for scientific research under Article 89. It has been noted above that DPAs are wary of interpreting this provision broadly, precisely because there is a built-in assumption that the data subject has voluntarily 'given up' the special protection for sensitive data by making them available to the public. Residual safeguards will still be found in parts of the GDPR to protect the data and afford the data subject some rights, but these safeguards are not nearly as robust as they are for other permitted exceptions such as explicit consent and scientific research.

As such, it is insufficient for a controller merely to inform data subjects that their permitted exception for processing special category personal data is one of those listed under Article 9(2). There needs to be an active fulfilment of the criteria, no matter the permitted exception invoked. In the case of Article 9(2)(e), the active intent to manifestly make the data public must come from the data subject rather than the controller. In other words, the data subject would need to initiate the activity of making their genetic data manifestly public by approaching the platform, and then uploading or otherwise performing the activity of making the data public. And, there would need to be a clear intent that the data subject is aware that doing so would make their data accessible to anyone. To ensure this, a controller might consider advising the data subject to actively signal their acknowledgement of a statement along the lines of: 'By uploading your genetic data onto our data-sharing platform, you acknowledge that you are making your data manifestly public—that is, openly available for anyone to access—and that it may be downloaded, shared, and used by us and other persons, for research and non-research purposes alike'.

Setting out these strict tests for invoking Article 9(2)(e) means that the potential compliance incentives as highlighted above would be largely neutralized and thus much less exploitable as a loophole for evasion of controller duties. Given this, it seems likely that open genomic data platforms and downstream users will be more inclined to rely on a lawful basis under Article 6(1) that is consent, public interest, or legitimate interests, and a permitted exception under Article 9(2) that is explicit consent or scientific research, even if it affords them less flexibility—though, as our earlier examples of PGP and openSNP suggest, there is scope for 9(2)(e) to be invoked—if designed and implemented appropriately. Between the two exceptions, undoubtedly explicit consent affords more control to data subjects, while scientific research affords more control to researcher-controllers. For those controllers relying on the scientific research provision under Article 9(2)(j), this would also oblige them to have more control over the data (eg technical and organizational safeguards, potential access restrictions) and afford less rights to data subjects under Articles 12–23. However, while our analysis leads us to conclude that the exception under Article 9(2) may often be one *other than* manifestly public self-disclosure, this does not automatically mean researchers (as distinct from other kinds of downstream users) intending to use the data already available on the platform *must always* ask the data subject for consent each time they use the data. This is because research purposes are presumed to be 'compatible purposes' under Article 5(1)(b), meaning that researchers may in most cases use the data based on the same original legal basis at the point of initial collection—whether it is consent, public interest, or legitimate interests.[70] This interpretation should allow researchers to make use of uploaded genomic data without having to repeat the process of obtaining the data subject's consent again (assuming the legal basis was consent and the original consent was drafted appropriately to account for a specific, defined, and acceptable range of research purposes), but remain under the data protection duties associated with consent (notably the right to erasure). We suggest this approach, driven by consent at Articles 6(1)(a) and 9(2)(a), will better allow individuals to maintain ultimate control of their sensitive data. After all, once sensitive data are manifestly made public by the data subject—put out into the open for all to access and use—it is difficult if not impossible to undo it, much less control downstream uses. Phrased another way, 'data subject beware' must be the operating ethos for Article 9(2)(e).

Conclusion

In this article, we charted the contours of a hitherto little-discussed provision in European data protection law: the provision that allows processing of special category personal data on the basis that they have been made manifestly public by the data subject. This

70 Recital 50 GDPR.

provision has appeared in law since the 1995 DPD, and currently is found in GDPR Article 9(2)(e). We were curious as to why such little guidance has been provided on this provision over the years, given its potential to be of significant value to research organizations that want to process health and genetic data without having to rely on other provisions that might be considered more onerous or problematic, such as explicit consent. At the same time, we were concerned that lack of guidance could lead to misuse or abuse of the provision. We therefore analysed the limited guidance available from DPAs and the case law, and considered its application in the context of genomic data and open data sharing. We proposed a three-arm legal test to determine when Article 9(2)(e) can be lawfully invoked. Our analysis led us to conclude that the exception must be construed narrowly, even by the general standards of Article 9(2). Unless each element of the legal test is met, it would be neither legally nor ethically appropriate to invoke this exception. This does not mean that Article 9(2)(e) can never be successfully invoked. Indeed, it may be available in research and non-research contexts alike, but it does mean that the invocation must be initiated and led throughout the process by the data subject, rather than the controller or processor. Legal certainty in this respect will only become more important as the EU is planning to establish a Common European Health Data Space[71] and to promote data altruism[72] as part of the European Strategy for Data. Applying the right set of rules to genome sharing platforms will be a key step towards building trust in encouraging data donation for the common good.

Acknowledgements

The authors would like to thank Regina Becker and Mark Taylor for their comments on a previous draft. Jiahong Chen's work was supported by the Engineering and Physical Sciences Research Council (grant numbers EP/M02315X/1, EP/T022493/1).

doi:10.1093/idpl/ipab005

71 European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European Strategy for Data (COM(2020) 66 final), 29–30 <https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf> accessed 4 February 2021.

72 Ibid 13. See also European Commission, 'Proposal for a Regulation on European data governance (Data Governance Act)' <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act> accessed 4 February 2021.