1. Introduction

In May 2018, the General Data Protection Regulation (GDPR) \(^1\) will come into effect, establishing a new set of rules for data protection in the European Union. The GDPR will replace the 1995 Data Protection Directive, building upon the key elements and principles of the Directive while adding new rights for individuals and new obligations on data controllers and data processors.

Compared to the 1995 Directive, the GDPR addresses de-identification in a more nuanced way. While the GDPR maintains the same high standard for achieving anonymization, it recognizes the existence of different levels of de-identification and it explicitly adds references to an intermediate form of de-identification – namely, pseudonymization.

In this paper, we use the terms “de-identification,” “pseudonymization,” and “anonymization.” For the purposes of this discussion, we use “de-identification” as a general term that includes the full spectrum of methods, from simple pseudonymization to full anonymization.

“Pseudonymization” commonly refers to a de-identification method that removes or replaces direct identifiers (for example, names, phone numbers, government-issued ID numbers, etc.) from a data set, but may leave in place data that could indirectly identify a person (often referred to as quasi-identifiers or indirect identifiers)\(^2\). Applying such a method, and nothing else, might be called “simple pseudonymization”. Frequently, security and privacy controls designed to prevent the unauthorized re-identification of data are applied on top of simple pseudonymization to create strong pseudonymization.

Pseudonymization is defined in the GDPR as:

> the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.\(^3\)

For the purposes of this paper, when we refer to pseudonymization, we assume some form of strong pseudonymization that would meet the GDPR definition and include the controls or “technical and organizational measures” referred to above.

Pseudonymized data remains “personal data” and is therefore subject to the requirements of the GDPR. But the GDPR provides some regulatory incentives to adopt pseudonymization, and there are therefore some significant benefits to employing it. Specifically, pseudonymizing data can help an organization meet some of the GDPR requirements, but it does not fully release the organization from them.

By contrast, “anonymization” as used in this paper refers to an even stronger form of de-identification. For the purposes of this paper, we will assume that strong anonymization methods are being used, and that these methods would be considered acceptable by European Data Protection Authorities (DPAs). In Appendix A we provide more details about what such methods entail. Fully anonymized data that meets the legal bar set by European data protection law is no longer “personal data” and is therefore not subject to the obligations of the GDPR at all. Thus, the benefits of pseudonymization pale in comparison to the benefits of full anonymization.


\(^2\) Another term that is sometimes used to mean the same thing as pseudonymization is “tokenization”.

\(^3\) GDPR Article 4(5).
To illustrate this point, we will look at 11 key GDPR obligations to see how pseudonymization and anonymization affects their applicability. The next section begins with a summary table listing these obligations, followed by a discussion of each. More details on some of these obligations are included in the Appendices.

2. Different Obligations under the GDPR

The following table summarizes the benefits of each type of de-identification and whether each obligation applies:

<table>
<thead>
<tr>
<th>GDPR Obligation</th>
<th>Identified</th>
<th>Pseudonymized</th>
<th>Anonymized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Provide notice to data subject</td>
<td>Required</td>
<td>Required</td>
<td>Not Required</td>
</tr>
<tr>
<td>2. Obtain consent or have another legal basis</td>
<td>Required</td>
<td>Potentially helps</td>
<td>Not Required</td>
</tr>
<tr>
<td>3. Give right to erasure / right to be forgotten</td>
<td>Required</td>
<td>Depends on strength</td>
<td>Not Required</td>
</tr>
<tr>
<td>4. Other data subject rights (access, portability...)</td>
<td>Required</td>
<td>Depends on strength</td>
<td>Not Required</td>
</tr>
<tr>
<td>5. Basis for cross-border transfers</td>
<td>Required</td>
<td>Required</td>
<td>Not Required</td>
</tr>
<tr>
<td>6. Data protection by design</td>
<td>Required</td>
<td>Partially met</td>
<td>Not Required</td>
</tr>
<tr>
<td>7. Data security</td>
<td>Required</td>
<td>Partially met</td>
<td>Not Required</td>
</tr>
<tr>
<td>8. Data breach notification</td>
<td>Likely</td>
<td>Less likely</td>
<td>Not Required</td>
</tr>
<tr>
<td>9. Data retention limitations</td>
<td>Required</td>
<td>Required</td>
<td>Not Required</td>
</tr>
<tr>
<td>10. Documentation / recordkeeping obligations</td>
<td>Required</td>
<td>Required</td>
<td>Not Required</td>
</tr>
<tr>
<td>11. Vendor / sub-processor management</td>
<td>Required</td>
<td>Required</td>
<td>Not Required</td>
</tr>
</tbody>
</table>

3. Analysis of Obligations

3.1. Notice to Data Subjects

A basic principle of data protection law is transparency and the obligation to provide notice to data subjects regarding the collection, use, and disclosure of personal information. Under the GDPR, data controllers must provide extensive details to data subjects whenever they process personal data. The text of the GDPR makes no distinction between fully identified personal data and pseudonymized personal data. Thus, the full range of mandated disclosures apply to any use or other processing of pseudonymized data. By contrast, because anonymized data is no longer considered personal data, none of the notice obligations apply to uses of data that has been fully anonymized.

---

4 GDPR Articles 13, 14 and 15. A full list of those details is included in Appendix B.
3.2. Consent and the Legal Basis for Processing

The GDPR, like its predecessor, the 1995 Data Protection Directive, requires there to be a legal basis to process personal data. The most well-known basis is the explicit consent of the data subject. However, under the GDPR, obtaining explicit consent can be difficult; and in some scenarios, such as research, big data analytics, and machine learning, obtaining explicit consent may be impractical or impossible.

However, there is language in the GDPR that sets out criteria for when a secondary use of data (such as for research or analysis) can proceed on a basis other than the consent of the data subject – in particular where the “processing for another purpose is compatible with the purpose for which the personal data are initially collected.” One of the key criteria to be used in determining whether such processing can proceed is “the existence of appropriate safeguards, which may include encryption or pseudonymization.” Thus, the use of pseudonymization, at least in some circumstances, can help enable data processing for secondary purposes without the need to obtain the explicit consent of the data subjects.

The Article 29 Working Party has made clear that the processing of personal data to fully anonymize such data is “compatible with the purpose for which the personal data are initially collected” and therefore does not require an additional legal basis. And once the data is fully anonymized, it is outside the scope of data protection law, and data controllers have no need to have or articulate a legal basis for further processing.

3.3. Right to Erasure / Right to be Forgotten

Article 17 of the GDPR creates a new right for data subjects to request that personal data about them be deleted. This right is referred to as the right to erasure, or the “right to be forgotten.” If certain criteria are met, data controllers are required to respond to such requests and to erase the personal data “without undue delay.” Further, if the controller has made the personal data public, it may be obligated to take “reasonable steps” to inform other controllers that may be processing the personal data of the erasure request.

Many organizations are finding that implementing this so-called “right to be forgotten” to be among the most onerous legal, technical, and operational challenges in their GDPR compliance efforts. Locating all copies of such personal data in all systems requires detailed data mapping efforts. Creating a scalable ability to granularly delete data related to a particular individual often require re-architecture of data systems and the development of new tools. Determining when data must be deleted under the GDPR can require case-by-case review of the facts and a legal determination based on those facts. And operationalizing the ability to receive erasure requests, authenticate the individuals making the

5 GDPR Article 6(1).
6 See Article 6(1)(a). Article 4(11) of the GDPR defines consent as being “freely given, specific, informed and unambiguous.” Compared to the definition in the 1995 Data Protection Directive, the GDPR definition adds the requirements that consent be “unambiguous,” which could be interpreted as raising the bar on what may constitute valid consent.
7 See Article 6(4). See also Article 5(1)(b).
8 GDPR Article 6(4)(e).
9 See Opinion 05/2014 on Anonymisation Techniques, at 7 (“the Working Party considers that anonymisation as an instance of further processing of personal data can be considered to be compatible with the original purposes of the processing but only on condition the anonymisation process is such as to reliably produce anonymised information in the sense described in this paper”).
requests, determining how to respond based on different criteria, locating and purging all copies of such data, and being able to demonstrate and document that such data has been fully erased will normally require a number of new resources and processes.

However, Article 12(2) of the GDPR specifies that if the data controller can demonstrate that it is not in a position to identify the data subject from the data it holds, it need not comply with a request to erase data. This provision reflects the common-sense proposition that if the data controller cannot reliably tie the data it holds to the individual asserting this right, it will be unable to comply.

However, many of not most pseudonymization methods will not result in this exemption applying. Many implementations of pseudonymization are readily reversible by a data controller, and others may contain indirect identifiers that could allow the data controller to match the data if the data subject making the request supplies sufficient data to allow the match. Therefore, the pseudonymization in such cases would not meet the standard reflected in Article 12(2). But there could be some implementations of strong pseudonymization that would. Thus, whether pseudonymization will result in relief from these GDPR obligations will depend on the strength of the method and implementation employed. And organizations asserting that the exemption applies will need to demonstrate or prove that their pseudonymization methods meets the standard.

For example, if a data analytics company maintains only the pseudonymized NHS number gender, and age of a patients participating in a clinical study, it would legitimately be difficult for them to identify the records that belong to a specific individual pursuant to an erasure request. In such a case there are insufficient indirect identifiers in the data.

Of course, de-identification that is even stronger, such that it meets for bar for full anonymization, will mean that the data will also be free of the obligation to delete.

### 3.4. Data Subject Rights

In addition to the right to erasure, or right to be forgotten, the GDPR gives data subjects a number of additional rights – including the rights of access, rectification, and data portability, as well as rights to object to the processing of personal data or obtain a restriction of such processing under certain circumstances.10

Here too, as with the right to be forgotten, the implementation challenges can be profound. But Article 12(2) of the GDPR also applies here and specifies that if the data controller can demonstrate that it is not in a position to identify the data subject from the data it holds, it need not comply with the Articles setting out these additional data subject rights. But as discussed above, it would have to be a particularly strong form of pseudonymization for this exemption to apply. If de-identification is even stronger, such that the data is fully anonymized, it will be outside the scope of the GDPR and therefore free of these obligations.

### 3.5. Cross-Border Data Transfers

Both existing European privacy law and the GDPR restrict the transfer of personal data outside the European Economic Area, except under certain conditions.11 For example, personal data may be transferred to the small number of jurisdictions that the European Commission has found to have “adequate” data protection regimes in place, transfers are allowed if subject to contracts that contain

10 See GDPR Articles 15(3)&(4) (right of access), 16 (right of rectification), 18 (right to restriction of processing), 20 (right to data portability), and 21 (right to object to processing).

11 See GDPR Article 44-49.
the model clauses approved by the Commission, or transfers may take place if the recipient is part of the EU-U.S. Privacy Shield. Both fully identified and pseudonymized personal data are equally subject to these restrictions. By contrast, these cross-border transfer restrictions do not apply to fully anonymized data.

3.6. Data Protection by Design and by Default

A new requirement imposed by the GDPR is referred to as “data protection by design and by default.” This new set of rules requires data controllers to “implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.”

This provision applies to both identified and pseudonymized personal data, but given that pseudonymization is called out as a key example of the types of measures required under this provision, data controllers can conclude that pseudonymizing data will at least partially satisfy this requirement, and that the number of additional measures to protect the data typically will be lower for pseudonymized data than for fully identified data. For fully anonymized data, no such measures are required because the data is no longer subject to this requirement.

3.7. Data Security

Controllers and processors handling personal data are obligated under the GDPR to implement measures sufficient “to ensure a level of security appropriate to the risk.” In gauging the level of risk posed by personal data, the level of de-identification applied is certainly a relevant factor. Thus, pseudonymized data will typically pose a lower risk than fully identified data, and therefore the level of security measures required will normally be reduced for pseudonymized data. In fact, the text of the GDPR suggests that pseudonymization itself can be thought of as a security measure that safeguards personal data. Here too, fully anonymized data is different in that it is no longer subject to the GDPR requirements and therefore the security obligations do not apply.

3.8. Data Breach Notification

The GDPR introduces new requirements to notify supervisory authorities and/or data subjects in the event of a breach of personal data. Supervisory authorities must be notified “unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons.” And data subjects must be notified if “the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons.” As with data security, the risk assessment for these provisions will certainly take into account the level of de-identification of the data. In the event of a data breach, fully identified personal data will almost always pose a greater risk than if that data were pseudonymized. Thus, while the need for notification is likely in the event identified data is breached, it is less likely if pseudonymized data is

---

12 GDPR Article 25(1).
13 GDPR Article 32(1).
14 See, for example, Article 32(1)(a) which lists pseudonymization as one of the “appropriate technical organization measures,” and Article 6(4) which refers to “appropriate safeguards, which may include encryption or pseudonymisation.”
15 GDPR Article 33(1).
16 GDPR Article 34(1).
breached. Here too, however, fully anonymized data does not trigger a breach notification obligation at all.

### 3.9. Data Retention Limitations

Data minimization principles reflected in the GDPR require that personal data not be retained longer than necessary to carry out the legitimate purposes of processing. Thus, data controllers must evaluate their needs to retain data and establish appropriate retention schedules for the personal data they hold. Data controllers may argue that from a policy standpoint, they should have flexibility to retain pseudonymized data longer, particularly where the data is highly useful even if not strictly necessary. But those arguments to not relieve the data controller from the obligations to assess and establish data retention timeframes for all personal data they hold – whether identified or pseudonymous.

With respect to data retention, full anonymization of data is considered the functional equivalent of deletion, and fully anonymized data may be kept indefinitely.

### 3.10. Documentation and Recordkeeping Obligations

The GDPR imposes far more documentation and recordkeeping obligations on data controllers and processors than is the case under current EU data protection law. In almost all cases, these obligations apply equally to the processing of both fully identified and pseudonymized data. By contrast, they do not apply to anonymized data.

### 3.11. Vendor or Sub-Processor Management

A number of requirements in the GDPR apply to the use of vendors, processors, or sub-processors who handle or access personal data. Some dictate certain provisions that must be in a contract between a data controller and a data processor. Others apply directly to data processors who are processing data on behalf of a controller. Still others regulate the use of sub-processors and the obligation to pass through certain requirements to such entities. Collectively, these requirements obligate organizations to have robust vendor management programs in place to carefully manage vendors who may touch personal data and ensure such data is protected as it is passed from one entity to another. These provisions apply equally to identified and pseudonymized data. However, they do not apply to anonymized data.

### 4. Conclusions

The above discussion and the summary chart in the introduction make clear that pseudonymized data is far more similar to identified data than it is to anonymized data in terms of the GDPR obligations that apply to it. While pseudonymization can form part of an overall GDPR compliance strategy in certain

---

17 See, for example, Article 30. Appendix C describes the key documentation requirements in more detail.
18 There is a narrow exception under Article 30(5) that applies to small organizations (under 250 employees) that do not process sensitive personal data and only occasionally process personal data at all. In such a case, the organization must to a risk assessment to determine whether the Article 30 documentation requirements apply, and the use of pseudonymization could play a factor in such assessment.
19 See GDPR Article 28(3).
20 See, e.g., Article 29 limiting the scope of data processors’ activities to that which is directed by the data controller or as required by law, and the recordkeeping obligations of processors under Article 30(2).
21 See GDPR Article 28(2)&(4).
cases, it does not result in complete relief from GDPR obligations in the way anonymization does. Thus, organizations should not confuse the limited advantages of pseudonymization with the far more sweeping advantages of anonymization.

5. About the Authors

**Mike Hintze**

Mike Hintze is a partner at Hintze Law PLLC. As a recognized leader in the field, he advises companies, industry associations, and other organizations on global privacy and data protection law, policy, and strategy. He was previously Chief Privacy Counsel at Microsoft, where, for over 18 years, he counseled on data protection compliance globally, and helped lead the company’s strategic initiatives on privacy differentiation and public policy. Mike also teaches privacy law at the University of Washington School of Law, serves as an advisor to the American Law Institute’s project on Information Privacy Principles, and has served on multiple advisory boards for the International Association of Privacy Professionals and other organizations. Mike has testified before Congress, state legislatures, and European regulators; and he is a sought-after speaker and regular writer on data protection issues.

Prior to joining Microsoft, Mike was an associate with Steptoe & Johnson LLP, which he joined following a judicial clerkship with the Washington State Supreme Court. Mike is a graduate of the University of Washington and the Columbia University School of Law.

Privacy Analytics Inc. is a client of Hintze Law PLLC, and support for Mike Hintze’s contribution to this White Paper has been provided by Privacy Analytics Inc. Neither Mike Hintze’s representation of Privacy Analytics Inc. nor his contribution to this White Paper serves as an endorsement of any technology, products or services of Privacy Analytics Inc.

**Khaled El Emam**

Dr. Khaled El Emam is the founder of Privacy Analytics Inc. and Director of Real World Evidence Solutions. As an entrepreneur, Khaled helped found five companies involved with data management and data analytics. He has worked in technical and management positions in academic and business settings in England, Scotland and Japan.

Khaled is also a senior scientist at the Children’s Hospital of Eastern Ontario (CHEO) Research Institute and Director of the multi-disciplinary Electronic Health Information Laboratory (EHIL) team, conducting academic research on de-identification and re-identification risk. He is a world-renowned expert in statistical de-identification and re-identification risk measurement. He is one of only a handful of individual experts in North America qualified to anonymize Protected Health Information under the HIPAA Privacy Rule.

In 2003 and 2004, Khaled was ranked as the top systems and software engineering scholar worldwide by the Journal of Systems and Software based on his research on measurement and quality evaluation and improvement.

Previously, Khaled was a Senior Research Officer at the National Research Council of Canada. He also served as the head of the Quantitative Methods Group at the Fraunhofer Institute in Kaiserslautern, Germany.

Khaled was one of the first Privacy by Design Ambassadors recognized by the Ontario Information and Privacy Commissioner. He previously held the Canada Research Chair in Electronic Health Information at
the University of Ottawa and is an Associate Professor in the Faculty of Medicine at the University. He has a PhD from the Department of Electrical and Electronics Engineering, King’s College, at the University of London, England.
Appendix A: ANONYMIZATION METHODS

The Article 29 Working Party has provided recent guidance on data anonymization techniques  in which they recommended two approaches: the data must meet three specific criteria or a re-identification risk analysis must be performed to demonstrate that the risk is acceptably small. The three criteria are that the data must not have any of the following properties:

- **Singling out**: possibility to isolate some records of an individual in the dataset
- **Linkability**: ability to link, at least, two records concerning the same data subject or a group of data subjects (in the same database or in two different databases);
- **Inference**: the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes

An analysis of the working party opinion has argued that meeting these criteria strictly would result in data sets with limited utility. For example, it would be difficult to create anonymized longitudinal data. Therefore, as a practical matter, the risk-based approach is the one that is applied.

Risk-based anonymization methods are consistent with recommendations from the Information Commissioner’s Office in the UK, anonymization guidance from European Medicines Agency, the privacy commissioner of Ontario, the Expert Determination methods under the HIPAA Privacy Rule in the US, and other governmental, academic, and professional associations and groups. In this

---

29 Institute of Medicine, ‘Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk’ (2015); The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, ‘Accessing Health And Health-Related Data in Canada’ (Council of Canadian Academies 2015); PhUSE De-Identification Working Group, ‘De-Identification Standards for CDISC SDTM 3.2’ (2015); Mark Elliot and others, Anonymisation Decision-
appendix we provide an outline of risk-based anonymization methods that have been specifically developed and applied for health data.

**Basic Principles of Risk-based Anonymization**

There are three fundamental concepts underpinning risk-based anonymization methods.

The first concept differentiates among the different types of information in data. Data may have direct identifiers, which are things like a patient’s NHS number or social security number. Direct identifiers are assumed to directly identify a patient with a high probability of success. These are typically pseudonymized or removed to create pseudonymous data. This type of data is still considered personal information. Another type of information would be quasi- (or indirect) identifiers. These are things like demographics (e.g., age, gender, race), socio-economic information (e.g., income, and years of education), and clinical events (e.g., rare diagnoses, and hospital visits). There is evidence that this kind of information can still identify individuals. Dealing with this kind of information can produce anonymous data. Figure 1 below presents further examples of direct and quasi-identifiers.

**Examples of direct identifiers:** Name, address, telephone number, fax number, MRN, health card number, health plan beneficiary number, VID, license plate number, email address, photograph, biometrics, SSN, SIN, device number, clinical trial record number

**Examples of quasi-identifiers:** sex, date of birth or age, geographic locations (such as postal codes, census geography, information about proximity to known or unique landmarks), language spoken at home, ethnic origin, total years of schooling, marital status, criminal history, total income, visible minority status, profession, event dates, number of children, high level diagnoses and procedures

Figure 1: Direct and Quasi-identifiers


Therefore, the act of anonymization is focused on the quasi-identifiers only. The assumption is that pseudonymization has already been applied to address re-identification risks from direct identifiers.

The second is that risk-based methods are quantitative. The quantity that is being measured is the risk of re-identification of an individual in the data. The initial step is to set an acceptable threshold for this risk. This means that the acceptable risk of re-identification is going to be some value larger than zero. Setting the threshold at zero risk means that no useful data will be shared at all.

The actual risk of re-identification is then measured on the data. This measured value is compared to the threshold. If the measured risk is above the threshold then the data is not considered anonymous. If the measured risk is below the threshold then the data is considered anonymous. If the data is not anonymous then various transformations can be applied to bring the measured risk below the threshold. These transformation may include generalizing certain values in the data (for example, generalizing a date of birth to a year of birth), or suppressing certain values in the data that make individuals stand out.

The general measurement process is illustrated in Figure 2.

---

**Figure 2:** Overview of the risk measurement process.

The third concept pertains to the context of the data. The actual risk of re-identification is a function of both the data and the context. The context represents the security, privacy, and contractual controls that are in place. For example, one context can be a public data release (e.g., an open data initiative). Another context would be a researcher who analyzes the data in a very secure enclave. These are two very different contexts and the risk of re-identification is different in each of these, even for the same data.
The exact definition of the context is illustrated in Figure 3. It consists of characteristics of the data recipient / holder, the contract or data use agreement, and the data itself 33.

![Figure 3: Definition of the data context.](image)

This is illustrated in the diagram in Figure 4. Here we see that the overall risk is a function of both, the data risk and the context risk. When expressed as probabilities, the overall risk of re-identification is the multiplication of these two numbers.

This means that the same data can have different levels of risk if it is processed in different contexts. But it also means that the same data can have different risk levels as it moves from one organization to another in the same data flow (i.e., over time). For example, if the data moves from an organization performing analytics to a say, a researcher, the risk may be low in the first instance but increase in the second instance after the transfer.

---

33 El Emam (n 47).
Figure 4: Re-identification risk is a function of data risk and context risk.
Anonymization Process

The steps of the combined pseudonymization and full anonymization process are summarized at a high level in Figure 5. Once the variables in the data have been classified into direct and quasi-identifiers, the direct identifiers can be pseudonymized, and then the quasi-identifiers can be dealt with to anonymize the data.

Documentation of the process and results is important to ensure that questions about the assumptions can be answered at a later date if needed, and to allow replicability of the analysis.

---

**Figure 5:** The anonymization process.
Appendix B: ADDITIONAL DETAILS REGARDING THE REQUIRED NOTICE TO DATA SUBJECTS UNDER THE GDPR

Articles 13, 14, and 15 of the GDPR set out a long list of items that organizations must include in privacy notices provided to individual data subjects. In this respect, the GDPR represents a significant change from the 1995 Data Protection Directive that it replaces, which specified a much more limited set of information that must be included in a privacy notice. Under the GDPR, notices must include:

- The identity and the contact details of the controller and, where applicable, of the controller’s representative;\(^{34}\)
- The contact details of the data protection officer, where applicable;\(^{35}\)
- Where personal data is obtained from a source other than the data subject:
  - The types of personal data obtained,\(^{36}\) and
  - The source(s) “from which the personal data originate, and if applicable, whether it came from publicly accessible sources”;\(^{37}\)
- Where the personal data is collected from the data subject, whether providing the data is required, including:
  - Whether it is a requirement necessary to enter into a contract,
  - Whether it is otherwise required by statute or contract, and
  - The possible consequences of the failure to provide such data;\(^{38}\)
- The intended purposes of processing the personal data;\(^{39}\)
- The legal basis for the processing;\(^{40}\)
- Where the legal basis for processing is “the legitimate interests pursued by the controller or a third party under Article 6(1)(f),” a description of those interests;\(^{41}\)
- Where the legal basis for processing is “the consent of the data subject under Articles 6(1)(a) or 9(2)(a),” the existence of the right to withdraw such consent at any time (which will not affect the lawfulness of any processing that occurred before such consent is withdrawn);\(^{42}\)
- Where personal data is used for automated decision making, including profiling, referred to in Article 22(1) and (4), the existence of such processing, meaningful information about the logic

---

\(^{34}\) Id. arts. 13(1)(a), 14(1)(a).

\(^{35}\) Id. arts. 13(1)(b), 14(1)(b). See id. art. 37, for the requirements for designating a data protection officer.

\(^{36}\) Id. arts. 14(1)(d), 15(1)(b).

\(^{37}\) Id. arts. 14(2)(f), 15(1)(g); see also id. Recital 61 (“Where the origin of the personal data cannot be provided to the data subject because various sources have been used, general information should be provided.”).

\(^{38}\) Id. art. 13(2)(e).

\(^{39}\) Id. arts. 13(1)(c), 14(1)(c), 15(1)(a). Note that “processing” is defined broadly, and includes any collection, use, or sharing of personal data, see id. art. 4(2).

\(^{40}\) Id. arts. 13(1)(c), 14(1)(c); see also id. art. 6 (listing the legal bases for processing personal data).

\(^{41}\) Id. arts. 13(1)(d), 14(2)(b).

\(^{42}\) Id. arts. 13(2)(c), 14(2)(d).
involved, and the significance of the processing and any anticipated consequences for the data subject;  

- “The recipients or categories of recipients of the personal data, if any;  

- The period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;  

- The existence of the right of a data subject to:  
  - Request from the controller “access to and rectification or erasure of personal data” or,  
  - Object to the processing of personal data or obtain a restriction of such processing under certain circumstances;  

- Receive data he or she has provided to the controller in a structured, commonly used and machine-readable format, and transmit that data to another controller (data portability);  

- The right to lodge a complaint with a supervisory authority; and,  

- Where the controller intends to transfer personal data to a third country or international organization, the fact of such transfer and either:  
  - The existence or absence of an adequacy decision by the [European] Commission, or  
  - In the case of transfers based on “suitable safeguards” under Articles 46, 47, or 49(1)(b) (such as contractual provisions or binding corporate rules), a description of such safeguards and how to obtain a copy of them.  

\[\text{43 Id. arts. 13(2)(f), 14(2)(g), 15(1)(h).}\]  
\[\text{44 Id. arts. 13(1)(e), 14(1)(e), 15(1)(c).}\]  
\[\text{45 Id. arts. 13(2)(a), 14(2)(a), 15(1)(d).}\]  
\[\text{46 Id. arts. 13(2)(b), 14(2)(c), 15(1)(e); see also id. art. 15 (right of access), art. 16 (right to rectification), art. 17–44 (right to erasure).}\]  
\[\text{47 Id. arts. 13–15. The right to object applies to processing based on Article 6(1)(e) (“necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller”) or Article 6(1)(f) (“necessary for the purposes of the legitimate interests pursued by the controller or by a third party”), or for the purposes of marketing. Id. arts. 6(1)(e)–(f), art. 21(1)&(2). The right to obtain a restriction on processing applies under four narrow circumstances described in Article 18(1). See id. An organization may choose to specify these circumstances in its privacy statement in order to avoid implying a broader right to object or restrict processing than is provided by the GDPR.}\]  
\[\text{48 Id. art. 12(7). See also id. art. 20, for the scope of the data portability obligations.}\]  
\[\text{49 Id. arts. 13(2)(d), 14(2)(e), 15(1)(f).}\]  
\[\text{50 Id. arts. 13(1)(f), 14(1)(f); see also id. art. 15(2).}\]
Appendix C: DOCUMENTATION AND RECORDKEEPING OBLIGATIONS UNDER THE GDPR

Article 30 of the GDPR sets out specific records that must be retained by a data controller or processor with respect to the processing of personal data. For data controllers, these records include:

• the name and contact details of the controller and, where applicable, the joint controller, the controller’s representative, and the data protection officer;
• the purposes of the processing;
• a description of the categories of data subjects and of the categories of personal data;
• the categories of recipients to whom the personal data have been or will be disclosed;
• where applicable, transfers of personal data to a third country or an international organization, including the identification of that third country or international organization, and the documentation of suitable safeguards as applicable;
• where possible, the envisaged time limits for erasure of the different categories of data;
• where possible, a general description of the technical and organizational security measures referred to in Article 32(1).

Similar requirements are set out for data processors. Organizations that operate as both controllers and processors would need to maintain records for both types of processing activities.

Additionally, the requirements under Article 35 for “data protection impact assessments” will result in additional mandated documentation of any data processing activities of the organization that potentially involve a “high risk” to the privacy rights of individuals. As a practical matter, organizations are likely to conclude that some form of documented assessment will need to be conducted for new technologies that involve the collection or use of personal data, for data analytics that involve large volumes of personal data, or other data processing that is not obviously low risk.

Other GDPR obligations, such as the notice requirements discussed in Appendix B above, will also require extensive documentation.