WHITE PAPER

GDPR Compliance & The Elastic Stack

TABLE OF CONTENTS

This white paper is provided for informational purposes only. It does not offer legal or audit advice. This white paper should not be relied on as a complete or accurate statement of the law. An organization's compliance with GDPR may be dependent on many factors outside the scope of this paper, ranging from its privacy policies and practices to its information security controls and organizational structures. For a complete and accurate statement of law or for legal advice for a particular situation, the reader should consult a competent attorney.

INTRODUCTION

The introduction of the European Union's (EU) General Data Protection Regulation (GDPR) heralds a wholesale change in the way thousands of organizations, worldwide, protect and use personal data. Arguably, this is the most pivotal change in data privacy regulation in 20 years. GDPR affects every organization, regardless of location, that captures and stores data on persons in the EU. All companies, large or small, must comply if they handle personal data of EU residents.

Purpose

The purpose of this white paper is to introduce the key concepts within the new regulation, plus offer suggestions as to how Elastic users can deploy Elastic Stack features to help them meet GDPR requirements related to the handling of personal data.

Background

Replacing the previous 1995 EU Data Protection Directive (and 1998 Data Protection Act in the UK), GDPR was developed in recognition of the increasing need to protect the rights and personal data of each individual EU resident. As the volume, depth, and breadth of personal data proliferates, it's easy to see why. GDPR is becoming increasingly recognized as regulation that will be leveraged to stem the increasing number of damaging data breaches reported across a variety of sectors. While previously compliant organizations may find many similarities to the earlier Directive, GDPR brings in some significant changes to the way personal data can be handled, rules on how breaches must be reported, and hefty penalties for non-compliance.

The European Commission believes that by homogenizing data protection laws throughout the single market, and by putting in place a more transparent and simpler legal landscape in which to operate, this will save corporations €2.3 billion each year, collectively.

With the deadline for compliance fast approaching, some claim corporations have a long way to go with preparations. According to recently published research from international law firm Paul Hastings, as of early January 2018, less than 39% of organizations (and 47% in the US) have created internal GDPR task forces. In addition, less than 40 percent are working with third-parties to conduct gap analyses and give counsel on compliance.

According to the UK Information Commissioner's Office, if an organization is already compliant with the current EU directive, then most of its approach to compliance will remain valid under GDPR and can be used as the starting point from which to build. However, there are new elements and significant enhancements to take into consideration.

As of this writing, there is no official GDPR compliance certification nor are there legal enforcements to use as a guide.

GDPR PRIMER

GDPR Affected Establishments

EU and Non-EU establishments may be affected by GDPR depending on their business models, geographical reach, and the subjects from which they control or process data. An easy way to think about this is to apply one of the following four classifications to organizations:

GDPR-Affected:

- EU Establishments (includes UK for now)
- Super-regional or global establishments with EU presence
- Establishments without EU presence, but with EU customers/visitors/subjects

Not GDPR-Affected:

- Establishments with no EU presence and no EU customers/visitors/subjects

Relevant GDPR Entities

GDPR defines roles or personas in terms of Data Subjects, Data Controllers, Data Processors, Sub-processors, and Authorities. Also, it defines a formalized role within the organization called the Data Protection Officer (DPO). The key roles are listed below:

- **Data Subject:** Persons in the EU
- **Data Controller:** Controls purpose and means of processing data. Direct responsibility to data subject and data protection authority
- **Data Processor:** Acts on instructions of Data Controller. Direct responsibility to data subject and data protection authority

Data in Scope for GDPR

Unlike other regulations, GDPR does not use the term "PII" (Personally Identifiable Information), but rather uses the term "Personal Data." In this white paper, we'll use capitalized "Personal Data" to refer to personal data as defined by GDPR.

Personal Data means any information relating to an identified or identifiable natural person ("data subject").

GDPR also defines "special categories" of personal data that have additional restrictions and requirements (racial, ethnic, religious, political, biometric, etc.). Additionally, Recital 10 of the GDPR equates these special categories of personal data with the term "sensitive data." In this white paper, we'll use the capitalized "Sensitive Personal Data" to refer to data that falls into these special categories.

"Identifiable Person" is a natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that person. [Chapter 1, Article 4(1)]

Protecting the Rights of Data Subjects

GDPR seeks to build on some of the key pillars of the current Data Protection Directive by significantly enhancing the rules around the processing and storage of personal data. GDPR includes the following rights for individuals:

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- The right not to be subject to automated decision-making including profiling

Many of these rights, existing within the older EU Directive, have been strengthened. What's new in GDPR are three key rights:

A) **The Right to Erasure:** This is designed to give individuals the right to be forgotten if the data held on them is no longer needed; the individual withdraws consent or objects to its use; there's a legal reason to remove it or it has been processed unlawfully; the data is related to a child / children. There is a responsibility for the operator to remove data and inform third-parties of the change.

B) **The Right to Restriction of Processing:** This is designed to make it easier to contest the accuracy or lawful processing of data where there is an objection regarding the legitimacy of the processing. If an individual asserts this right, his or her data can only continue to be processed for defense of legal claims or with the consent of the individual or for the protection of the rights of another person or an important public interest.

C) **The Right to Data Portability:** This protects the rights of individuals to access all data held by third-parties and to be able to view it in a commonly used format. Under this right, individuals can ask one company to share personal data with another.

Granting New Access

GDPR requires that data controllers and data processors must be more transparent about how they collect data, how they process it, and how they intend to store it. This must be communicated in a clear and unambiguous way. Under GDPR, individuals have the right to access any information an organization holds on them, to know why it's being processed, who it can be accessed by, plus where (and for how long) it is stored. GDPR expects organizations to provide direct, secure access for people to review what information is held.

Handling Data Breaches

According to a report from IBM Security and the Ponemon Institute, the average total cost of a data breach is an estimated $3.62 million (or €2.9 million). The rules for handling data breaches within the GDPR framework are clear: organizations must inform their local data protection authority of a breach within 72 hours of detection.

As many organizations fail to detect breaches in a timely manner, it is seen as a difficult criteria to follow and the onus will be on IT departments to put in alerting systems that are capable of handling larger amounts of structured and unstructured data that come from multiple different sources. Next-generation fraud detection and alerting methodologies will become crucial in detecting threats before they have time to do damage.

The Cost of Non-Compliance

A major change in GDPR is the way that non-compliance will be penalized; the severity of the penalties should not be underestimated. Those who fail to comply could face a penalty of up to €20 million or 4% of a company's annual worldwide revenue, whichever is higher. When looking at recent Information Commissioner's Office (ICO) fines issued (maximum penalty of £500,000) then scaling according to GDPR rules, it's clear how much harsher the penalties for non-compliance could be. Incredibly, the total fines issued by the ICO in 2016 amounted to £880,500 — this amount would become tens of million under the new GDPR rules. It's pays to be compliant!

Securing Personal Data

The security requirements in GDPR are general in nature and avoid prescribing specific technologies or practices.

"...the controller and the processor shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk...." [Chapter IV, Section 2, Article 32]

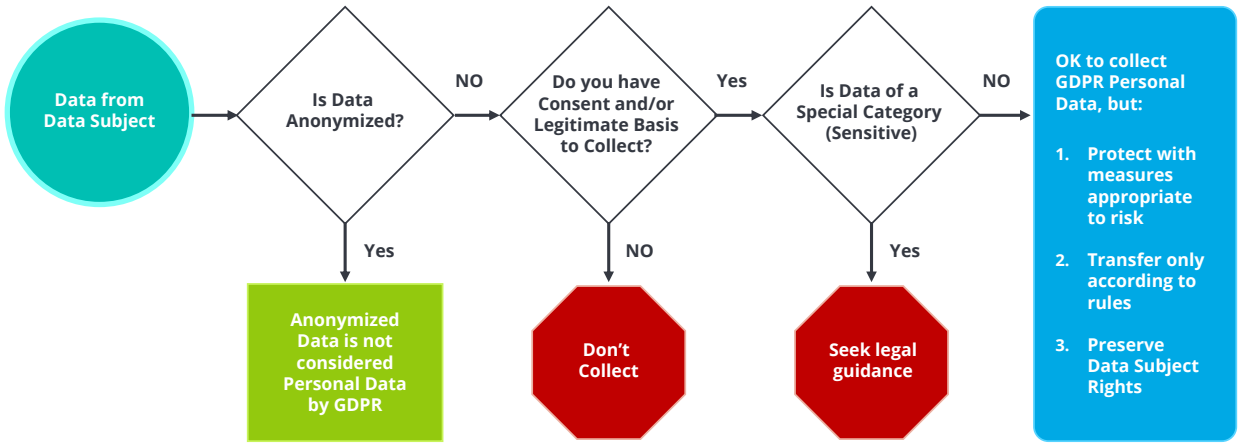
Those familiar with other information security frameworks such as SOC and ISO 27000, will recognize the terms "technical and organizational measures" to be a sweeping term describing all the things an organization's people, processes, and technology must do in order to provide the confidentiality, integrity, and availability of data.

Transferring Personal Data

Moving Personal Data across borders can be complex. Transfers of Personal Data out of the EU to a country that is not deemed to provide an adequate level of protection are only permitted if the controller or processor provide appropriate safeguards as described in the GDPR. These safeguards may include standard data protection clauses adopted by the European Commission (i.e., "Model Clauses"), binding corporate rules, or an approved self-certification program such as the EU-US Privacy Shield.

GDPR Handling of Personal Data in a Nutshell

The simplified diagram below summarizes the decision process a GDPR Affected organization should consider when determining how it treats Personal Data.

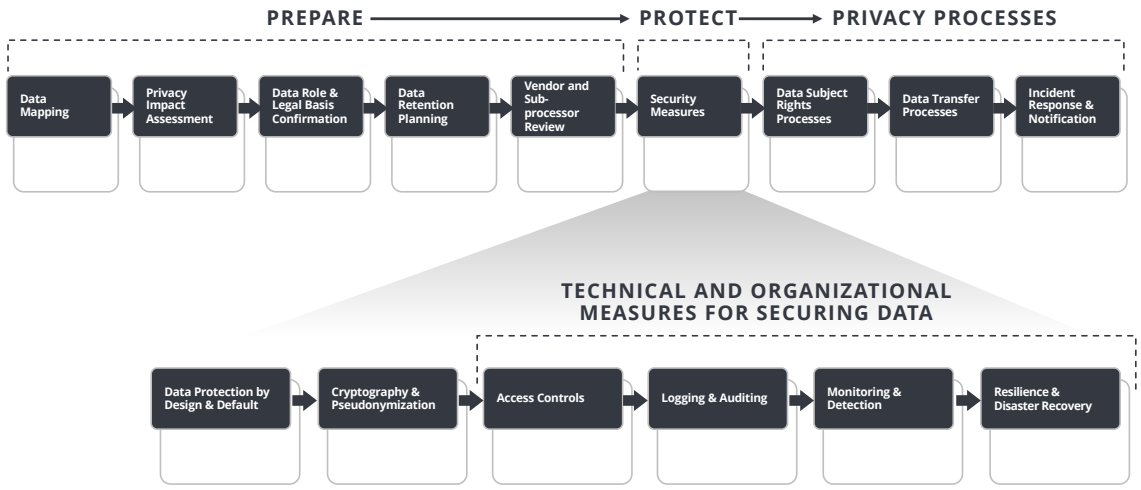


HANDLING GDPR PERSONAL DATA: A THREE STAGE PROCESS

With a basic understanding of GDPR in in mind, next the discussion turns to the more pragmatic side of compliance.

- How to **Prepare** an organization for controlling or processing Personal Data
- How to **Protect** any such Personal Data
- **Privacy Processes** to preserve the rights of the Data Subject over their data

The following diagram illustrates a typical process flow that an organization will follow while starting and maintaining a GDPR compliance initiative.



The process includes three stages, the subtasks of which are as follows:

Prepare for Handling Personal Data

The *Prepare* stage must be done at the initiation of the GDPR compliance cycle, but should be documented and integrated into an organization's ongoing Information Security processes and repeated at regular intervals.

- **Data Flow Mapping:** The process of identifying and documenting all the data flow processes within your organization that control or process Personal Data.
- **Privacy Impact Assessments [Chapter IV, Section 3, Article 35]:** The risk assessment process clarifies the level of risk associated with loss or disclosure of the Personal Data associated with a particular data flow. This assessment helps determine the appropriate level of protection required.
- **Data Role Confirmation:** For each data flow identified, confirm whether the organization's role is Data Controller, Data Processor, or both.
- **Confirm Legal Basis for Processing:** In this step, confirm the basis upon which the organization is authorized to collect Personal Data. In general, this step requires that for any data flow where Personal Data is collected, that there is a documented expression of consent by the Data Subject allowing the intended use of the Personal Data or there is a legitimate purpose for the collection and processing of the data..

- **Personal Data Retention Planning [Chapter III, Section 1, Article 12]:** For each identified data flow, determine how long Personal Data is stored. This is a good time to start thinking about how to delete data once its retention period has been reached and/or if the data subject might subsequently request it to be erased.
- **Vendors and Sub Processor Review:** The organization's responsibilities as Data Controller or Data Processor don't stop when data is passed to a sub-processor. In fact, under GDPR, the data controller or processor remains responsible for any acts or omissions of the sub-processor. So, it is important that an organization ensures that the sub-processor can adequately protect the Personal Data and that there is an appropriate data processing agreement in place to enforce such protection.

Protect Personal Data

The *Protect* stage is all about implementing the appropriate level of protection for the Personal Data associated with each data flow identified in stage one above. Security of Processing is the overarching theme of [Chapter IV, Section 2] Article 32, whereby GDPR states "…the controller and the processor shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk…" and can be summarized with the following set of steps:

- **Data Protection by Design & Default [Chapter IV, Section 1, Article 25]:** This step involves checking each data flow to ensure that wherever and whenever possible that anonymization and pseudonymization of Personal Data is maximized, while distribution of Personal Data is minimized.
- **Cryptography & Pseudonymization [Section 28 and Chapter IV, Section 2, Article 32]:** When data cannot be anonymized, using encryption and other pseudonymization techniques helps meet the requirement for data protection by design.
- **Access Controls [Chapter IV, Section 1, Article 25]:** One of the central tenets in securing data is the use of access controls to ensure that only authorized persons can access Personal Data.
- **Logging and Auditing [Chapter IV, Section 1, Article 25]:** As in many information security frameworks, ongoing logging and auditing must be performed to ensure that the access controls above are being implemented effectively.
- **Monitoring and Detection:** Again, similar to existing frameworks, this step requires ongoing monitoring and detection of threats that might affect the security of infrastructure components and applications involved in processing Personal Data.
- **Resilience and Disaster Recovery [Chapter IV, Section 2, Article 32]:** To guard against Personal Data loss, infrastructure and processes must be designed to preserve data integrity in the event of system disruptions or failures.

Privacy Processes for Handling Personal Data

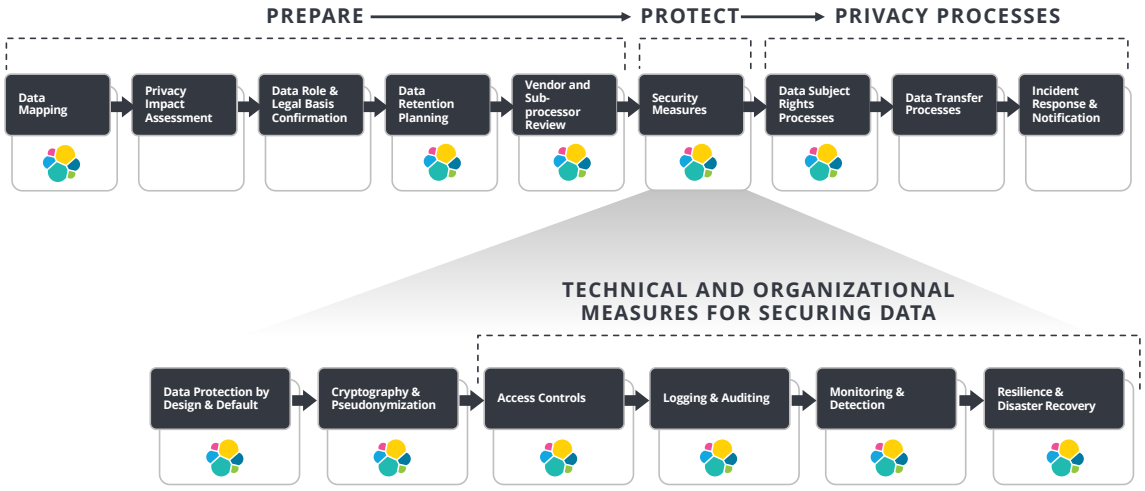
The *Privacy* stage is where an organization defines and implements a series of internal and external processes that will allow it to comply with specific GDPR regulations involving protection of data subject rights and controlled data transfers to internal or sub processor locations.

- **Maintaining Data Subject Rights:** As listed above, GDPR defines several rights for data subjects. An organization must put in place people, processes, and technology to enable the organization's ability to honor a Data Subject's requests to exercise their rights, such as their right to erasure.
- **Data Transfers Outside of the EU [Chapter V, Articles 44 – 45]:** During preparation for GDPR, this step involves the creation of Data Processing Agreements (DPAs) to be executed between controllers and processors as well as processors and sub-processors.

- **Incident Response and Notification:** This step includes an organization's internal processes for communicating detected incidents, as well as notifying the data processing authorities in the event of a data breach.

ELASTIC STACK FEATURES HELP MEET GDPR REQUIREMENTS

In the diagram below, the Elastic logo indicates the GDPR implementation steps in which the user can deploy the Elastic Stack features to help them meet GDPR requirements relating to the handling of Personal Data.



Using the Elastic Stack in the *Prepare* Stage

In this stage, organizations must identify all the data flows where Personal data is controlled or processed. Elastic Stack can be utilized to help with GDPR readiness in the following steps:

- **Data Flow Mapping:** Mapping data flows is the first step in GDPR preparation, and if an organization is unable to identify relevant data flows, the GDPR initiative may be incomplete and/or ineffective. Depending on where an organization is storing Personal Data today, it may make sense to import or ingest a copy of all such data into Elasticsearch, where its powerful and fast full-text search capabilities will enable quick identification of tables, queries, reports, or applications that rely on Personal Data.
- **Personal Data Retention Planning:** It is best to include this step in the Prepare stage, since this is when an organization must decide how long it will store the Personal Data it collects. GDPR specifies limited retention, and GDPR-affected organizations are required to delete Personal Data when it is no longer needed (or when the Data Subject withdraws consent). The retention of Personal Data stored in Elasticsearch can be easily managed through index management. Elasticsearch supports time-based indices — that can be deleted after the retention period has expired — using tools such as Ansible, Chef, or Puppet, as well as by using Elastic Cloud Enterprise (ECE), Elastic's central orchestration software, which can be used to manage a fleet of Elasticsearch clusters and address challenges that come with multi-tenancy, such as version control and data retention.

- **Vendor and Subprocessor Review:** Today's supply chains may extend to hundreds or thousands of vendors and subprocessors. With GDPR potentially requiring a Data Processing Agreement with each of them, the ability to index and perform full-text search instantly through thousands of Agreements can effectively facilitate vendor status reports and, more importantly, enable proactive vendor programs.

Using the Elastic Stack in the *Protect* Stage

In this stage, there are different scenarios in which the Elastic Stack can be utilized to help with GDPR readiness:

A) Securing the Elastic Stack when it is used as a datastore for Personal Data

B) Assisting with security measures by acting as a centralized security logging and analytics platform when Personal Data may be stored in another data platform. Note that in this case, the Elastic Stack may also be considered a Personal Data store since log data itself may contain Personal Data.

In either case, Elastic Stack features can help meet GDPR requirements for most steps within the Protect Stage. Deployed on secure systems, the Elastic Stack version 6 configured with X-Pack commercial extensions can be used to meet the following security requirements in this stage:

- **Data Protection by Design & Default:** If an organization is considering using the Elastic Stack as a datastore for Personal Data, the capabilities of ECE, Elastic's central orchestration software, can put the organization on the GDPR track from the start. The principle of data protection by design is about treating Personal Data like a valuable secret by limiting access, maintaining accuracy, ensuring data is secure, and limiting retention. Unlike traditional data architectures with one massive data store and volumes of complex overlapping data access controls (required for allowing access to only certain data by various projects), ECE makes it practical to instantiate new Elasticsearch clusters for each project, plus include only data relevant to that project in its cluster. This architecture enables the minimization of Personal Data — another GDPR principle for protecting Personal Data.
- **Cryptography & Pseudonymization:** Part of securing Personal Data means employing multiple levels of protection to ensure that data is not lost, destroyed, or disclosed to unauthorized individuals. Elasticsearch supports deployment on systems that have enabled system-wide disk-based encryption at rest. This makes it less likely that an unauthorized person accessing the underlying file system will be able to access cleartext Personal Data. Another GDPR principle for securing Personal Data is Pseudonymization, which is defined as "...the processing of personal data in such a way that the data can no longer be attributed to a specific Data Subject without the use of additional information." Logstash, a key component of the Elastic Stack, includes a set of capabilities in its Fingerprint filter plugin that can be used to implement pseudonymization — this capability is useful in both scenarios A and B listed above.
- **Access Controls:** To prevent unauthorized access to Personal Data stored in an Elasticsearch cluster, there must be a way to authenticate users. This means that a user is validated for who they claim to be. X-Pack security features provide a standalone authentication mechanism that enables quick password-protection of a cluster. If there is an external authentication mechanism to manage users in the organization, such as LDAP, Active Directory, or PKI, X-Pack security features are able to integrate with those systems to perform user authentication. X-Pack security features also include IP-based filtering. Also, it is easy to whitelist and blacklist specific IP addresses or subnets to control network-level access to the Elasticsearch cluster. However, to meet GDPR requirements, simply authenticating users isn't enough there needs to be a way to control what data users can access and which tasks they can perform. X-Pack security features enable control to authorize users by assigning access privileges to roles, and assigning those roles to users. This role-based access control (RBAC) provides the ability to specify which user(s) can perform read and/or write operations on the Elasticsearch indices holding Personal Data, plus allows implementing separation of duties for individuals authorized to access the data.

- **Logging and Auditing:** In scenario A, where the Elastic Stack is used as a data store for Personal Data, X-Pack security features enable an organization to maintain an audit trail by auditing security events. The audit log produced by Elasticsearch makes it easy to see who is accessing a cluster and what they're doing. By analyzing access patterns and failed attempts to access a cluster, insights can be gained into attempted attacks and data breaches. In Scenario B, where the Elastic Stack is not the primary store for Personal Data, the Elastic Stack can be used as the centralized logging platform for managing security-related logs from throughout the organization's infrastructure and application base. In this scenario, the Elastic Stack is used as a security analytics solution.
- **Monitoring and Detection:** In both scenarios, the Elastic Stack can be a key component of implementing the monitoring and threat detection for Personal Data stores. Usually, deployments follow basic principles of monitoring data store health, monitoring log continuity (ensuring that the flow of logs from data stores and other infrastructure is not interrupted), plus detecting malicious and suspicious activity within the environment (such as cyber attacks). X-Pack monitoring features help administrators keep a close watch on the health of the Elasticsearch Cluster. X-Pack alerting features enable automated monitoring of log continuity and notifications when interruptions or failures occur. Along with X-Pack alerting features, X-Pack machine learning jobs and Kibana dashboards can be the basis for an organization's threat detection platform, enabling ongoing security monitoring and interactive threat hunting as part of a security analytics solution.
- **Resilience & Disaster Recovery:** Whether as a data store for Personal Data or as a centralized logging platform to help with securing Personal Data, Elasticsearch has been designed from the start to be a distributed data store and search engine. It scales horizontally to handle extremely high event rates, while automatically managing how indices and queries are distributed across the cluster for smooth operations. The Elasticsearch cluster architecture includes replicas of index components (shards) for resilience and failover. Snapshot and Restore functions are built-in to allow for convenient backups. X-Pack security features help to preserve the integrity of data with message authentication and SSL/TLS encryption.

Using the Elastic Stack in the *Privacy Processes* Stage

- **Maintaining Data Subject Rights:** When a Data Subject exercises their right to erasure, or withdraws their consent to allow the collection of their Personal Data, one of the biggest challenges can be how to actually find that data. Elasticsearch can enable quick identification of the Data Subject's Personal Data in tables, queries, reports, or applications. Elasticsearch features, such as the Delete By Query API and the Update By Query API enable a team to take the appropriate action to satisfy these GDPR requests.

CONCLUSION

With the deadline for GDPR compliance looming, the race is on for all organizations to get ready. Heavy fines for non-compliance, as well as breach notification requirements with tight timelines, add to the aggressive implementation deadline. This means that if organizations haven't already begun GDPR preparations, then they must begin working on this now.

Many reports have indicated the cost to become GDPR compliant will spiral — many businesses have planned or are planning to bring on new, permanent teams to handle the transition. Becoming GDPR compliant will require focus and it's certainly true that a certain amount of an organization's resources will be occupied with the challenge.

As shown, it's possible to use the Elastic Stack technology to hasten the process and ensure data management processes are fit-for-purpose for the long term. Specifically, using the Elastic Stack as a data store for GDPR Personal Data provides organizations with a strong starting position for building a GDPR-compliant data store with security, access controls, resilience, and disaster recovery capabilities. The inclusion of Elastic Cloud Enterprise can further align an organization's data processing efforts with the GDPR principles of protection by design and personal data minimization.

Additionally, the Elastic Stack can be used as a centralized logging system to implement the GDPR principles of protection by design, cryptography and pseudonymization, plus the technical and organizational measures for protection of Personal Data, including access controls, logging and auditing, as well as monitoring and detection that can help implement an effective overall approach to GDPR compliance.

NEXT STEPS WITH ELASTIC

Prepare to meet GDPR requirements with Elastic Stack features. Utilizing Elasticsearch, Logstash, Kibana, and the X-Pack can help meet requirements in many of the steps organizations must take on their path to GDPR readiness. For more details, or to speak to an Elastic representative, please use the contact details below.

ELASTIC: EUROPE HEADQUARTERS

Rijnsburgstraat 9-11

1059 AT, Amsterdam

General +31 20 794 7300

Sales +31 20 794 7310

info@elastic.co

sales@elastic.co

ELASTIC: US HEADQUARTERS

800 West El Camino Real, Suite 350

Mountain View, California 94040

General +1 650 458 2620

Sales +1 650 458 2625

info@elastic.co

sales@elastic.co

Learn more about Elastic services, including support, consulting and training.

RESOURCES

Official EU GDPR Web Site - https://www.eugdpr.org/

WP-GDPR021318