On 25 May 2018 the GDPR will be in full effect within the European Union. Organizations are struggling to implement all the GDPR requirements in a timely manner and are finding themselves in a swamp of technical and data-driven challenges that come along with the GDPR. Organizations are for example required to document where all their personal data is stored, what it is processed for and who it is being transferred to. The level of detail that is required is very hard to produce in a short timeframe and without the proper tools nearly impossible. Indica, in collaboration with KPMG, has developed a tool that can help organizations overcome a large part of the technical challenges that they are faced with when addressing these technical and data-driven GDPR requirements.

S.X. Idema MSc LLM
is a Manager at KPMG.
idema.stephan@kpmg.nl

Data driven challenges of the General Data Protection Regulation

A sustainable and automated solution

INTRODUCTION

In 1995 we were living quite differently. We ordered our products from big paper catalogues, television commercials or through the telephone. We arranged our bank transactions by actually going to a bank. We went to the post office to send a letter or a postcard. To sell our own personnel things, we used the bulletin board at the local grocery store. Life in 1995 was quite different than life as we know it in 2018. Why outline 1995 you might think? In 1995 the European Directive 95/46 was ratified by the European Commission and laid the foundation for data protection within the European Union. To date this is still the foundation of many locally implemented privacy laws within the European Union. This piece of legislation was drafted and ratified in a time period where Facebook and Google did not yet exist and less than one percent of the world population had access to the internet, which only contained very basic information.

During the last twenty years a lot has changed in the way we communicate and in how we do business. The internet, mobile telephony and computers have developed substantially over the past two decades, providing us more and more possibilities in every way imaginable. These aforementioned developments have also had a huge impact on the amount of data that it is now being processed. This data also includes a lot of personal information that has been captured during those two decades. And the further we have integrated our life with the digital highway, the more personal data has become available for (commercial) organizations. The legislation that was drafted in 1995 did not foresee such a rapid change in our society and therefore a change was obviously required in order to protect European citizens. This updated piece of legislation has now arrived and is known as the General Data Protection Regulation.

GDPR AT A GLANCE

On 16 April 2016 the European Commission ratified the long awaited General Data Protection Regulation. The regulation was ratified after four years of designing, discussing and negotiating its contents and applies to all organizations that are processing personal information. The new legislation came into effect as of 16 April 2016, but will be enforced from 25 May 2018. The legislation will cause quite some changes with regard to the current privacy legislation that was implemented based on the 1995 directive.

The main changes involve two general aspects: firstly, create more accountability for the organizations that are processing personal information and secondly, put more control in the hands of the data subjects - the individuals of whom the data is being processed.

Accountability

The European Commission wants organizations to be able to demonstrate their level of compliance with privacy regulations. This means that organizations must be able to show that they have control over their processing of personal data. A few examples in the regulation are the data processing inventory, executing privacy impact assessments and appointing a formal data privacy officer. The regulation will also force organizations to have legal contracts with their third-party data processors. Organization must implement these processes and activities before 25 May 2018.

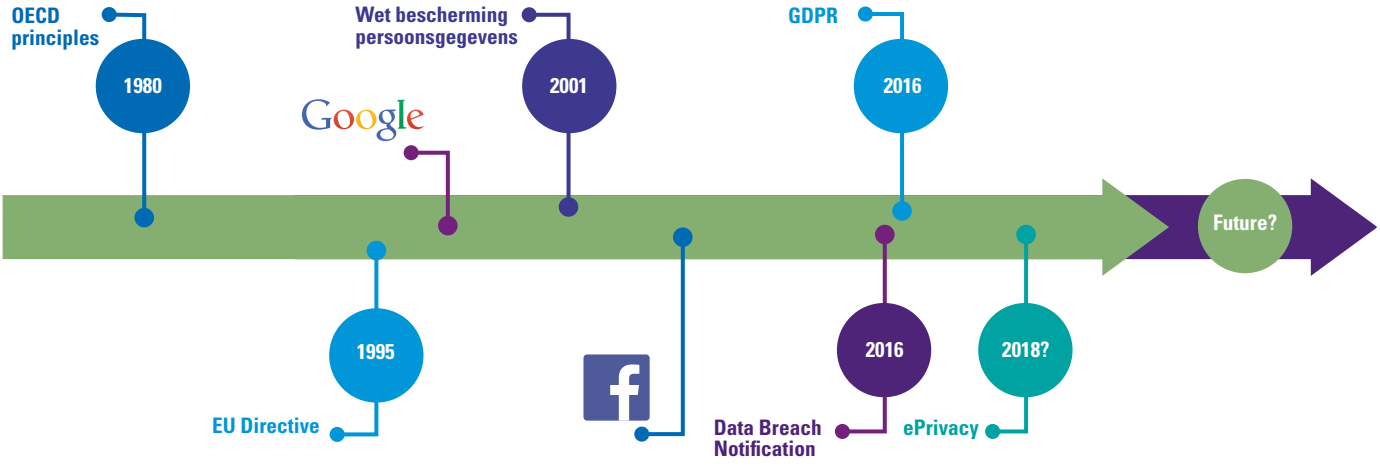
Data Subject Rights

Along with the accountability, the European Commission wanted to give data subjects more control over their personal data. This can be achieved by providing them with more tools and rights that they can execute against organizations that process their personal information. The 1995 directive already had a few of these rights embedded in the legal structure, such as the right to access and the right to correct personal information. The GDPR adds a few additional rights for the data subject, such as the right to erasure, in case the subject wants all his personal data removed and the right to data portability, to name but two.

Most new legislative requirements are administrative activities or process-related activities that organizations need to implement before 25 May 2018. Implementing such processes and activities may be a time-consuming process, but they are not necessarily complicated or of a sophisticated nature and therefore not difficult to imple-

During the last 20 years technology has drastically changed the way we communicate and do business

Figure 1. Timeline regarding data privacy regulations.



ment within the organization. There are however also a few requirements in the new legislation, especially for larger organizations, that require a complicated data driven exercise to become fully compliant with. These requirements include for example the data processing inventory (art. 30 GDPR), and being able to comply with data-subject rights (art. 15 till 19 GDPR). Both of these challenges have one thing in common, which is that they both require a deep understanding of and control over the (personal) data within the systems and applications in the organization, let alone unstructured data files.

GDPR CHALLENGES FROM A DATA PERSPECTIVE

Data privacy is about being in control of your personal data processing activities. If you do not have insight into which personal data is being processed within the organization, where this processing is taking place, where the data is stored and for what reason this data is being processed, then it will be hard for a company to demonstrate that it is in control of its personal data processing. Without such insight, it will also be quite a challenge to assess privacy risks within the organization and to control them accordingly.

To answer the above challenge, the European Commission has implemented article 30 in the GDPR. Article 30 obliges organizations to implement a processing inventory for personal data.

Building such an inventory from scratch will be quite a challenge for organizations that have never initiated such a process in the first place. The data processing inventory applies to both structured and unstructured data sources that are being used in the business processes. Detailed information about data processing may be readily available at hand for some key systems or applications within the organization, but this may not be the case for smaller applications or less frequently used systems, let alone for unstructured data sources on file servers or in e-mail attachments.

Article 30 of the GDPR states:

1. Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility. That record shall contain all of the following information:
   - **a.** Name and Contact Details of the Controller;
   - **b.** Purpose of data processing;
   - **c.** Categories of Data Subjects and Categories of Personal Data;
   - **d.** Categories of recipients to whom personal data will be disclosed;
   - **e.** Data transfers to third countries or international organization;
   - **f.** (Optional) – Data retention schemes for the erasure of the data;
   - **g.** (Optional) – A general description of taken measures and safeguards with regards to the data protection.

Structured data sources

Structures data sources such as the HRM, Finance, and CRM systems are the main key information systems containing personal data that will have to be documented in the data processing inventory. The challenge however is to get a complete overview of all applications and systems within the organization that process personal information. Then the associated processes that use this data will have to be identified to determine the purpose of processing and if any third parties are involved in this process. Gathering this information can be quite an extensive task if this has never been performed before. Things will become even more challenging if the various information systems interface with each other and transfer or duplicate data. These flows also have to be documented in the data inventory.

Unstructured data

Creating an inventory from structured data seems to be quite a challenge, but once all data fields and attributes are identified, the full dataset should be covered. For unstructured data this is not quite the case, since the data is obviously not structured in a database or table format. Personal data could be scattered all over the document and also in open text data fields.

Unstructured data are documents such as PDFs, Excel files, lists of names, etc. These files can be stored on a file server, can reside in e-mail attachments or are attachments in CRM systems or other systems HRM. Open text fields in a database can also constitute unstructured data. Some of this data is automatically 'searchable', some data, such as PDF documents are sometimes not. This makes it more complicated to determine if any personal information is involved or not.

Data Subject Rights

One of the two key pillars of the GDPR, given the continuous growth in processing of personal data, is to protect EU citizens against the further processing of their personal information and that their personal data is handled with due care. To enforce such a due care treatment, the EU wants to give the data subjects more tools or legal grounds to gain more control over their personal data. This control has been given a place in the GDPR under articles 15 to 19, where the data-subject rights have been documented. There are several different rights that a data subject can exercise, which have been summarized in the text box.

The right to access and the right to erasure might become quite a challenge for organizations who process personal data in a lot of different systems. A data inventory may help in identifying which systems are actually relevant with respect to extracting or removing data, but identifying the exact records of the data subject who is making the request is another story. It is also challenging to demonstrate that a complete set of records of this data subject have been given access to or have been removed within the organization.

Another problematic issue is that the GDPR also obliges an organization to respond to the request in a timely manner without delay. In practice this means that the request should be handled within approximately four weeks. When an organization only gets a handful of such requests during a year it might still be very possible to answer these in a timely manner, but probably very inefficiently. But in some sectors where a lot of consumer data is being processed for marketing related activities, it may become a trend that more and more data subjects will execute their right of erasure or access to their data within your organization. When the requests pile up, it will become quite a challenge to adhere to the timelines specified by the GDPR. This will then certainly become a big risk, since not adhering to subject-access rights will be a violation of GDPR compliance that will fall in category 2 of the sanctions provisions of the GDPR.

Overview of data subject rights

- *right to access:* the data subject has the right to gain insight into what data is being processed by the personal data controller;
- *right to rectification:* the data subject has the right to correct his personal data in case there is an error;
- *right to erasure:* the data subject has the right to have all his personal data removed;
- *right of data portability:* the data subject has the right to transport his data to another service provider;
- *right to restriction of processing:* the data subject can demand restrictions on the processing of his personal data;
- *right to object:* the data subject has the right to subject against automated processing of personal data.

INDICA GDPR – A TOOL FOR DATA DRIVEN GDPR CHALLENGES

As identified in the previous paragraph, organizations will face two key data driven challenges when implementing the GDPR: subject access rights and data inventory requirements. Life would be a lot easier if there would be continuous insight into all data that is being stored within the organization. It would be even better if this stored data would be easily searchable so that relevant documents can be found with little to no effort.

Indica eSearch

Indica eSearch is an enterprise search tool that enables organizations to index their data, documents and other digital files and enables them to search through the data pile effectively and efficiently. This search engine and tool that has been built to answer the demand for business search applications formed the basis for developing the Indica GDPR module.

Indica eSearch is an agentless tool that can be installed in one day. Indica runs either in the cloud or on a standalone virtual machine in the IT infrastructure. Indica gets read-only access to the data by giving it access as a user on the database, active directory or file server directly. The Indica account will then read the data it has access to and create an index of the identified data. This includes standard databases as part of ERP systems, CRM systems or other applications that run on a database layer. Indica can also index office files such as Word, Access, Excel, etc. PDF files can also be read and if required they can be OCRed to make them 'readable'. After the files have been indexed, the Indica search algorithm enables users to find their documents by using a key-word search.

The Indica eSearch forms the foundation for the GDPR Module. The GDPR Module uses the Indica indexing technology and search algorithm to identify Personal Identifiable Information (PII). By using logical expressions, artificial intelligence and intelligent search strings, the Indica GDPR module is capable of identifying numerous PII data attributes in the indexed data.

Indica for Data Subject Access Rights

When enterprise data is searchable and PII attributes can be identified, a tool such as Indica is going to greatly assist organizations in dealing with data subject requests from individuals. The subject access rights that have been discussed previously and are deemed to have a strong data-driven component are the right to access and the right to be forgotten (as outlined in the text box).

Simply typing in the name of the individual that is requesting information or is requesting deletion of his information is enough to provide the request handler with all the data that is stored about that individual within the organization. All documents, records and e-mails related to that individual will be shown in the Indica interface, because Indica will link the identified person to all the other records that belong to that person. The handler can extract the document and record overview and share them with the system owners and/or privacy officer to start gathering the information or start deleting the information.

About Indica and the partnership with KPMG

Indica was founded in 2013 by a small group of IT professionals who were looking for an answer to customer needs with regard to the exponentially growing data volumes and the challenges that come along with it. They decided to develop a tool to address these challenges and built a unique correlation algorithm which allows organizations to manage and index their data sources and enables them to locate their knowledge and information swiftly. This patented technology became the core of the future product which was highly appreciated on the market. Indica soon grew into a team of professionals with great competence in the areas of IT, data science, law, risk and compliance, and economics.

In 2014 KPMG Netherlands recognized the potential of Indica as an eDiscovery and Compliance tool, Indica and KPMG Netherlands established a sales and technological partnership and started jointly to develop eDiscovery and Risk & Compliance solutions. KPMG and Indica developed the Indica GDPR module as part of this partnership.

Indica for Data Inventory

The article 30 requirement for a data inventory is another data-driven challenge that comes from the GDPR. Indica is capable of identifying PII data in the indexed data sources of an organization. Indica also categorizes this data in the related PII attribute, such as name, telephone number, bank account number, social security number, credit number, and so on. With this algorithm, Indica will tell you for each information system that is connected to Indica, what personal information is stored, how much data is stored and who has access to this data. This information provides the data privacy officer with the basis for an art. 30 data inventory. Indica will provide the DPO with which PII data is being processed for each information system and can document the findings in the data inventory. The next step is to have a business owner of the information system verify the information and make sure the inventory is accurate and complete. Indica will not create the inventory autonomously but will provide the person responsible for the register with the information that he requires to set up such a register.

Apart from data processing activities that are captured in information systems, it is also possible that data processing activities are stored in unstructured documents, such as PDF files, Excel lists or other unstructured data formats. These data types pose a greater risk and challenge for the DPO, because there is usually a far lower level of control on these types of files. Indica can index fileservers or Sharepoint sites to identify PII data from these sources and can help the DPO to determine whether or not these identified documents should be reported in the data inventory, because they are an integral part of a process, or that these documents need to be removed from the file source.

Indica as a DPO Risk Management Tool / Monitoring Tool

Apart from providing an indexing mechanism, a search algorithm and a PII identification algorithm, the GDPR Module of Indica also has an advanced dashboard and workflow management system. These additional tools enable a data privacy officer to monitor the compliance of company data with GDPR requirements, to mitigate new potential data risks and to identify and reject false positives. In the example case below we will illustrate the added value of the Indica dashboard and workflow system. See Figure 2 for an overview of the Indica Dashboard.

Below I will describe a real-life case where Indica has proven it's added value with regards to the detection of GDPR risks for a company in the financial sector.

The GDPR exposes organizations to several new risks and challenges with regard to the management of their (personal) data

Indica real-life case

A medium sized organization, operating in the financial sector requested a proof of concept for Indica to identify potential risks in their unstructured data sources with regard to GDPR. Indica, together with KPMG set up an environment with the client to connect the data sources to the Indica GDPR platform. Indica indexed the data sources and identified personal information by using both pre-programmed PII tokens as well as industry specific keywords that were defined by KPMG.

In total approximately 100 gigabytes of data was indexed with Indica, containing approximately 400,000 files / data records. Of these 400,000 records of data, about 50% contained some personal information. Along with names and telephone numbers, more sensitive data was also discovered, such as medical records, passport data and even information about sexual preferences.

Of the approximately 200,000 files with PII, only 1200 records required further validations. After validation by a KPMG privacy professional, a total of 1090 files were deemed to be non-compliant with GDPR regulations, because no legal grounds for processing this information was present. The client has been advised on how to mitigate these findings most effectively by KPMG privacy professionals.

The client is currently working on better work instructions and the awareness of their employees with regards to privacy compliance.

CONCLUDING REMARKS

The GDPR exposes organizations to several new risks and challenges with regard to the management of their (personal) data, obviously with good reason. The volume of data keeps growing year after year and this creates more layers of complexity because organizations also want to do more with this data. A regulation such as the GDPR will now force organizations to gain more control over the processing of this data. Without any automated means it will be quite impossible to manage all these data flows, structures and archives. Indica provides such automated means for organizations that are having difficulty in achieving a good level of control over their (personal) data. Going forward such tooling will be a prerequisite in order to be able to demonstrate to the authorities that you are in fact in control of your data, and that you are fully aware of what data is being processed for what purpose. GDPR compliance will no longer be a case of 'tell me', but more a case of 'show me'. A tool as Indica will enable you to show control and compliance of your data with regard to GDPR. Of course, Indica will not provide a solution for all your GDPR challenges, but will surely enable you to tackle most of the practical implications with regard to personal information and the compliance thereof.

About the author

Stephan Idema MSc LLM is a Manager at KPMG Risk & Regulatory. He focuses on a wide variety of engagements involving complex legal and technology issues. These are predominantly engagements and projects regarding data privacy, intellectual property and software licensing. Stephan has an academic background in the fields 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 the 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.

Figure 2. One of the Indica GDPR dashboards that will help privacy officers to mitigate privacy risks regarding unstructured data.