Privacy pitfalls and challenges in assessing complex data breach incidents

In the past few years we have seen how the introduced data breach notification requirements have affected organizations in dealing with large scale data breaches. When looking at some practical example cases, we have seen that organizations struggle with gathering the right information about the nature and scale of the breach, especially information about the specifically affected individuals that they may need to be notified. Most challenges arise from both a data and a legal perspective. From a data perspective, organizations struggle in getting a complete and accurate overview of all affected data in the breach and the affected individuals. From a legal perspective, organizations need to assess to what degree the breached data can pose a high risk to the affected individual. A complex risk assessment needs to be performed and documented. This article will outline these challenges in detail and will conclude with recommendations for organizations on how to properly prepare themselves.

Stephan Idema MSc LLM is Senior Manager at KPMG Risk & Regulatory.
INTRODUCTION

In May 2018, the General Data Protection Regulation (GDPR) came into effect, forcing organizations to comply with a set of legal requirements regarding data breaches. Some countries, such as The Netherlands, have already implemented a similar regulation prior to this.1

When looking into the likelihood of a data breach and the likelihood of having to undergo these data breach proceedings seem to differ per country. In a recent report from DLA Piper, we read that about 25% (40k) of all data breaches within the European Union came from the Netherlands ([DLA20]). With only about 3.3% of the total EU population, this seems out of proportion in relation to the reported breaches in other EU member states. The most sensible explanation is that the Dutch data protection authority is more active in the Netherlands and the fact that organizations already had a reporting obligation since 2016 in the Netherlands. We see that Germany is the runner-up with 37k in reported data breaches. The UK, Ireland and Finland are in 3rd, 4th and 5th place respectively.

When we zoom in on the number of data breaches being reported in the Netherlands, we see that most data breaches (around 73%) concern relative straightforward cases of sending personal information to the wrong recipient, either via e-mail or via physical post, or the loss of a laptop or other data carrier (5%). The more complex data breach cases, however, are the ones that showcase that the legal requirements as set forth in Articles 33 and 34 of the GDPR are a big challenge. Data breaches are related to hacking, malware, phishing or other data theft (4%), or personal data that has been incidentally published or a leak in the system allowing unauthorized third-party access (7%).

THE LEGAL PLAYING FIELD

There are two main articles in the GDPR that cover the data breach notification: Articles 33 and 34. According to GDPR Article 33, in case of a personal data breach the data controller should report the incident to the supervisory authority within 72 hours, describing the nature of the breach, assessing the impact of the breach and describing measures taken. Article 34 states that in case of a high risk to the rights and freedoms of natural persons, the controller will communicate the personal data breach to the data subject without undue delay. The information to the data subject should clearly contain the nature of the breach and at least information about the likely consequences of the data breach and the measures taken or proposed to mitigate possible negative effects. Failure to comply with these regulatory requirements may cause high monetary and reputational damage to the organization.

The next section will further dive into the legal requirements in relation to the more complex data breach cases. These more complex cases will show the financial burden that organizations need to undergo by putting in hours of investigation, legal analyses and decision making in order to comply with all legal requirements. The fact that the supervisory authority is closely monitoring these high-profile cases makes it more important that these requirements are met.

CHALLENGES IN ASSESSING LEGAL REQUIREMENTS

In the introduction it was laid out that about 11% of the reported data breaches contain more complex cases (or at least, here in the Netherlands). When a data breach has a malicious source it already becomes quickly evident that the breach will fall into the category of a complex case. We can say the same about cases where there is a vulnerability or a leak regarding a database or server with consumer data. These kinds of cases have taught us in the last few years that the legal requirements of the GDPR can become a heavy burden in case an organization is ill prepared. The next section will – per requirement – outline the challenges an organization may be facing and the impact this will have on resources, timelines and in the end financial or reputational damage.

REPORTING TO THE AUTHORITIES

According to Article 33 of the GDPR, the data controller should notify the supervisory authority about the data breach within 72 hours. The notification should include the nature of the breach, the categories and approximate number of data subjects involved, and the categories and approximate number of personal data records involved. Next to that, the likely consequences for the individuals should be communicated and the measures to mitigate the negative consequences.

---

1 Already pre-GDPR, The Netherlands has implemented additional Articles to the Personal Data Protection Act regarding the reporting of data breaches to the authority, as per January 1, 2012 for telecom and internet service providers and January 1, 2016 for all organizations processing personal data.
Nature of the breach

The nature of the data breach may in most cases be very clear. In cases where data has been published incidentally or a leak has (potentially) allowed unauthorized third parties to access the data, it is fairly straightforward to explain the nature of the breach. Also, in most cases of malicious intent, the nature is quite clear, when communicated in generic terms (e.g. malware attack, hacking, theft of a physical hard drive). In some cases, where personal data is being compromised and the organization learns of this fact due to external sources, the nature of the breach may not be clear. A thorough incident response investigation will be required to determine this.

Scale of the breach

More challenging than determining the nature of the breach, will be the scale of the breach. The GDPR is fortunately asking data controllers to come up with approximate numbers. This may however be hard to estimate within 72 hours after discovery of the data breach. We have seen cases where multiple systems where compromised during a malicious attack. Consumer data was stored in these systems and often contain data records of the same individual across different systems. The overlap of data subjects and categories of personal data records will make it hard to determine a set of unique data subjects and data records to report to the authorities. Especially in cases where the master data management has not been up to par with industry best practices. Organizations with a large consumer base such as banks, pension funds or insurance companies will face this challenge. We will further dive into this in “Reporting to individual data subjects”.

Likelihood of consequences and mitigating measures

The likelihood of consequences is easy to identify on a generic level. When facing a breach with more generic data such as names, e-mail addresses, phone numbers, etc., the consequences can be found in the area of phishing and scamming activities. When more data is added, more threats and adverse consequences can become apparent, such as spear phishing, extortion, or targeted theft. When assessing the extent to which individual data subjects need to be notified, the analysis of these adverse consequences is critical in complying with the regulatory requirements of a data breach. The result of the analysis and the consequential decision making should be carefully documented. The challenges that come along with this assessment will also be laid out in “Reporting to individual data subjects”.

Complex data breach cases show that GDPR requirements in Articles 33 and 34 are a major challenge

Reporting the mitigating measures may be difficult to communicate to the authority within 72 hours of discovering the breach, since it is probably still being investigated. This also applies of course to all the other reporting requirements as discussed above. In order to meet data controllers halfway, the supervisory authority can allow data controllers to send a provisional or initial notification which can be adjusted or revoked at a later stage. When submitting a provisional data breach notification with regard to a large scale data breach, it is advisable to seek contact with the data protection authority and keep close communication with them with regards to the data breach. A data breach in itself is not (necessarily) a violation of the GDPR. Not handling a data breach in line with the GDPR requirements or without following up instructions from the data protection authority, however, is.

REPORTING TO INDIVIDUAL DATA SUBJECTS

To get a good understanding of the realistic challenges of an organization when it comes to the reporting requirements of the GDPR, we must picture large-scale data breaches, mostly from a malicious external source (such as a hack, data theft or malware related incident). For example, the data breaches of British Airways,

2 The Dutch Data Protection Authority allows data controllers to submit an initial data breach notification which can be revised afterwards.

3 British Airways was fined GBP 183 million, because credit card information, names, e-mail addresses were stolen by hackers, who diverted users of the British Airways website to a fraudulent website to gather the personal information of the data subject.
4 In March of 2020, the e-mail vendor of T-Mobile was hacked, giving unauthorized access to e-mail data and therefore personal information of T-Mobile customers. In 2018, unauthorized users also hacked into the systems of T-Mobile to steal personal data.

5 Equifax systems were compromised through a hack in the consumer web portal in 2017. Personal data of over one hundred million people were stolen. Personal data containing names, addresses, date of birth and social security numbers.

6 In 2018 and again in 2020, Marriott reported that their reservation system had been compromised. Passport and credit card number of 500 resp. 5 million customers were stolen.

### Table 1. Criteria to be considered when assessing the likelihood of a high-risk impact for the individual.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Explanation and further guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Type of Breach</td>
<td>For example, unauthorized access, unauthorized disclosure or loss of data. These types will have a different risk profile attached to them.</td>
</tr>
<tr>
<td>B. Nature, sensitivity and volume of personal data</td>
<td>The nature and sensitivity are already defined by the GDPR, but also personal data that may seem innocuous at first. When combined with other factors or data it may pose more serious risks to individuals, however.</td>
</tr>
<tr>
<td>C. Ease of identification of individuals</td>
<td>How easy would it be to identify the individual, based on the data (and with the help of other external sources)</td>
</tr>
<tr>
<td>D. Severity of consequences of individuals</td>
<td>What can someone with malicious intent do with the personal data? How harmful can the data be to the individual?</td>
</tr>
<tr>
<td>E. Special characteristics of the individuals</td>
<td>The breach may concern children or minority groups that are placed in a greater risk when the data is breached.</td>
</tr>
<tr>
<td>F - Special characteristics of the data controller</td>
<td>The nature of the data controller may also indicate something about the data subject. For example, if it is a mental health institution or a political party, which would tie the individual to special categories of personal data.</td>
</tr>
<tr>
<td>G. Number of affected individuals</td>
<td>The total number of affected individuals.</td>
</tr>
<tr>
<td>H. General points.</td>
<td>Other relevant higher risks.</td>
</tr>
</tbody>
</table>

Including the potential adverse consequences and what precautionary measures can be taken by the individual. The European Data Protection Board has provided guidelines about the criteria that should be considered when assessing the likelihood of a high-risk impact for the individual (see Table 1, [WP2916]).

When assessing whether or not an individual is exposed to a high-risk adverse event regarding their rights and freedoms, a data controller needs to look at each single individual and determine whether or not they should be notified as well as the content of the notification. From a data perspective, this can be a humongous challenge, especially when the data management practices are not up to par. Determining the level of risk for each data point can also take significant effort, especially when the data has aged. These challenges will be addressed in the upcoming section.

### Challenges from a data perspective

When looking into some specific data breach cases of the last two years, we have seen many challenges from a data perspective, for which a few examples will be given in this section. When the data controller is aware of which systems are compromised, the next step is to determine which personal data was stored in these information systems and what this data was about. When the level of maturity of data management practices within the compromised organizations is poor, then assessing the data will probably be the most challenging task in adhering to GDPR data breach reporting requirements.

Let’s focus on the idea that multiple systems containing consumer data have been compromised and the data quality / management within the organization is not of the highest standards. The following examples will provide master data challenges to determine for each impacted individual whether or not they should be notified.

For example, we will look into the fictional individual “Adam Smith”, who is a customer of an airline company. His personal information is in the customer master database, booking database, the payment database and in the Event database, where tickets for troubleshooting are stored. In these three systems, his personal data shows up as shown in Figure 1 (these are exaggerated examples).

When we want to identify Adam Smith and check the personal data that is being kept of him which has been compromised during the breach, we see different data from different systems. Even data from the same system may indicate different information.
Challenge 1: Aggregate all the data of the same individual “Adam Smith” into one single overview of all his personal data
Since the individual “Adam Smith” is present in four different systems and is even found twice in the master database, we need to aggregate the data to understand which personal data is stored and what risks he may be exposed to across different systems. Because Adam is registered differently across systems and each system contains different categories of data, there is no single unique identifier to use and identify one single Adam Smith. Ideally, each information system would have a reference to the same unique customer number. In practice, we have seen that this is not always the case and will be the root cause of complex data analytics to create a full picture of every single individual and their corresponding personal data records.

Challenge 2: Which “Adam Smith” data is the most recent data?
We may need to extract additional data from the systems to determine the registration date or last mutation date of the provided data record. Amending source data with additional data to determine the relevant records is going to be an additional layer of complexity to identify the unique users which should be notified under the GDPR.

Challenge 3: What is the age of the data records that is being shown?
A lot of companies struggle with the implementation of proper data retention procedures and controls. As a consequence, a lot of old data is still being stored in the information systems. In order to determine the relevance of this data, one needs a timestamp of when the data entered the system or when it was modified. Some data fields lose their relevance. For example, someone’s home address, telephone number or credit card may change in the course of 10 years. The same applies to someone’s license plate or IP address. These may be irrelevant already within 3 years’ time. Someone’s social security number or medical records are however never subject to change.

Challenges from a legal perspective
In case it is identified what categories of data have been compromised, it would be best to create an overview of all different profiles and tie risk classifications to each profile and determine whether or not these individuals will be personally notified or not (of possibly by what medium). Such a matrix will look like this: (insert example)

To fill in the risk profiles and determine whether or not each individual should be notified, the following activities can be performed:

- Assess the value of the data (considering the nature of the breach)
- Assess the potential risks for the individuals
- Determine the impact of the age of the data (also, ‘actuality’, ‘accuracy’, or ‘timeliness’ of data)

Figure 1. Data challenge examples.
Assess the value of the data

The value of the breached data can play a pivotal role in assessing potential threats. With regards to this whitepaper, the value of data is defined as what people with malicious intent could profit by potentially selling the information they have gained from your systems. Looking into the black market or “dark web” is a good starting point to assess the value. The value of personal information on the black market depends on the type of information and the combination of data available on the same individual. After doing some initial research, we see that the price of personal data varies quite a lot. Some basic information about individuals may provide you with a few dollars per record but adding bank account or credit card data to that will increase the value significantly. In the Netherlands, too, we have seen cases of selling personal data in combination with license plates is quite valuable. Combinations of personal and health information is two or three times the value of financial information alone as there are many more opportunities for fraud or blackmail of wealthy customers. For some example price ranges, please refer to Table 2.

Assess the potential risks for individuals

The second step in the assessment is to identify the potential risks for each individual. The potential risks of the breached information can be categorized under (at least) three threat scenarios: Identity Theft, Scamming and Leaking/Blackmailing. We will provide examples for each threat scenario below.

Identity theft

Customer information can be used to impersonate a customer or employee.

• Acquiring funds or goods. Identity theft can be a tool to commit fraud to acquire funds or goods. The severity and impact to the individual in this is high, because the customers can suffer from financial losses, unless they can prove they are the victim of identity theft. An attacker could for example order subscriptions or goods online based on the personal information of the victim.

• Framing for (illegal) activities. Identity theft can also be used to ensure other illegal activities cannot be traced back to the person that committed them. An attacker could scam people on online platforms, such as “Marktplaats” (online Dutch marketplace and subsidiary of eBay), while impersonating a victim of the data breach. The stolen personal data is used to convince the person that is being scammed of the legitimacy of the scammer. As a result, the victims of the identity theft may be harassed by the victims of the scam ([App18]).

• Acquiring more personal information. An attacker can also contact organizations and impersonate a customer in order to obtain additional personal information about the customer. The attacker can directly request insight into the personal information kept by the organization based on GDPR, or ask questions to deduce personal information that the organization has of the victim. The information from the breach is used by the attacker to initially identify as a customer. Obtaining additional personal information is not the end goal, but a means of achieving another goal in one of the three categories. An attacker could for example attempt to obtain the document number for an ID card or driver’s license, which can then be used to create a fake digital copy of such as document. This can then be used in other identity theft schemes that require a copy of such a document, such as renting buildings ([Sama16]).

Scamming

The stolen information can be used in different ways to scam the customer whose data has been stolen.

Table 2. Market worth of privacy data.

<table>
<thead>
<tr>
<th>Data</th>
<th>Worth per record (approximately)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name, address, date of birth and Social Security Number</td>
<td>$0.50-$7 ([Stac17], [TrMi15], [CyRe18])</td>
</tr>
<tr>
<td>Name, address, Social Security Number (SSN) and one or more pieces of banking information*</td>
<td>$15 ([Secu16])</td>
</tr>
<tr>
<td>Personally identifiable information (name, address, phone number, SSN, date of birth (DoB), bank account data etc.)*</td>
<td>$40-$200 ([Armo18])</td>
</tr>
<tr>
<td>License plate information with owner name, address</td>
<td>$55-$110 ([Hofm19])</td>
</tr>
<tr>
<td>Medical records**</td>
<td>$0.03 ([Sama16]) - $10 ([Hume14]) - $250-$1,000 ([DHHS19])</td>
</tr>
</tbody>
</table>

* Banking information refers to more information than just the account number.

** The amounts for medical records vary hugely. Some reports come up with numbers up to $1,000, which mainly refers to complete medical records/files ([DHHS19]), other sources such as Intel only claim that a single record is worth $0.03 to $2.42.
Generic scams. General contact information can be used to send spam, perform phishing attempts and attempt other generic scams. The impact of such generic scams on the victims depends on the success rate of the scams.

Tailored scams. Personal information, such as age and medical information, can be used to perform more tailored scams or target weaker groups. For example, older people generally have less digital experience, making them an easier target and chronically ill people are generally more willing to try new things to improve their health. Again, these techniques can be used to obtain money or credentials from the victims.

“Spear” scamming. More personal information, such as a BSN and medical information, can be used to attempt to convince the customer that the attacker is from an organization where the victim is registered or from an authority such as the police. The attacker achieves this by providing the victim with information about them, which should generally only be known to such organizations. Providing personal information about the victim increases the credibility of e-mails, letters and other interaction with the victim. Social engineering techniques can for example be used to trick people into transferring money or providing login credentials for online accounts.

Leaking or blackmailing
The stolen information can be leaked, or the victim can be blackmailed for money or other gains.

Sensitive information. In case of available personal information, such as medical information of high-profile individuals, like celebrities and politicians, the stolen information can be leaked, or the individual can be blackmailed with the threat of leaking the information.

Threatened identity theft. Victims can also be blackmailed with the threat of identity theft. This would have a high impact on them. The leaking of information can result in reputational damages for the victim, whereas blackmailing can result in either financial damage or reputational damage.

Determining the impact of the age of the data
Regulators, legal cases or current black market prices unfortunately do not tell us anything about the relevancy of the age of the data records. It is important to assess to what extent leaked personal data that is relatively old, can still impact the individual and to what extent the notification can help them to take steps to protect themselves from the effects of the breach. Even though leaked personal data is relatively old, the risk still exists that the breach may lead to physical, material or non-material damage for these individuals.
This is especially applicable to cases where sensitive personal data is leaked, such as health data. The question that needs to be asked with regard to the impact of the breach for relatively old personal data is: could the breach still result in identity theft or fraud, physical harm, psychological distress, humiliation or damage to reputation of the individual? ([WP2916]).

To help make this assessment, some general statistics can provide some guidance (see Table 3).

**Document decisions and communication**

When the assessment concerning the risks for the individuals subject to the data breach has been completed, a communication matrix can be created to determine the risk level for each case (personal data types and data age) and if this risk level meets the threshold of ‘high risk’ as stated in Article 34 of the GDPR. It is very important to document and substantiate the assigned risk level and why it is – according to your analysis – below or above the indicated threshold of Article 34 of the GDPR. This will be your core rationale of why you are going to notify an individual or not.

When the data is prepared and the legal analysis and risk analysis have been completed, a communication scheme can be set up. There are different methods for reaching out to individual data subjects. This can be done by physical mail, e-mail, SMS or even by telephone. Depending on the available contact information and the efficiency, a decision can be made.

When sending out large communications to individual data subjects, one can expect to receive some sort of a response from those individuals. The individuals may have questions about the data loss, they may want to execute their privacy rights, or they may want to have their data deleted. It is highly recommended to anticipate these scenarios by setting up a call center to answer questions and gather subject requests, set up a specific e-mail address to gather subject requests and complaints and more importantly, reserve resources to follow up on an expected peak of access and deletion requests under the GDPR.

**CONCLUSION AND HOW TO PREPARE**

When a more complex and/or large-scale data breach occurs, an organization is under heavy stress and pressure, regardless of the legal requirements as set forth by the GDPR. Acknowledging this will help organizations understand why it is critical to assess your internal procedures, data management maturity and incident response capabilities thoroughly. Then one will have a decent understanding to what degree these challenges can be resolved effectively and efficiently in a timely manner. When we look at the root causes for the delays and challenges to adhere to the legal requirements that come with a data breach, it is recommended to assess how prepared you are, on the following topics:

- (Master) Data Management: What is the quality of your master data and is your organization able to create insights on a person level (rather than a product or process level), which customer information is stored?
- Data Retention: Which records of your customers are you keeping and how are your data retention policies carried out in practice? Do you have insights in the age of your data and whether or not you should still have this data of your customers?

---

**Table 3. Statistics on data age.**

<table>
<thead>
<tr>
<th>Data</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>The average period of an individual living at the same address</td>
<td>10 years ([Appe18])*</td>
</tr>
<tr>
<td>The average period of an individual having the same telephone number</td>
<td>Unknown</td>
</tr>
<tr>
<td>The average period of an individual owning the same car</td>
<td>4,1 year ([WP2916])*</td>
</tr>
<tr>
<td>The average period of an individual having the same bank account no.</td>
<td>Unknown**</td>
</tr>
<tr>
<td>Validity of a Dutch passport for individuals under the age of 18</td>
<td>5 years</td>
</tr>
<tr>
<td>Validity of a Dutch passport for individuals of 18 years or older</td>
<td>10 years</td>
</tr>
<tr>
<td>Social Security Number</td>
<td>As long as the living individual is a Dutch resident.</td>
</tr>
<tr>
<td>Medical diagnosis</td>
<td>Depending on the condition, this may affect the duration of the life of the individual, and in case of hereditary diagnoses: possibly the life of a family.</td>
</tr>
</tbody>
</table>

* The statistics differ per age category. The general tendency is that younger people change their address or car more often than older people. The statistics were published by the Central Bureau of Statistics (CBS) of The Netherlands.

** This statistic is not available, in 2018 however, 91,000 people changed banks and in 2017, 67,000 people changed banks in the Netherlands. The statistics were published by the central bureau of statistics of the Netherlands.
• Data Minimization: Which records are you keeping of your customers? Are there additional records being kept (for example in open text fields or document upload features) that should not be stored and retained of these individuals, according to your policies?
• Do you have proper contact details of your customers? Are you able to contact them in an efficient manner and is this contact information up to date?
• Do you have a data breach procedure? Are you testing or evaluating this procedure and is it robust enough to handle more complex data breach incidents? This may include a crisis management plan and follow-up communication plans.
• Is your data encrypted at rest and at transit? Are you using encryption techniques that are robust enough to prevent unauthorized access to (potentially) leaked or stolen data?
• Do you have insight in what data you are processing for third parties and can you isolate this data from the data for which you are data controller? What are the liabilities in the data processing agreement between you and the third party of whom you are processing data?
• You can offer a credit monitoring service for victims of a data breach, to monitor whether or not identity theft has taken place.
• Do you have a cyber insurance to cover incidents like these?

The above topics are certainly not a limitative set of steps that need to be taken, but merely a guiding set of questions you might want to ask yourself when preparing for a data breach.

It can be concluded that no data breach is the same and that every case has its unique characteristics, but when handling large sets of data and applying the same legal requirements, the challenges will be of similar nature and can be properly prepared.

References

About the author
Stephan Idema MSc LLM is Senior Manager at KPMG Risk & Regulatory and focuses on a wide variety of complex legal and technical matters, such as data privacy, intellectual property and software licensing. He is the head of KPMG’s data privacy team. Stephan has an academic background in the field of information studies and Law. He graduated in 2009 as a Master of Science in Business Information Systems at the University of Amsterdam. In 2011 he finished his Master of Law with a focus on IT and privacy law, at the University of Groningen. In 2016, Stephan finished his Executive Master of IT Audit at the University of Amsterdam.