ALSTON & BIRD

Data Protection Advisory

The Importance of Accurate Retrieval of Data Subjects' Personal Data in Complying with GDPR Individual Rights Requirements[1]

By Jan Dhont, Peter Swire & DeBrae Kennedy-Mayo[2]

The General Data Protection Regulation, which enters into effect on May 25, 2018, goes considerably beyond existing law in setting forth individual rights that allow data subjects to control how their personal data is used. This White Paper addresses a fundamental issue for implementing individual rights – how can those companies who process data ensure that they uniquely identify data subjects when administering their data subject rights?

In responding to individual rights requests, the ad hoc measures that many companies have employed to date may in many instances no longer be sufficient to comply with the GDPR. Companies that process personal data face greatly increased potential fines.[3] These companies – both the controllers who determine how personal data may be used, and the processors who act on their behalf – thus have strong reason to discover and implement effective measures to respond to requests to uphold individual rights.

This White Paper briefly describes the individual rights that are most salient under the GDPR, including the right of data subjects to access their personal data and rectify inaccuracies in such data. It then examines key technical issues for pulling together the relevant data in a company's many databases, while excluding the irrelevant data. The Paper highlights two crucial goals for processing individual rights requests:

- One requirement is for accurate "entity resolution," which means linking the relevant data with each person.
- Another requirement is to achieve this entity resolution while acting consistently with the many other requirements of the GDPR, including data minimization and avoiding the violation of other data subjects' rights, such as ensuring that personal data is released to the correct data subject in the context of a data access or portability request.

[1] Research support for this White Paper was provided by Senzing, Inc. The opinions expressed here are those of the authors and do not constitute legal advice

[2] Jan Dhont is a partner in the Brussels office, leading Alston & Bird's Privacy & Data Security Practice in Europe. Peter Swire is Senior Counsel to Alston & Bird, and the Holder Chair of Law and Ethics at the Georgia Tech Scheller College of Business. DeBrae Kennedy-Mayo is a consulting attorney with Alston & Bird, and a research faculty member at the Georgia Tech Scheller College of Business.

[3] The requirements imposed by GDPR are not limited to companies, but include organizations and governments. Article 4(7), 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, 27 April 2016 (hereinafter "GDPR") (defining "controller" to include "natural or legal person, public authority, agency or other body").

This White Paper analyzes the use of entity resolution to implement “federated fetch” – the ability to fetch data as required from all of a company’s diverse databases while minimizing the risk of inaccurately processing the wrong data. Under guidance from a key regulatory body, the Article 29 Working Party, there is a strong case that this approach provides high-quality compliance with the individual rights requirements of the GDPR, while being implementable at modest cost. Other approaches for compliance should be assessed against the general principles of EU data protection law, including accuracy of data and data minimization.

1. INDIVIDUAL RIGHTS REQUESTS UNDER THE GDPR

Under the GDPR, data subjects have several individual rights, including: the right to access one’s own data; the right to rectification; the right to erasure; the right to restriction of processing; the right to data portability; and the right to object to processing.[4] A company must ensure timely handling of individual rights requests that it receives.[5] A data subject has the right to obtain confirmation from the company, within one month of receipt of a request, on action taken by the company in response to the request.[6] It is quintessential for the administration of these rights that companies accurately pinpoint what personal data corresponds to a particular data subject. Without accurate identification of the personal data of data subjects, companies expose themselves to information security incidents and associated liability exposure.

A. The GDPR Article 15 - Subject Access Requests

Article 15 confers on data subjects the right to obtain confirmation from the company as to whether the company is processing personal data about the data subject. This request is also known as Subject Access Requests or “SARs”. In case a failure to comply with a request is due to the fact the company is not able to identify or retrieve the relevant data, this may constitute a violation of Article 15 of the GDPR and may give rise to the highest fines under the GDPR (i.e., up to 4 percent of the company’s global turnover).[7] To handle the requirements under this article, the company must know where relevant data resides and, once located, whether that data relates to the data subject making the SAR.

B. The GDPR Article 16 - Right of Rectification

Article 16 allows the data subject to have inaccurate personal data about him/her corrected and to have incomplete personal data completed. For a company to comply, it must be able to identify inaccuracies (e.g., company records reflect two people when in fact they are one) or supplement incomplete data (e.g., a credit history is missing positive information about the data subject).

[4] The data subject also has the right not to be subject to a decision based solely on automated decision-making, including profiling, but this right is more limited in scope and will not be elaborated on here. Article 22(1), GDPR; see Article 15, 16, 17, 18, 20, 21 and 22, GDPR.

[5] Article 12(3), GDPR.

[6] Note that this delay is extendable by two additional months where necessary. See Article 12(3), GDPR.

[7] Failure to comply with a SAR may give rise to an administrative fine of 4% of worldwide turnover or 20M Euros, whichever is higher. See Article 83(5)(b), GDPR.

C. The GDPR Article 17 - Right to Erasure (Right to be Forgotten)

Article 17 describes the Right to Erasure, which is more commonly known as the Right to be Forgotten. This right enables a data subject to request the deletion or removal of personal data where there is no compelling reason for its continued processing. Importantly, the right can be triggered either when a data subject objects to the processing, when he withdraws his consent, or where the personal data was unlawfully processed. If a data subject objects to processing on the grounds that data about the data subject held by the company is inaccurate, this could provide a basis for triggering the right to be forgotten. To fully comply with the requirements of GDPR in relation to this right, the company must be able to identify the data subject at issue in the deletion request within a company's multiple systems. The company must also identify whether there are reasons to reject the deletion request and maintain the data, for instance when further processing of that data is necessary for compliance with a legal obligation, or for reasons of public interest including public health.[8]

D. The GDPR Article 18 - Right to Restriction of Processing

Article 18 enables data subjects to restrict processing of personal data. A data subject can, for instance, exercise this right if he/she contests the accuracy of personal data held by a company, or in case processing is unlawful but the data subject does not want the data to be erased. In the case of contesting accuracy, when processing is restricted, a company is permitted to store the personal data, but not further process it until it has verified its accuracy.[9]

For all of these individual rights under the GDPR, the company must be able to identify the data subject's personal data across its information systems. In addition, the company must be able to verify whether there are limitations to (part of) the relevant data, preventing it from being subject to the individual right request (for instance, in case the data is protected by intellectual property rights or in case it is mixed up with another data subject's data). The best way to approach this requirement systematically and within the GDPR-prescribed time periods, is to employ technology.

2. TECHNOLOGICAL APPROACH FOR COMPANIES TO HANDLE REQUESTS FROM DATA SUBJECTS

For the individual rights of the data subjects to be exercised and handled, companies need to have an effective method for single-subject search – to produce the records that respond to the request, while screening out all the other records. The critical challenges are that the data is scattered and inconsistent. In this setting, "scattered" means that the company typically has many systems to review, numbering in the hundreds for large organizations. Most companies are already aware of the challenges of finding the

[8] Article 17(3), GDPR.

[9] Note that the company, during this period of verification, also retains the right to process the data for the defense of legal claims, for important reasons of public interest, for the protection of another data subject's rights, or with the consent of the data subject. See Article 18(2), GDPR.

relevant data across all the company's systems. The problem of "inconsistent" data is perhaps less familiar. "Inconsistent" means that the data subject's personal data is often entered with variations that must be resolved.[10] As a simple example, some of the records of William Smith may be stored under Will Smith or Bill Smith. In addition, there are many people who share that name, so the company needs to determine whether each record is actually attributed to the William Smith making the data subject request. A combination of two methods, entity resolution and federated fetch, effectively allows companies to carry out the mandates of the GDPR, and to accomplish this within the required time-limits.

A. Entity Resolution Assists the Company in Finding the Correct Person

With the many information systems utilized by a company in its numerous departments, there must be a method to address the multiple variations of a data subject's name as well as other identifying information that the company holds about the data subject such as address, phone number, and government-issued identification number. This task is known as entity resolution.

Entity resolution allows the company to focus on more than the specific name of the data subject to ensure that all personal data about a specific data subject is located. It takes into account multiple variations of the name – married name, misspellings, derivations, and transliterations of name. For example, if a data subject reports the name of "Jan Peeters," this method may report back instances of the male names John Peeters, Johannes Peeters, and Jann Peters as well as the female names Janice Peeters, Janet Peeters, and Jan Peeters Martin for further analysis. It also takes into account multiple addresses, phone numbers, dates of birth, etc. -- and their misspellings and numerical errors and transpositions.

In smaller organizations, entity resolution has often been done informally, such as based on knowledge of an employee or a manual assessment of which information matches which person. In companies with more data about more people, the challenge of accurate entity resolution has been quite substantial.[11]

To illustrate the entity resolution issue, let's imagine that the company that receives a request from Jan Peeters at 1234 Bedford Road has numerous systems that might hold information on that person. These systems could be related, for instance, to the payroll department, a loyalty program, an applicant pool, and employee complaints. The customer loyalty program could have a record of Jan Peeters at 1234 Bedford Road, while the applicant pool's record of Jan Peeters is located at 1235 Bedford Road. Payroll lists a Janice Peeters, with government-issued identification number 123-45-6789. Utilizing the entity resolution approach, the company can have a provisional guess that Jan Peeters #1 is the data subject related to the payroll record and the applicant pool's

[10] See Michelle Levin, Chapter 9 – Data Subjects' Rights, European Privacy: Law and Practice for Data Protection Professionals, p. 131-132, IAPP (ed. Eduardo Ustaran, 2012) (explaining that some of the practical difficulties under Directive 95/46/EC of handling data subject requests related to individual rights are: "coping with the volume of information that needs to be searched;" "identifying the personal data of the individual making the request, whilst not infringing the right to privacy of third parties that are also identified in the data;" and "identifying and complying with the applicable time limits").

[11] For more detailed discussion of entity resolution, see Senzing, *Finding the Missing Link in GDPR Compliance*, 2018, available at https://senzing.com/gdpr/ (last visited April 2018).

record, while there is not enough data about Janice Peeters with the government ID number to assign that record to Jan Peeters #1. An effective entity resolution system examines the available data sources to create a searchable index – key to improving the likelihood of accurately linking the records of one person, while excluding the records of others.

One metaphor for an entity resolved searchable index is the traditional card catalogue in a large library, with one or more library cards each indicating the location of one book.[12] In the current example, imagine that the first two records are placed together, with a rubber band around them. Based on the original information that is available, the third record is kept separate.

To achieve accurate single-subject search, it is important to entity resolve new information as it comes in. As a first example, the company may learn later that there really is one Jan Peeters at 1234 Bedford Road, while a different Jan Peeters lives next door at 1235. At that point, the rubber band comes off, and the two library cards are separated. As another example, suppose that the company later learns that the address of Janice Peeters, with government ID number 123-45-6789, is actually 1234 Bedford Road (i.e., the address corresponding to Jan Peeters). With that new information, the rubber band would go around all three cards. The company should then include the Janice Peeters information together with the Jan Peeters information as the available data shows they now apply to the same person.

B. Federated Fetch Locates Personal Data with a Minimum Amount of Processing

An entity resolved searchable index solves two key challenges: scattered and inconsistent data. The index approach addresses the problem of inconsistency, because records are placed together via an entity resolution process (rubber bands around cards from the catalog that appear to belong together). This entity resolution reduces the risk of missing records, and of improperly placing records together that do not belong together. The challenge of scattered data is overcome as the searchable index reveals pointers to source systems and record locators. We use the term “federated fetch” to describe this process of having pointers in the index (cards in the card catalog), with the pointers providing the means to precisely query specific systems for specific records.

In considering the challenge of locating one person’s records across many systems, imagine a continuum between one highly centralized database, on the one hand, and numerous very decentralized systems, on the other. The centralized database is sometimes called a “data lake.” This data lake approach raises concerns under the principle of data minimization – personal data is gathered in all of the sub-systems of the company, then replicated (processed) into the data lake, and processed again every time it is touched in the data lake (e.g., marketing analytics). The data lake approach also increases the magnitude of harm if there is a data breach – “big data” can lead to “big data breach.” With the data lake, a single breach risks compromising all of the personal data held by the company.

[12] A traditional card catalogue would often have multiple cards that pointed to the same book, listed, for instance, by title, author, and subject matter. Each of the cards would provide the location of the same book.

At the other end of the spectrum, consider a searchable index with the federated fetch approach for the goal of accurate single-subject search. The term "federated" here means that personal data (including events and transactions) is held within each sub-system, and not centralized into a data lake. The idea of "fetch" fits the metaphor of the card catalogue, where the card catalog provides a mechanism for looking up the correct book (correct data subject), while keeping the books themselves (personal data about the data subject) out of the central repository. This federated fetch approach honors the principle of data minimization far better than a data lake, because the data is not processed multiple times as it enters and leaves the data lake(s). The searchable index and federated fetch approach also reduces the risk of data breach. A breach of the card catalogue does not actually reveal the personal data held by the company – no one can read the entire book simply by seeing the card and the book's location.[13]

When combined, an entity resolved searchable index and federated fetch allow a company to locate all of its personal data about a data subject in a quick timeframe, even when the company holds large amounts of data across numerous systems. A high-quality system of entity resolution reduces the likelihood of returning any personal data that actually pertains to a different data subject, and reduces the likely magnitude of harm if there is a data breach.

3. REDUCING RISKS RELATED TO HANDLING DATA SUBJECT REQUESTS BY EMPLOYING TECHNICAL MEASURES FOR DATA CONFIDENTIALITY, DATA MINIMIZATION, AND COST SAVINGS

The technical measures of entity resolution, central searchable index and federated fetch provide an effective approach to finding all of a data subject's personal data and providing it to the subject in a timely and cost-effective manner. As will be discussed below, if companies do not use technology to handle data subjects' individual rights requests, they risk violating their requirements under the GDPR.

A. Overview of Legal Requirements to Use Technical Tools to Meet Requirements of the GDPR

Through its overarching goals, the GDPR provides a framework for how companies should conduct searches to respond to requests by data subjects. The goals of the GDPR include minimizing processing of data subjects' personal data and conducting searches about data subjects for narrow purposes to have the least impact possible on the data subjects. Article 5 of the GDPR codifies a series of data protection principles – data minimization[14], purpose limitation[15], storage limitation[16], accuracy[17], accountability[18], and

[13] As discussed further below, "federated fetch" can be distinguished from an industry practice sometimes called "federated search." See infra Part 3B, 3C.

[14] Data Minimization – limit processing to the personal data necessary for obtaining the processing purposes. See Article 5(1)(c), GDPR.

[15] Purpose Limitation (including archival purposes) – only process personal data for specific, explicit and legitimate purposes. See Article 5(1)(b), GDPR.

[16] Storage Limitation – only retain personal data for as long as necessary to attain processing purposes. See Article 5(1)(e), GDPR.

[17] Accuracy – ensure timely rectification or erasure of inaccurate data. See Article 5(1)(d), GDPR.

[18] Accountability – the ability to demonstrate compliance. See Article 5(2), GDPR.

integrity and confidentiality.[19]

i. The GDPR Mandates Use of Technical Tools By Its Adoption Of Data-Protection-By-Design Requirements

In its data-protection-by-design provisions, Article 25 of the GDPR requires "appropriate technical … measures" to implement the data protection principles outlined in Article 5, in an "effective manner" that achieves the appropriate safeguards for processing the data.[20]

For companies with substantial databases of personal data, the reference to "technical measures" calls into question a manual, ad hoc approach for searching systems for electronic personal data of the requestor. Unlike the Data Protection Directive which is silent on the concept of data privacy by design, the GDPR specifically states that companies should employ technical approaches to ensure compliance with individual rights.[21]

ii. Article 29 Working Party Provides Advice on Using Electronic Tool For Extraction of Data

With regard to locating the personal data of a data subject who has made a request related to individual rights under the GDPR, using a technical tool for the electronic extraction of data is a first step in following the mandates of the Regulation.[22]

In its guidance, the Article 29 Working Party acknowledges the lack of explicit instruction in the GDPR of how a company should deal with the large quantities of information it holds about data subjects.[23] As to using an electronic tool to comply with the GDPR, the Article 29 Working Party provides the following guidance:

> "On a technical level, data controllers should explore and assess two different and complementary paths for making portable data available to the data subjects or to other data controllers:

[19] Integrity and Confidentiality – ensure appropriate security of the personal data. See Article 5(1)(f), GDPR.

[20] Article 25(1), GDPR.

[21] Article 25 has no exact equivalent in the Data Protection Directive. See Lina Jasmontaite, et al., "Implementation of Data Protection by Design and by Default: Framing Guiding Principles into Applicable Rules," available at https://edps.europa.eu/ipen-workshop-2017_en (last visited April 2018). [ENISA has guidance on Privacy Enhancing Technologies (PETs) – the precursor to Data Protection by Design. See Privacy by Design, ENISA's Data Protection, available at https://www.enisa.europa.eu/topics/data-protection/privacy-by-design (last visited April 2018); Privacy Enhancing Technologies, ENISA's Data Protection, available at https://www.enisa.europa.eu/topics/data-protection/privacy-enhancing-technologies (last visited April 2018).]

[22] Guidelines on the Right to Data Portability, p. 16, Article 29 Data Protection Working Party (April 5, 2017), available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233 (last visited April 2018). Although the Article 29 Working Party provided this guidance specifically in relation to data portability, it makes sense that the guidance can be scaled more generally to the requirements in the GDPR relating to individual rights of data subjects. See Stronger Protection, New Opportunities – Commission Guidance on the Direct Application of the General Data Protection Regulation as of 25 May 2018, p. 6-7, European Commission (Jan. 24, 2018), available at https://ec.europa.eu/commission/sites/beta-political/files/data-protection-communication-com.2018.43.3_en.pdf (last visited April 2018). [Guidance from Article 29 Data Protection Working Party is currently available in several areas, and not yet released in others. See Guidelines from the Article 29 Working Party, Justice and Consumers, European Commission, available at http://ec.europa.eu/newsroom/article29/news.cfm?item_type=1360 (last visited April 2018).]

[23] "The GDPR does not explain how to address the challenge of responding where a large data collection, a complex data structure or other technical issues arise that might create difficulties for data controllers or data subjects." Guidelines on the Right to Data Portability, p. 18.

> - a direct transmission of the overall dataset of portable data (or several extracts of parts of the global dataset);
>
> - an automated tool that allows extraction of relevant data.
>
> The second way may be preferred by data controllers in cases involving [ ] complex and large data sets, as it allows for the extraction of any part of the data-set that is relevant for the data subject in the context of his or her request, [and] may help minimising risk[24]…"

In sum, the Article 29 Working Party supports the deployment of a technical tool to extract data to handle a data subject's rights request, particularly when companies have a complex system environment.

iii. Guidance from European Data Protection Supervisor (EDPS)

As discussed below for the index of pointers, one part of the electronic extraction of information related to a data subject is the use of a central registry or index. The European Data Protection Supervisor (EDPS) has issued guidance for government entities to comply with data protection rights of data subjects that are similar to those found in the GDPR. In its guidance, the EDPS "strongly recommends" that government agencies keep a central registry of records to further their requirement to make records available upon request – a requirement with similarities to the company requirement under the GDPR to comply with data subjects' individual rights.[25]

In making this recommendation, the EDPS highlighted multiple benefits of a central registry including allowing better response to requests for records, making records easier to compare for quality control, and ensuring internal compliance with legal requirements.[26]

B. Single Subject Search with Entity Resolution Is a Technical Solution To Address Data Accuracy Concerns When Responding To Data Subject Requests Related To Individual Rights

Companies seeking to handle requests of data subjects exercising their individual rights under the GDPR must be concerned with data accuracy. Companies must be able to search all their systems, and to identify only the personal data of the correct data subject.

i. Legal Requirements For Accuracy

Article 5 of the GDPR states that personal data must be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they

[24] Guidelines on the Right to Data Portability, p. 16.

[25] Accountability on the Ground Part I: Records, Registers and When to Do Data Protection Impact Assessments, p. 4, European Data Protection Supervisor (February 2018), available at https://edps.europa.eu/sites/edp/files/publication/18-02-06_accountability_on_the_ground_part_1_en_0.pdf (last visited April 2018).

[26] Accountability on the Ground Part I: Records, Registers and When to Do Data Protection Impact Assessments, p. 7-8.

are processed, are erased or rectified without delay."[27]

This provision of the GDPR focuses on the data protection principle known as accuracy or data quality. The principle of accuracy mandates that the company not maintain inaccurate information about a data subject; such information may create a wrong image of a data subject and adversely impact his/her rights and freedoms. To the extent the company holds inaccurate information about the data subject, the erroneous personal data must be promptly removed or corrected. In doing so, the GDPR requires the company to take "every reasonable step" to rectify inaccuracies in the data.[28]

ii. Single Subject Search Using An Entity Resolved Index Addresses The Data Protection Principle Of Accuracy

Single subject search using an entity resolved index is critical to implement the GDPR's requirement of data accuracy including the differentiation between personal data belonging to the requestor and personal data concerning other data subjects. Let's say Jan Peeters contacts the company in an effort to exercise his data subject rights. The company learns that it has one listing for Jan Peeters in its payroll records and another listing for Janice Peeters in its customer database. If these two records pertain to two distinct persons, the company will create an inaccuracy in its records if it tags both of these records as belonging to the same data subject. Clearly, the error will be further exacerbated if the company releases personal data to the data subject that does not actually belong to him or her.

Utilization of a single subject search against an entity resolved index is an effective means to resolve the potential problem of identifying personal data as belonging to the wrong data subject; importantly, this potential problem can be addressed prior to the determination harming the data subject. When a company attempts to comply with the mandates of the GDPR, single subject search using an entity resolved index is a reasonable step to undertake that will dramatically reduce the likelihood of errors being added into the company's data and of the company releasing data to the wrong data subject.

C. Entity Resolution Is a Technical Solution To Address Data Confidentiality Concerns When Responding To Data Subject Requests Related To Individual Rights

Other relevant provisions are the data integrity and confidentiality requirements set forth by the GDPR. These requirements apply both to the data shared with or released to the data subject and the internal procedures for a company to ascertain what data pertains to the data subject.

i. Legal Requirements For Data Integrity And Confidentiality

Article 5 of the GDPR states that personal data must be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or

[27] Article 5(1)(d), GDPR; see Recital 39, GDPR ("…Every reasonable step should be taken to ensure that personal data which are inaccurate are rectified or deleted.").
[28] Article 5(1)(d), GDPR.

damage, using appropriate technical or organisational measures."[29]

This provision of the GDPR highlights two important data protection principles: data integrity and confidentiality. The principle of data integrity ensures that a data subject is not harmed by inaccurate or incomplete information. The principle of confidentiality mandates that information not be disclosed to unauthorized data subjects and not be subjected to unnecessary processes. This means that a company should, as a default, keep personal data within the organization and not release or reveal it to outside persons or organizations without a proper legal basis. Furthermore, personal data must also be kept confidential within the company, i.e., access should only be granted on a "need-to-know" basis (deployment of so-called "role-based access").[30]

ii. Entity Resolved Indexes And Federated Fetch Address The Data Protection Principles Of Integrity And Confidentiality

Both an entity resolved index and federated fetch implement the GDPR's requirement of data integrity and confidentiality related to a data subject's request. Entity resolution provides an important means to limit or eliminate both over and under collection of data that a particular data subject would get access to. By way of example, let's assume that the company has a listing for Jan Peeters in its payroll records and a listing for Janice Peeters in its customer database. If these are two distinct people, the company will "over-collect" if it tags both of these records as belonging to the data subject who is requesting to exercise his or her individual rights under the GDPR. Conversely, if Jan Peeters in its payroll records and a listing for Janice Peeters in its customer database are the same person, the company will "under-collect" if it does not identify both of these records as belonging to the requestor.

Within the company, compared to a centralized data lake, the searchable index and federated fetch approach dramatically limits access to the data by employees who are located outside the department where data is housed. Any search for the data subject's requested personal data is accomplished using the central searchable index so there is minimal exposure of the data itself. This system for structured and minimized access accommodates the GDPR requirements of data integrity and confidentiality.

D. Index Search With Federated Fetch Is a Technical Solution To Address Data Minimization Concerns When Responding To Data Subject Requests Related To Individual Rights

Data minimization is a concern for companies who are seeking to handle requests of data

[29] Article 5(1)(f), GDPR; see Recital 83, GDPR ("In assessing data security risk, consideration should be given to the risks that are presented by personal data processing, such as accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed which may in particular lead to physical, material or non-material damage.").

[30] Article 5(1)(f), GDPR; see Guidance on Data Portability, p. 9, 11 ("[The right to data portability] shall not adversely affect the rights and freedoms of others… This [ ] condition is intended to avoid the retrieval and transmission of data containing the personal data of other (non-consenting) data subjects to a new data controller in cases where these data are likely to be processed in a way that would adversely affect the rights and freedoms of the other data subjects. (Article 20(4) of the GDPR.)").

subjects exercising their individual rights under the GDPR. Companies are required to search all their systems, and to limit the impact of the processing on the personal data of the data subject.

i. Legal Requirements For Data Minimization Under The GDPR

Article 5 of the GDPR mandates data minimization. This term means that personal data shall be "limited to what is necessary in relation to the purposes for which they are processed."[31] In the discussion of data protection by design in Article 25, the GDPR requires "appropriate technical… measures" designed to help implement the data protection principles - such as data minimization - to safeguard the processing of personal data.[32]

During the identification of personal data that is relevant to the data subject requesting to exercise individual rights under the GDPR, companies must limit personal data collection, storage, and usage to data that is needed for carrying out this purpose.

ii. Index Search With Federated Fetch Addresses The Data Protection Principle Of Data Minimization

In order to meet the GDPR goal of data minimization, single subject search against a searchable index is the superior technology to use to locate the records of data subjects, employing data minimization techniques by design and lowering the risk for data breach.

a. Index Search with Federated Fetch's Implementation of Data Minimization Technique

Index search with Federated fetch is one technical tool for a company to handle the GDPR's requirement of data minimization related to a data subject's request. In a searchable index and federated fetch model, a record of a search for a data subject is logged in the index once.[33] This

[31] Article 5(1)(c), GDPR; see Recital 39, GDPR ("…limited to what is necessary for the purposes for which they are processed…"); Recital 78, GDPR ("… the controller should adopt internal policies and implement measures which meet in particular the principles of data protection by design and data protection by default. Such measures could consist, inter alia, of minimising the processing of personal data, …").

[32] Article 25(1), GDPR; see Lina Jasmontaite, et al., "Implementation of Data Protection by Design and by Default: Framing Guiding Principles into Applicable Rules," IPEN Workshop 2017 (June 9, 2017), available at https://edps.europa.eu/ipen-workshop-2017_en (last visited April 2018); cf. Google v. AEPD, Judgment of the ECJ (Grand Chamber), Case C-131/12, para. 58 (May 13, 2014) ("…it cannot be accepted that the processing of personal data carried out for the purposes of the operation of the search engine should escape the obligations and guarantees laid down by Directive 95/46, which would compromise the directive's effectiveness and the effective and complete protection of the fundamental rights and freedoms of natural persons which the directive seeks to ensure…"), available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0131 (last visited April 2018).

[33] See Toby Segaran & Jeff Hammerbacher, Beautiful Data: The Stories Behind Elegant Data Solutions, O'Reilly (2009), p. 113 ("Let's say that directories, indices, and catalogues are all basically the same thing: a thing used to locate other things. . . . After the user is provided a pointer, the activity becomes 'federated fetch.' Note the difference between a federated search (not useful) and a federated fetch (useful)."), available at https://books.google.com/books?id=zxNglqU1FKgC&pg=PA113&lpg=PA113&dq=federated+fetch&source=bl&ots=DEBQQ6bCkG&sig=f9dDki3x_Y586zc6JYApTAlwn1A&hl=en&sa=X&ved=0ahUKEwjVz6OTnefZAhWomeAKHZBNBkUQ6AEIYjAI#v=onepage&q=federated%20fetch&f=false (last visited April 2018); see also Rossitza Setchi & Ivan Jordanov, Knowledge-Based and Intelligent Information and Engineering Systems, 14th International Conference, KES 2010 (Sept. 2010), p. 600 ("Directories are locator services; they return references after being provided with query terms. Provisions of references turns the next series of activities to "fetch," and on the WWW that is equivalent

is sufficient to document compliance with the requirement to conduct the search for the data subject, but is also not more than is necessary to ensure data subject rights and achieve accountability.[34]

Federated fetch contrasts with federated search, a different technological approach for searching multiple databases. In federated search, the query is sent from a centralized location to each of the other databases being searched. The databases receiving the query thus log the search, which will be based on the name of the data subject and associated personal data. The databases then return plain text of all of the possibly relevant personal data to the centralized location. In federated search, therefore, each of the remote databases receives personal data from the centralized query, and the centralized location then receives personal data from remote databases. All of this personal data is typically logged both centrally and remotely, regardless of whether relevant data is found in a particular remote database. This means that if a company has 100 systems then 100 copies of the request would be created by the search.[35] Federated fetch, by contrast, involving the transfer only of pointers, minimizes data processing compared to federated search.

b. Index Search and Federated Fetch Lessens Risk of Data Breach

Along with data minimization in the central index, federated fetch lowers the risk of data breach, newly governed in the GDPR under Articles 33 and 34. If a breach does occur with the centralized directory, the breach would include only pointers (typically hashed). Compared with breaches of personal data, breaches of only hashes are far less likely to pose a risk to individuals' rights.[36]

Other approaches for responding to a data subject's request have higher risks from a data breach perspective. One approach is for a company to have all personal data held in a single, consolidated database, i.e. the "data lake" approach. In the event of a breach, all of the personal data in the centralized database could be exposed.

to 'federated fetch.'"), available at https://books.google.com/books?id=LKb0VAZKYz4C&pg=PA600&lpg=PA600&dq=federated+fetch&source=bl&ots=k6zgcmk--4&sig=NGVIrkLWiuhhdS0agMa2KzPxJX0&hl=en&sa=X&ved=0ahUKEwjVz6OTnefZAhWomeAKHZBNBkUQ6AEIYDAH#v=onepage&q=federated%20fetch&f=false (last visited April 2018).

[34] See Article 5(2), GDPR ("The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 [detailing the data protection principles] ('accountability')").

[35] It is worth mentioning that another approach which does not involve a technological tool is a manual search of the systems. This would appear to be the least attractive approach to achieve the company searches required under the GDPR. Both the federated fetch and the federated search begin with the step of organizing the systems in such a way that all of the company's data is searched and then carries out a unified search for the requested information. The manual search risks omitting one or more systems from the search as well as having multiple people searching different systems with less than consistent search terms or methods. The results of such a manual search are at best unreliable.

[36] See Article 32, GDPR; see also Guidelines on Personal Data Breach Notification Under Regulation 2016/679, p. 15-19, Article 29 Data Protection Working Party (Feb. 6, 2018) (discussion of hashed data), available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612052 (last visited April 2018).

Similarly, in an approach using federated search, each database in the company receives a full text of queries, often accompanied by additional identifying information such as address and email. Thus, a breach in any of the 100 systems that logged the query would contain personal data that might require breach notification.

For the risks from data breach, therefore, a searchable index and federated fetch reduces the scope of data lost in any breach and creates lower risk that a breach will trigger notification requirements. The data minimization that results from federated fetch thus appears to comply better than other approaches for responding to individual rights requests.

E. Risk of Not Using State-of-the-Art Technology When It Comes At Low Costs

When requiring technological solutions, the GPDR instructs companies to take into account several factors. The most notable of these are the "state of the art" and "the cost of implementation." This identical language can be found both in the data protection by design provision of Article 25 and Article 32's security of processing requirements.[37] This language reads as follows:

> "Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons …"

The GDPR particularly mandates that the "state of the art" and "the cost of implementation" should be taken into account when a company is determining the appropriate technical measures to use.[38] As presented in this White Paper, entity resolved indexes and federated fetch provide state of the art single-subject search to accomplish locating the personal data of the data subject who has requested to exercise individual rights under the GDPR. At present, options on the market allow for large companies as well as micro, small, and medium sized enterprises (SMEs) to acquire this technology for a modest price. In addition, the searchable index with federated fetch approach adds a light layer to a company's system. The technological layer is not overly burdensome and can be operated without highly technical operators.

[37] Recital 78 to the GDPR provides guidance on how "state of the art" is to be implemented, making no mention of the associated cost. This recital instructs that companies are to give "due regard to the state of the art" when selecting a product or service to process data to ensure that they are "able to fulfil their data protection obligations." Recital 78, GDPR ("When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations.").

[38] The language regarding "the state of the art" and "the cost of implementation" is found in both Article 25 and Article 32 – Security of Processing. Article 25(1), GDPR; Article 32(1), GDPR. This language is not found in Article 24 – Responsibility of the Controller. Article 24, GDPR ("Taking into account the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons …"); see Felix Bieker and Marit Hansen, "What are DPAs Expecting from Controllers to Comply with Article 25 GDPR?," IPEN Workshop (Sept. 9, 2016), available at https://edps.europa.eu/sites/edp/files/publication/16-09-09_bieker_dpbd_ipen_ffm_en.pdf (last visited April 2018).

CONCLUSION

To comply with the goals and requirements of the GDPR, companies should carefully examine their systems' abilities of responding to individual rights requests such as the right to access. As discussed here, the technology of single subject search via an entity resolved searchable index, when properly implemented, appears to be a close fit to the requirements and guidance for compliance with the GDPR. The GDPR, given its all-encompassing obligations, drives the use of technology to achieve compliance.

ALSTON & BIRD

www.alston.com