DOCAUTHORITY

White Paper

Guide To Intelligent GDPR Compliance

Executive Summary

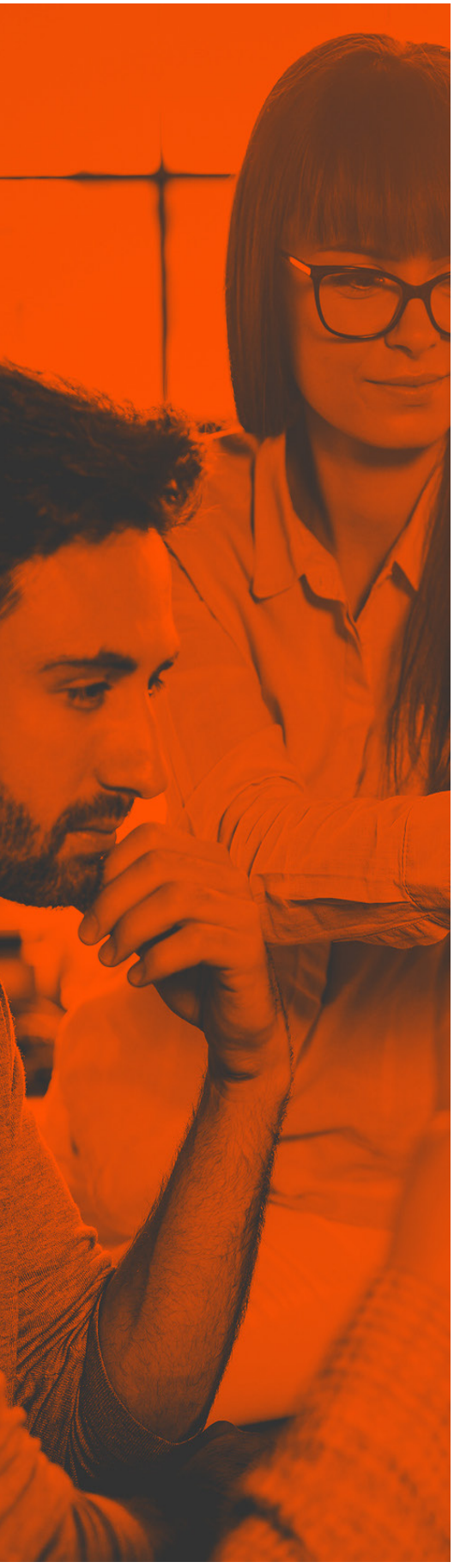
Organizations around the world are scrambling to understand and comply with the new European Union (EU) General Data Protection Regulation (GDPR) which impacts not only EU companies, but anyone who handles the data of EU citizens. The regulation's broader definition of what constitutes personal data poses a challenge for the large volumes of unstructured data that many organizations handle. How do organizations effectively find and protect this broader class of personal data in the often-chaotic proliferation of unstructured data?

DocAuthority's artificial intelligence-based engine functions as a human would, but at warp speed, to make sense of the morass of unstructured data, discover the data subject to GDPR, amongst other compliance regulations, and provide a structure for an effective operational response. Classify, mitigate, and remediate hundreds or thousands of files with a single click. Additionally, the system will automatically apply to new data created for "evergreen" compliance. With DocAuthority, you don't need an army of data management professionals to comply with GDPR.

Introduction

The GDPR will fundamentally change the way private or personal data is stored, processed, and transmitted. Since the GDPR applies to both EU-based organizations and any organization that stores, processes and/or transmits the personal data of EU citizens, regardless of their geographical location, it will apply to most organizations across the globe. Non-European companies handling EU personal data will likely be the most affected, since they may not have needed to implement comparable levels of privacy controls in their business practices.

Non-compliance has the potential to lead to huge fines - the greater of 20 million Euros or 4% of the corporation's total annual worldwide revenue. EU supervisory authorities (SAs), like the European Data Protection Supervisor (EDPS) and the Information Commissioners Office (ICO), are responsible for investigating non-compliance and administering the fines. The supervisory authority leading the investigation may be dependent on the location of the data controller, or the data handling practices of either the data controller or processor. The GDPR directs organizations to understand what and where personal data is stored and how it's processed and ultimately transmitted.



The regulation has a particularly broad view of what is considered personal data:

- Personal Data includes obvious categories such as full name, home address, email address, national identification number, passport number, vehicle registration license number, credit card numbers, date of birth, birthplace, telephone number, etc. It also includes less obvious data types such as IP addresses, application user IDs, global positioning system (GPS) location data, Internet related 'cookies', media access control (MAC) addresses, unique mobile device identifiers (UDID), international mobile equipment IDs (IMEI), etc.

- Sensitive Personal Data includes personal data that contains ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, and data concerning health.

The GDPR also introduces directives such as enabling any EU citizen to submit a request to any organization for his/her personal information to be disclosed, updated, and/or deleted within one month of receiving the request. It is therefore extremely important for organizations to be able to quickly and easily identify and find personal data that they store, process, or transmit.

The volume of unstructured data in most organizations, now potentially subject to the GDPR, presents an enormous capacity challenge. Personal data and sensitive personal data can exist anywhere in unstructured data. And the nature of data flows within an organization is dynamic, making routine identification and classification exceptionally difficult. Many organizations will struggle with this requirement. What is needed is a solution that can quickly and accurately find such data, organize and classify it, and enable easy management and remediation with existing resources.

DocAuthority offers a solution to discover information in most languages, with no requirement to preprogram keywords or regular expressions, that employs artificial intelligence (AI) to automate the discovery, identification, and classification of sensitive unstructured data assets. DocAuthority's AI can infer the business context and purpose of files and can group them accordingly. Grouping by business context and purpose allows data owners or IT to classify hundreds or thousands of files with just one click and to quickly ascertain whether access permissions are correct or need to be modified, speeding up the initial impact assessment. Additionally, segregating data subject to GDPR supports a prioritized response, addressing the most critical requirements first. Grouping files according to usage patterns will enable removal of files that are no longer in use. There is no need to spend money protecting irrelevant data.

With DocAuthority, you will be able to:

- Identify all relevant GDPR related data within business files.
- Apply pertinent classification based on sensitivity, risk, criticality, sovereignty, retention, and access permissions amongst many other factors.
- Control specific data which can be deleted, processed, stored, shared and other controlling actions.
- Prioritize data to be protected.
- Review and correct data access permissions to meet least privileged requirements.

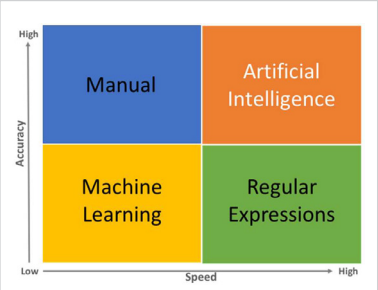
Another key challenge in complying with GDPR is the accountability inherent in the regulations. Organizations must be able to demonstrate how they are complying with the data protection principles. Governance programs must not just guide how to protect data and how to respond to breaches, but provide auditable evidence that sensitive data is being safeguarded, and that these safeguards are incorporating "privacy by design" rather than as an afterthought.

Furthermore, data processors are now subject to these requirements as well as data controllers, and face fines for non-compliance. Due to this change, contracts and agreements between data controllers and processors will need to be reviewed and updated to comply with the new regulation, and better protect customer, employee, and consumer data, as well as all parties involved.

DocAuthority enables tracking of progress in GDPR compliance activities, such as correcting access permissions, enabling encryption or anonymization of personal data, and removing unused files. Organizations can demonstrate the steps they have taken to comply with the GDPR by utilizing the reporting capabilities within the system as well as leveraging the embedded tags to drive actual data centric workflows.

Why Artificial Intelligence Beats Legacy Data Management and Classification Approaches

DocAuthority's approach to discovering and organizing unstructured data is built on an AI engine that uses deep learning to build a new navigation framework, the data map, based on business content and context. AI has two key advantages over legacy approaches: accuracy and faster time to value. Both advantages dramatically reduce the overall cost of ownership.



To date, the most accurate way of organizing and categorizing unstructured data is to do it manually. However, this approach is both time-consuming and costly. For example, assuming an organization has 100 million files and it takes 5 minutes to read, understand, and classify each file, and they hire staff to work 24 hours a day at $50 per hour, it would take 347,000 days and over $400 million to complete the project. Assuming files are added or edited at a rate of 1% per day would take an additional 3,500 days and $4 million to address each new day of additional or edited data. The manual approach is simply not feasible.

Technology leveraging regular expressions to detect certain types of sensitive data, such as credit card numbers and personal identification numbers, can speed up the process, but reduces overall efficiency due to a high false positive rate. Even systems combining detection approaches, such as including keywords and data fingerprinting, can yield false positive rates as high as 50%. Any time you gain in initial analysis is lost in addressing incorrect results. Improving the accuracy of these methods requires a significant amount of tuning, which also adds time and cost to the overall project.

Additionally, while these types of technologies can be used to find defined sensitive data patterns and demonstrate if the sensitive data is protected, they don't provide an easier way to navigate unstructured data nor effectively detect the new kinds of individual data (e.g. ethnic origin, political opinions) now considered sensitive personal data under the GDPR.

Machine learning provides a significant improvement over regular expressions and related technology, dropping the false positive rate from a high of 50% to 30%. However, machine learning systems need to be trained before they can begin analysis. The training process can take a significant amount of time since determining and gathering the training examples requires involvement from business stakeholders. Then the system output must be designed, the training data must be transformed into input to the system, and the effectiveness of the system evaluated and tuned before it can be implemented on real data – said basically, the upfront time and effort is too long relative to any reward.

DocAuthority's AI engine functions as a human would in reviewing files and making organization and classification decisions based on content and context. However, DocAuthority can process these files at a rate of 10 files per second. Following our example above, this reduces overall processing time to 115 days, versus 347,000 days; or considering this as a percentage, it equates to performing the same process in only 0.03% of the time.

DocAuthority's AI accurately organizes files according to a business taxonomy template encoded in the system. The actual false positive rate is less than 0.1% based on numerous production installations to date. The template is customizable, but does not require customization before the scan and analysis begins. Instead of requiring training, the AI trains itself as it works through the files, identifying sensitive content, expanding the taxonomy, and building more complicated file relationships as necessary.

Following the initial AI analysis, data trustees (typically 1-2 individuals designated by functional business departments) will validate the initial findings per business category, make any necessary updates, and tag data according to their business needs with the system for low maintenance governance, risk mitigation and/or remediation activities going forward. Tags are completely customizable and the system makes it easy to tag many thousands of files with just one click. Any new files detected on subsequent scans will be automatically organized and tagged according to the validated structure. The organized data is then searchable for key elements such as data owner, sensitive content, and data subject to GDPR-related requests.

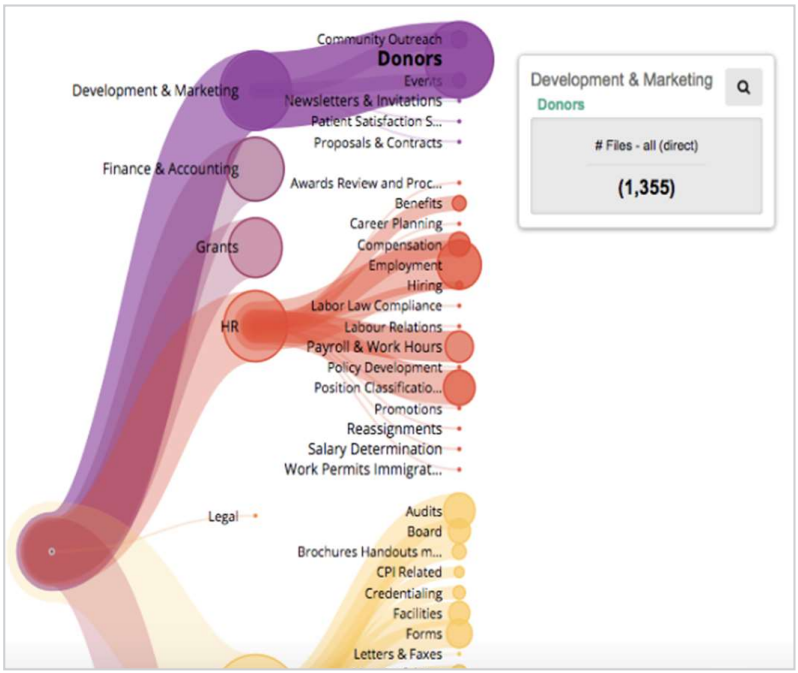
Step One: Data Discovery

The first step is to discover and organize all unstructured data in the organization. The DocAuthority AI performs all the work in this phase. All the organization's unstructured data can be discovered and organized through a series of scans either geographically or by business function. The DocAuthority system averages a rate of 10 files/second, which can be used to estimate how long it will take to understand all your business files. You will not need to define anything within the system, provide any samples of what you are looking for, nor design any inputs or outputs. This scanning time can be exponentially accelerated by applying incremental parallel virtual hardware.

The DocAuthority AI methodically opens, scans, and closes each file, then forms an opinion of the file based on content and context. The AI applies 1,500 discrete attributes when making decisions about business data. Using a template within the system, the DocAuthority AI builds a business function taxonomy and uses it to organize similar files into the structure, regardless of actual file location. For example, you may have financial statements that are maintained by your Finance department. There may also be financial statements that are maintained at a Business Unit level. The DocAuthority AI would discover these across the organization and group them all under Financial Statements, making it easy to find all financial statements across the organization.

It is important to note that the files are not moved from their locations. Instead, the DocAuthority system creates a data map to enable quicker and more effective navigation based on file content and context, and business purpose, rather than location. Customers typically find files with sensitive data in locations they would have never thought to look. This approach allows files to be matched, and subsequently managed by, their true business owners without disrupting existing business processes.

Step Two: Supporting GDPR Gap Assessment and Data Protection Impact Assessment



The DocAuthority-created data map (shown left) serves as a tool to more effectively navigate your organization's data. The DocAuthority system's role-based access control allows you to administer who can view and manage what types of data to support compliance with regulatory controls and company policies.

By clicking on the graphic, you can navigate the data flows, drill down into areas of interest, and even navigate to individual files. You can use the map to navigate your organization's unstructured data to identify:

- Data flows and authorized and unauthorized data usage
- Data locations, especially for sensitive data
- Data access permissions
- Missing controls (e.g. encryption, anonymization, DLP, etc.)

DocAuthority has uncovered some surprising findings with our customers to date – from organizations where all employees have access to all data, to organizations with substantial stores of data that haven't been touched in over 10 years.

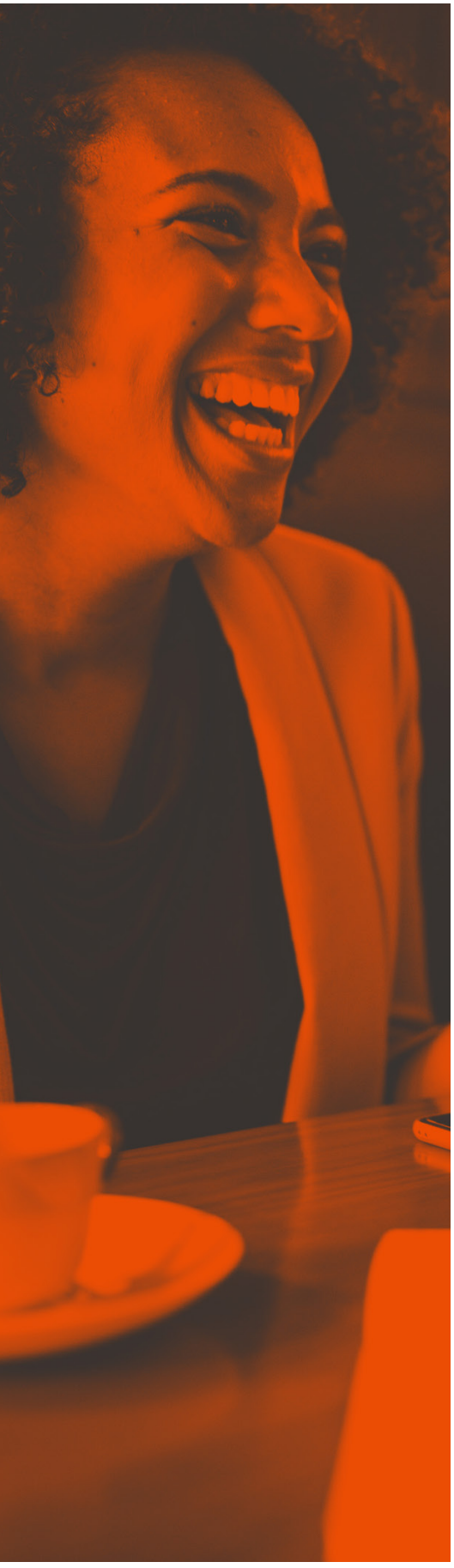
Step Three: Triage, Mitigate, and Create Response Roadmap

One of the outcomes of the Data Protection Impact Assessment should be a sense of what is critical to address first. The file tagging available in DocAuthority will help prioritize and triage mitigation and response planning. Based on your review of the file groupings created by the DocAuthority AI, identify and use the tags to segregate what is most sensitive and important in the organization. From there, assess the following:

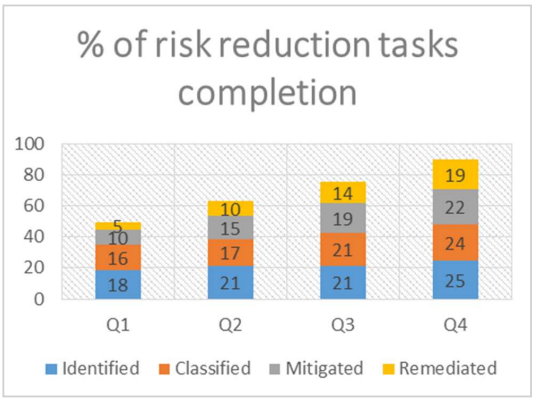
- Is file access sufficiently restricted?
- Are permissions least privileged?
- In what countries is the data stored and how may the GDPR impact that location?
- Are the files adequately protected?
- Can any of the files be archived or deleted because they are past their useful lifecycle?
- How can tags be applied to ensure the critical personal data be easily found in response to a "Right to be forgotten" request from a data subject?

Ask these questions for all your high priority data groupings and then develop a prioritized roadmap to address access permissions, data retention, and compensating controls. Identify which users and groups have the most access to sensitive data and make sure they are adequately trained to handle it. DocAuthority is designed to support remediation efforts by automating access permissions management, retention, and connection to other controls such as encryption and DLP systems.

One of the key advantages that DocAuthority offers in terms of GDPR compliance is the "evergreen" nature of the organization and classification structure. Any new information detected in subsequent scans is slotted into the existing organization structure and appropriately tagged, streamlining ongoing management and maintaining compliance. Once you organize the unstructured data, it stays organized with minimal effort.



Step Four: Track and Report Progress



Finally, identify the Key Performance Indicators (KPIs) and Key Risk Indicators (KRIs) and design trend charts within the DocAuthority system to manage and report on risk reduction and compliance progress. You can start with assessing the overall progress of identifying, classifying, mitigating, and remediation GDPR data. This effort can be tracked over time to demonstrate how you are bringing your GDPR subject data into compliance. These metrics can be monitored overall and at a country-specific or departmental level. They are derived from activities taken at the file level, and are therefore consistent and accurate across the organization.

Other recommended indicators include:

- Percentage of sensitive, GDPR-subject data in each country
- Risk reduction per country
- Risk reduction per department
- Percentage of each remediation activity (e.g. encryption, anonymization)
- Percentage of files in each stage of GDPR response (e.g. recommended, approved, remediated)

Conclusion

With DocAuthority, you will be able to:

- Understand the extent of GDPR-subject data in your organization
- Prioritize which data is the most critical and in need of protection or data handling changes
- Determine the resources needed to change processes or adopt measures to protect it
- Develop a roadmap and success criteria
- Measure and demonstrate progress against compliance.

To experience first-hand how DocAuthority can support your GDPR compliance program, contact us to schedule a 1-2 week discovery scan. We will scan and organize a subset of your data to prove the effectiveness of our artificial intelligence engine and show how to leverage it to more quickly to understand the scope of your GDPR needs and take action to comply with the new regulation in an efficient manner.

Contact us at info@docauthority.com to request our "Pre-Installation Checklist".

About Us

The explosion of data is a huge problem for organizations. Now with GDPR-like regulations coming into place, company data is now also a massive compliance risk. DocAuthority enables you to turn these compliance requirements into business opportunities and use them to dramatically improve all aspects of unstructured data usage, management and governance. DocAuthority's revolutionary and patented AI engine quickly and efficiently identifies and creates an inventory of all of your business data with the precision of 99.99%. With ease, you can now accurately identify both data's risk and its value and automate its ongoing classification, protection and retention while improving accessibility and quality.

Email

info@DocAuthority.com

Americas

+1 844 362-2884
3340 Peachtree Road,
Atlanta,
GA 30326

EMEA

+ 44 333 050 3241
Hamilton House,
Mabledon Place,
London
WC1H 9BB

APAC

+66 843 162 785
Sukhumvit Central
Business District,
51 Sukhumvit Road
Soi 8,
Klong Toey,
Bangkok
10110
Thailand

Israel

+972 1801 220 508
Ha-Tidhar St 15,
Ra'anana,
4365713
Israel