### Contents

1. Executive overview 3
2. The need to inventory personal data (discover) 3
   - Data is not always easy to find 4
3. Data loss prevention as a technical measure to inventory personal data 5
4. Forcepoint’s DLP solution is fit for purpose 5
   - Data inventory capabilities 5
   - Forcepoint DLP can scan many different locations to find sensitive data 6
   - Management, monitoring, reporting and incident response capabilities 6
   - Integration with third party technologies 5
5. More detail and guidance around Forcepoint’s DLP solution 7
   - Configuring Forcepoint DLP for GDPR policies 7
   - Configuring Forcepoint DLP to fingerprint PII contained in structured data 7
   - Configuring Forcepoint DLP to inventory personal data 8
   - Reviewing the results of your personal data inventory 9
   - Responding to a data subject access request 9
6. Next Steps 10

**DISCLAIMER**
Although Forcepoint has made every effort to ensure the accuracy of this paper which has been prepared in good-faith, Forcepoint cannot accept any responsibility whatsoever for any consequences that may arise from any errors or omissions, or any opinions given. This paper does not constitute legal advice and Forcepoint makes no representation or warranty, express or implied, regarding its products including without limitation fitness of its products for a particular purpose. In no event will Forcepoint be liable for any direct, indirect, incidental, consequential, special, or punitive damages related to this paper. The information provided in this paper is the confidential and proprietary intellectual property of Forcepoint, and no right is granted or transferred in relation to any intellectual property contained in this paper.

Copyright © 2017 Forcepoint. All Rights Reserved.
Executive overview

The EU General Data Protection Regulation (GDPR) has been approved and is set to replace the previous EU Data Protection Directive. The new regulation will come into effect in May 2018 and will require organizations to put a much stricter focus on data protection.

The GDPR provides broad guidance around all aspects of data protection, but is not prescriptive in terms of the information security requirements needed to adequately protect private data.

In this series of papers, Forcepoint provides an interpretation of the regulation and how it maps to Information Security technology, and more specifically, how Forcepoint technology can be used to aid your compliance efforts.

Each paper will focus on key areas where technical controls play a part in gaining or demonstrating compliance of the GDPR.

In this paper, we examine the need for organizations to protect and maintain personal data inventories either as part of the initial phase of meeting the new compliance requirements and impact or as part of the ongoing operational phase of lawful processing personal data.

Before we get started, let’s discuss what is defined as ‘personal data’.

From Article 4 section 1: ‘Personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

This is a broader definition and includes on-line identifiers like IP address and email address for example.

The need to inventory personal data (discover)

The GDPR requires the Data Controller and Data Processors to understand its processing activities.

This means that organizations will need to understand not only how and where data is being used and transmitted, but also where it is stored. In terms of data discovery, the action of processing should be mapped to a system or person and recorded in addition to the data.

Data processing is defined in Article 4, section 2 as: ‘Processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

There are articles within the GDPR that relate specifically to the need for inventorying personal data:

Chapter 2 (Principles), section 3 (Rectification & Erasure):

- Article 17 (Right to erasure / ‘right to be forgotten’): ‘The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay’.
- Article 20 (Right to data portability): ‘The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format’.

The articles above refer to some of the rights data subjects have on the Controller [and Processors] over their personal data. Any data subject access request will mean that controllers must respond in a timely manner (within one month of the receipt of the request). In order to do this, organizations must know where this data exists, or be able to find it quickly across their infrastructure.

Chapter 4 (Controller and Processor), section 1 (General Obligations):

- Article 24 (Responsibility of the Controller): (1) ‘The controller shall implement appropriate technical and organizational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation’.

The article above explains that Controllers need to gain oversight of lawful processing of personal data (which includes the storage of this data), using technical and organizational measures. Ongoing inventorying of data can be an effective way to maintain this visibility.

In addition, Chapter 5 (Articles 44 – 50) are focused on the ‘Transfers of personal data to third countries or international organizations’. This section explains the conditions of when personal data can be transferred or processed outside of the EU, including Article 46: (Transfers subject to appropriate safeguards).
Invariably, when Forcepoint conducts data risk assessments for these organizations that are at the early stage of their Data Protection program, sensitive data is found in places that it shouldn’t be: in third party cloud services, laptop devices, or even file shares in publicly available parts of their network. This visibility provides organizations with the opportunity to remediate privacy violations and refresh data protection policies and processes, which can improve the overall level of governance around personal data processing.

DATA IS NOT ALWAYS EASY TO FIND

It is worth considering the two main types of stored data: structured and unstructured. Structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spreadsheets. Conversely, unstructured data is not limited to a fixed field and can exist anywhere in a file (for example, a word processing document, presentations, web pages, multimedia files, etc.). According to many experts, the majority of data within an organization (up to 90%) is unstructured, which makes it much harder to inventory.

Organizations will not only need to inventory personal data operationally, as part of the day to day data subject access requests detailed above, but it should also be looking to proactively inventory personal data as part of the preparation or planning phase of scoping out the effort and resources required to undertake GDPR compliance. Many of the conversations Forcepoint is having with organizations today are at the early stages of discussion around GDPR, and these organizations are seeking to understand the extent of their data flows (how far data has been distributed and duplicated throughout their organization).

As a cloud services provider, Forcepoint takes its responsibility as a processor very seriously; that’s why Forcepoint’s cloud infrastructure is certified to international standards that are recognised within the industry. For example, we hold ISO27001, ISO27018 and CSTAR level 2 certifications. In addition, Forcepoint actively participates in The Privacy Shield framework:
https://www.privacyshield.gov/participant?id=a2zt0000000TNRtAAO
Data loss prevention as a technical measure to inventory personal data

Data Loss Prevention (DLP) is a common technical measure to help organizations inventory personal data. The term ‘Data Loss Prevention’ is a little misleading, as prevention of data loss is not the only use case for organizations using this technology. DLP is in fact a proven technology that helps organizations to understand their data and digital footprint, providing visibility into how their data is processed (at rest, in motion and in use).

Many analysts support this view: in a recent study published by Osterman Research, ‘GDPR Compliance and Its Impact on Security and Data Protection Programs’ and a recent paper from Gartner, ‘Focus on Five High-Priority Changes to Tackle the EU GDPR’, DLP is identified as a top technical control for GDPR.

DLP technology must be able to inventory data wherever it resides in a modern organization, whether it’s located on the local area network or, as is increasingly the case, off-network. Local file stores might include Mac or Windows workstations, email inboxes, databases or file shares or network attached storage. Remote locations could include laptops owned by an increasingly mobile workforce, third party file sync and share or storage cloud service providers, or cloud email and collaboration systems like Microsoft Office 365, for example.

A mature DLP solution must not only be able to scan structured or unstructured data for key words or alphanumeric patterns (like credit cards or national tax IDs), but it must be able to use more complex methods of data scanning, including fingerprinting and optical character recognition (OCR). It leverages the broadest set of analysis tools for structured and unstructured data formats, including personal information (PII) or intellectual property (IP) across both structured or unstructured formats, including intellectual property, and provides a ‘signature library’ that the DLP technology can use as a reference to identify this content again as it is processed within the organization (including the inventory or discovery of personal data).

4 Forcepoint’s DLP solution is fit for purpose

Forcepoint has been a leader in the DLP market for many years, both in terms of market share and feature set, as validated by key analysts1 globally. In fact, Gartner rated Forcepoint as the highest in terms of compliance in their Critical Capabilities for Enterprise DLP report published in April 2016.

Forcepoint’s DLP2 solution has a broad set of features that have been developed over many years to solve some of the real problems and challenges that organizations large and small face around protecting personal data and intellectual property.

DATA INVENTORY CAPABILITIES

- Forcepoint DLP is able to identify personally identifiable information (PII) or intellectual property (IP) across both structured and unstructured data formats, including personal data contained within images using Optical Character Recognition (OCR). It leverages the broadest set of analysis tools in the industry:
  - ‘Described’ content – this is content that the DLP solution can be programed to look for without seeing it before (e.g., you don’t have to have a record of every credit card in existence, you just have to be able to recognise the pattern that credit cards follow):
    - Lexical (keyword) analysis.
    - Alphanumeric patterns (using ‘regular expressions’) like credit cards or national tax IDs.
  - ‘Learned’ content – this is content that the DLP solution is shown beforehand and can be programed to look for again (e.g., a citizen name and address, medical records or business planning documents).
  - ‘Described’ content – this is content that the DLP solution is shown beforehand and can be programed to look for again (e.g., a citizen name and address, medical records or business planning documents).

1 Gartner Magic Quadrant, Forrester Wave, Gartner’s Critical Capabilities for Enterprise Data Loss Prevention report.
For a more detail around these capabilities, please refer to the third paper in this series entitled, "The Need for Preparation to Report Personal Data Breaches in a Timely Manner". In this paper we provide an in-depth analysis on how Forcepoint’s DLP and insider threat technologies can streamline the response and remediation efforts during a data incident.

**Forcepoint DLP** provides organizations with a comprehensive set of tools to be able to visualise, manage and respond to data protection incidents.

- Hierarchical management and access to management information & reporting:
  - In the process of monitoring users as they process personal data, organizations will generate and collect personal data on the processor (or employee). The privacy laws that are designed to protect the citizen’s personal data are also applicable to the user, too.
  - Forcepoint DLP has support for tiered access to the management console, including reporting and incident work flow.
  - When the inevitable happens and a device gets lost or stolen, it is imperative to know exactly what data was stored on that device in order to understand regulatory obligations and impact.
  - Data hoarding can also be an early sign that an individual may have a more sinister motivation, such as the intent to steal data as they leave employment.

- Reporting, logging and alerting:
  - The Forcepoint DLP solution centralises the policy management and reporting within the Forcepoint Manager. This is the same console where existing Forcepoint customers are able to manage Web and Email security gateway policies, too.

> Critical, PreciseID is not limited to exact matches, but is also able to match against partial derivative content (like copy/paste) that has been taken from previously fingerprinted data.

> Machine Learning technology searches for data that 'looks like or is related to' the fingerprinted sensitive data. This technology is very effective at protecting newly processed sensitive data within the organization and its various departments that resembles the already trained data set.

**FORCEPOINT DLP CAN SCAN MANY DIFFERENT LOCATIONS TO FIND SENSITIVE DATA**

- **Endpoint devices** (Windows & macOS)
- **Microsoft Exchange** (inc. online version)
- **Microsoft SharePoint** (inc. online version)
- **Shared Storage**
- **Box** (Cloud scan)
- **Databases**
  - Mac and Windows endpoint devices. When combined with the Forcepoint DLP agent, organizations can inventory personal data on local and remote devices. This is particularly important as workers become more and more mobile. In addition, many users consciously or unconsciously ‘hoard’ data onto their devices, putting the organization at increased risk.
  - When the inevitable happens and a device gets lost or stolen, it is imperative to know exactly what data was stored on that device in order to understand regulatory obligations and impact.
  - Data hoarding can also be an early sign that an individual may have a more sinister motivation, such as the intent to steal data as they leave employment.
  - Databases.
  - Local file shares.
  - Microsoft Exchange (& user mail stores) and SharePoint servers.
  - Cloud applications (like Salesforce, Box or Office 365 OneDrive).
With over 20 prebuilt report templates specifically for data discovery, organizations are able to understand many aspects of data processing in their organization and respond to data incidents through incident management and specific or scheduled reporting.

Security analytics and risk prioritization:
- In 2016, Forcepoint released the industry’s first DLP security analytics capabilities into Forcepoint DLP.
- Incident Risk Ranking (IRR) is the first Forcepoint DLP feature that leverages our new integrated security analytics software appliance that is tuned for a DLP data set and designed to address common DLP operational challenges.
- IRR provides a stack ranked security operations report that highlights data theft DLP cases; i.e., deliberate attempts by users or systems to exfiltrate business critical data.
- At a high level, the analytics platform uses event clustering and grouping (linking related activity), organizational and employee level baselines (anomaly detection) and Bayesian belief networks (activity classification) to structure, classify and apply a risk score to DLP cases. IRR greatly reduces the time it takes incident responders to respond to data incidents.

Incident workflow for remediation:
- Console-based Incident Workflow enables responders to distribute incidents for review and remediation to data owners and business stakeholders by leveraging the built-in, role-based management capabilities of the Forcepoint Manager.
- In addition, email-based incident workflow makes it easy to distribute an incident for review and remediation to data owners and business stakeholders without needing to provide access to the DLP management system.

INTEGRATION WITH THIRD PARTY TECHNOLOGIES
Forcepoint DLP integrates with many other technologies including data classification providers, customers can use this integration to classify PII data discovered during an inventory. This could be useful as part of a broader classification project, or a Cloud migration project. Data can be marked as ‘sensitive: internal storage only’ for data that must not leave specific local areas, or ‘sensitive: specified datacenter only’ for data that can be migrated to specific or approve cloud providers only, for example. Forcepoint’s DLP solution can then make policy-based decisions on data at rest, in motion or at use based on the classification criteria.

Forcepoint DLP can integrate with other 3rd party data protection technical measures like encryption and digital rights management providers to provide deeper visibility or policy-based remediation of specific violations of privacy found within data at rest. These remediation actions can be accessed directly from the Forcepoint Manager console.

Forcepoint DLP integrates with SIEM solutions to assist with the visualisation and management of incident activities as part of a larger governance risk and compliance program too.

5
More detail and guidance around Forcepoint’s DLP solution

In this section we will demonstrate specific features and configurations of Forcepoint’s DLP solution that support the claims made within this document on the inventory of personal data.

CONFIGURING FORCEPOINT DLP FOR GDPR POLICIES

Prior to running the first discovery scan, GDPR policies must be configured so Forcepoint DLP can find the relevant and regulated data.

An organization must identify which countries it operates within in order to select the relevant pre-defined compliance and data protection policies.

Forcepoint provides a large library of predefined policies which have been developed by an experienced in-house policy research team.

Policies are provided to identify and classify personal information and address data protection regulations and legislations across 85 countries globally.

Each of these policies uses a regulation-appropriate combination of content and context, and lexical analysis to identify personal data. The accuracy and maturity provided by our PII policies is one of the leading reasons that Forcepoint is selected by our customers.

Below you can see a breakdown of personal data classifiers that are relevant to the GDPR:

Financial Classifiers:
- Credit cards (with an ability to select a classifier for Credit Cards prevalent in the EU).
- Financial data (generic, globally-relevant classifiers, and additionally, specific country set of classifiers – includes 28 countries across EMEA).
- Financial regulations (including specific EU Finance classifiers).
- Payment Card Industry (PCI, globally-relevant).

Privacy Classifiers:
- Generic EU Directive PII classifiers.
- Generic EU finance set of classifiers.
- 13 EU country-specific set of classifiers.
- Further 7 country-specific set of classifiers for countries in EMEA (outside of the EU).

Refer to this link to see more detail on all of the classifiers available in the FORCEPOINT DLP solution by default: https://www.websense.com/content/support/library/data/v83/policy_classifier/data%20usage%20policies.aspx
Most organizations will not want to rely on dictionary and regular expression-based policies alone. Many of Forcepoint’s customers also choose to fingerprint PII data contained in structured data formats (for example, databases, or CRM cloud service providers). This greatly increases the accuracy of detection of PII during personal data inventories and cuts down on false positives.

Figure 4 shows many of the pre-defined policies available to an organization.

**CONFIGURING FORCEPOINT DLP TO FINGERPRINT PII CONTAINED IN STRUCTURED DATA**

Most organizations will not want to rely on dictionary and regular expression-based policies alone. Many of Forcepoint’s customers also choose to fingerprint PII data contained in structured data formats (for example, databases, or CRM cloud service providers). This greatly increases the accuracy of detection of PII during personal data inventories and cuts down on false positives.

Figure 5a shows how an administrator can configure Forcepoint DLP to fingerprint personal data contained within Salesforce.com.

Figure 5b shows the results of a fingerprint scan.
CONFIGURING FORCEPOINT DLP TO INVENTORY PERSONAL DATA

Forcepoint DLP can be configured to scan for personal data in a variety of locations across an organization’s infrastructure including:

- Endpoint devices (Windows AND macOS)
- Microsoft Exchange (inc. online version)
- Microsoft SharePoint (inc. online version)
- Shared storage
- Box (Cloud scan)
- Databases

REVIEWING THE RESULTS OF YOUR PERSONAL DATA INVENTORY

The Forcepoint Manager displays the results of the inventory, highlighting where personal data is found. In this example, you can see results of scanning for personal data across many different location and storage types.

An incident responder or stakeholder in the workflow process is able to look into each record to see more detail including:

- Where the file or record is located.
- Who has access to the file containing the record.
- Detailed file properties including owner, and when the file was created and last accessed.

![Figure 6: Shows the results of a DLP Discover scan.](image)

Forcepoint’s DLP solution is able to detect a partial match from previously fingerprinted data. This could be where a user has copied and pasted an extract from a sensitive document; for example, even when copied to a different file format. When the DLP solution finds a partial match, it will provide a confidence score to minimize false positives.

Another key capability Forcepoint DLP provides for our customers is to be able to anonymize personal data and user data, so that only those authorized to see this data have access to it.
Figure 7: Shows an incident where text within an image is detected as personal data.

Figure 8 shows how personal data can be masked or anonymized so that it is not possible for the responder to see the personal data that is the subject of the incident or the user behind the data incident, unless he or she has the authorization to do so. Forcepoint DLP customers are able to use our built-in hierarchical management capabilities to put in place well defined policies and processes around this so that there are multiple business functions involved in the approval process.
Where personal data is found to be located in high-risk areas (like public file sync and share, or remote Apple macOS laptop device), or with inappropriate access rights, Forcepoint DLP provides a responder with the ability to initiate a new workflow to start the remediation process. Remediation scripts and actions could include:

- Escalate (to manager or another person).
- Move.
- Delete.
- Encrypt.
- Apply DRM.
- Apply masking.
- Apply Classification.
- Apply Pseudonymisation (for a test system for example).

**RESPONDING TO A DATA SUBJECT ACCESS REQUEST**

When a Subject Access Request (SAR) is received, the personal data inventory can be used to the advantage of the data controller. By searching the incident data set (the results of the personal data inventory detailed above in figure x), a controller can see quickly and easily where this data resides.

**5 Next steps**

For more information on the General Data Protection Regulation please visit our website: [http://www.forcepoint.com/gdpr](http://www.forcepoint.com/gdpr).

To learn why organizations must ensure they understand data flows and how DLP technology can assist them to manage and control personal data flows as part of meeting GDPR requirements, please read part 2 entitled, "Data Flow Mapping & Control".

To learn which technologies can assist organizations to respond to data breaches in a timely manner (within 72 hours of the controller becoming aware of the data breach), please read part 3 entitled, "Preparation to report personal data breaches in a timely manner".

To arrange a demonstration or to request a GDPR Risk assessment, please contact your local Forcepoint sales office here: [https://www.forcepoint.com/company/contact-us](https://www.forcepoint.com/company/contact-us).