GDPR Technology Mapping Guide

DATA FLOW MAPPING & CONTROL
# Contents

1. Executive overview 3
2. The need to gain oversight and control of personal data flows 3
3. Data loss prevention (DLP) as a technical measure to map and control personal data flows 5
4. Cloud access security broker (CASB) solution as a technical measure to map and control personal data flows 5
5. More detail and guidance around Forcepoint’s DLP solution 5
   - Data Flow Mapping, Monitoring And Control Capabilities 6
   - Management, monitoring, reporting and incident response capabilities 7
   - Integration with third party technologies 7
6. Detail and guidance around configuring Forcepoint’s DLP solution 8
   - Configuring Forcepoint DLP for GDPR policies 8
   - Configuring Forcepoint DLP to fingerprint PII contained in structured data 9
   - Configuring Forcepoint DLP to monitor personal data flows 9
   - Reviewing policy violations and alerts from monitoring personal data 11
7. Next Steps 12

---

**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 © 2018 Forcepoint. All Rights Reserved.
1 Executive overview

The European Union (EU) General Data Protection Regulation (GDPR) is now in effect which requires organizations to put a much stricter focus on data protection.

The EU 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 new 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 Paper 1 we discussed the need to inventory personal data either initially, in order to be able to scope out compliance efforts and to understand ‘sensitive data drift’, or as part of the day-to-day tasks when responding to data subject access requests.

In this paper, we identify why organizations need to understand data flows and how technologies like DLP are best suited to assist them to manage and control personal data flows as part of meeting GDPR requirements.

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.

2 The need to gain oversight and control of personal data flows

The GDPR requires data controller and data processors to understand its processing activities.

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’.

This means that organizations need to understand not only where data is stored, but also where data is being used and transmitted.

Figure 1: The evolution of IT infrastructure.
Modern business practices and demands have resulted in an evolution of IT infrastructure, whether it’s through an increasingly mobile workforce or the adoption of cloud infrastructure, data is now more distributed than ever. This means mapping the flow of data is becoming more challenging, but organizations still need to identify rogue business processes that expose areas of contractual risk with suppliers of unsanctioned IT such as file sync and share services, for example.

There are articles within the GDPR Journal that relate specifically to the need to map out and manage personal data flows:

**Chapter 4 (Controller & 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’.

Article 25 (Data protection by design and by default): ‘The controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects’.

**Chapter 4 (Controller and Processor), section 2 (Security of personal data):**

Article 32 (Security of processing): ‘The controller and the processor shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

- (a) The pseudonymisation and encryption of personal data; 4.5.2016 L 119/51 Official Journal of the European Union EN
- (b) The ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
- (c) The ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
- (d) A process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the processing’.

The articles above that are part of Chapter 4 detail some of the responsibilities of the processors and controllers with regard to technical and organizational measures to protect and monitor the lawful processing of personal data.

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).

Organizations with supplier relationships outside of the EU, where data may be processed, need to monitor these relationships and perform due diligence to ensure these relationships or data transfers are not in violation of the new EU regulation and do not put private data at risk. Organizations should therefore ensure they know when and where personal data flows; specifically to third countries or international organizations.

Organizations need to proactively map out and monitor data flows in order to minimise data loss or theft incidents. As part of the initial stages of meeting GDPR compliance, many organizations are working cross-functionally with business stakeholders to understand what data they are processing - both intellectual property or regulated (personal) data. This processing analysis must also include how this data flows inside and outside of the organization.

In many cases, the majority of data flows can be mapped out quickly without the need for relying exclusively on technology like data loss prevention (DLP). For instance, it is clear that human resources (HR) departments need access to employee personal data. It will also be fairly simple to map out how they need to process this information and where it needs to flow (for example, to and from a human capital management [HCM] software solution, or from spreadsheets). Personal data may need to be legitimately transmitted to third countries if the organization has chosen a cloud-based HCM service based in the USA, for example. Alternatively, it would be safe to assume that blocking the flow of employee personal data to third countries (or other recipients) not authorised would be a sensible ‘block by default’ approach to minimise unlawful processing, which would pose a potential breach in the GDPR.

Monitoring the flow of data must be done across the organizational infrastructure: not only at the obvious communications gateways (like web & email), but also at the endpoint, to cover other communications channels. This is especially necessary where the device is being operated outside of the organization (i.e. a remote worker); or if a user has configured local folders to sync with personal cloud-based file sync and share applications, such as Dropbox, for example.
Cloud access security broker (CASB) solution as a technical measure to map and control personal data flows

CASB is a term describing a solution that provides visibility and control of cloud-based application use (including the flow of data into these third party suppliers). Users often provision and use many different cloud applications, generally for innocent reasons like increasing their productivity and innovation. However, with the strict guidelines of the upcoming GDPR, these cloud services are a potential risk. Consider that many organizations simply don’t have the visibility into the extent of unsanctioned IT (shadow IT) or even the extent of their use of sanctioned IT. Reports can be generated to profile the amount of cloud application in use and, based on a number of risk factors (including how they process data), an overall risk score can be calculated. When integrated with enterprise DLP, organizations can manage and monitor the flow of personal data to both sanctioned and unsanctioned applications.

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 analysts globally. In fact, Radacati rated Forcepoint as a Top Player in their DLP Market Quadrant report published October 2017.

Forcepoint’s DLP 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 in protecting personal data and intellectual property.

Figure 3: Forcepoint’s DLP solution covers all areas where data is stored, accessed or transmitted.
DATA FLOW MAPPING, MONITORING AND CONTROL CAPABILITIES

- Forcepoint DLP is able to identify personally identifiable information (PII) or intellectual property (IP) across both structured and unstructured data formats, including text 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).
    - PreciseID fingerprint analysis profiles more complex data in 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].
      - Critically, 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 monitoring newly-created sensitive data as it flows across the organization from its various departments.
- The Forcepoint DLP solution suite provides your organization with the broadest view of data flows within and outside of the organization. Whether it’s across the network, outbound flows over email and web (and other communication channels), or from remote endpoint devices, or as it flows to sanctioned cloud applications like OneDrive for business, Forcepoint DLP solution delivers data visibility. In addition, Forcepoint DLP can provide users with visual aids to help them learn how to process personal (or other critical data) lawfully and responsibly by displaying pop-up messages as they interact with this type of data.

Network-based DLP: Forcepoint DLP can monitor data in motion by integrating with existing security gateway technologies like web security gateways, email security gateways or cloud access security brokers. In addition, Forcepoint’s own Web Security Gateway and Email Security Gateway already include a built-in integrated DLP module.

- Forcepoint’s DLP module integrated with the Forcepoint Web Security Gateway can perform policy-based inspection of encrypted web communications (SSL or HTTPS) and can leverage the URL categories into the policy logic.
- Forcepoint’s integrated DLP module contained with the Forcepoint Email Security Gateway can provide policy-based email encryption, allowing sensitive data to flow to legitimate or approved recipients.

Endpoint-based DLP: Forcepoint DLP provides an agent for both Windows and Apple Mac OS devices that monitors data in use. Endpoint DLP is one of the fastest DLP deployment modes because it solves the dilemma of protecting the increasingly mobile workforce. It can monitor data transmitted from the endpoint device and is especially effective at mapping out and controlling data flows to unsanctioned applications and locations, regardless of the location of the device. A good example of where endpoint-based DLP is effective is where a user legitimately stored personal data in a folder on their laptop device, but that folder is configured to synchronise with a cloud-based file sync and share service.

Mobile device DLP: Working with leading Mobile Device Management (MDM) providers, Forcepoint DLP can restrict sensitive information from being sent to any ActiveSync-enabled mobile device. For additional flexibility, when Forcepoint DLP is integrated with Forcepoint CASB, organizations can manage sensitive data being synchronised with both managed and unmanaged [BYOD] mobile and laptop devices.

Cloud-Based DLP: Forcepoint DLP can be used to monitor and manage personal data flows between both sanctioned and unsanctioned cloud applications.

In-line deployment: When deployed in-line [either integrated with Forcepoint’s email or web security gateways, Forcepoint CASB, or any other third-party gateway provider], Forcepoint DLP can be used to protect personal data flowing across these gateways [including encrypted traffic in the case of the Forcepoint gateways].
API-Based (Native) deployment: Forcepoint DLP can be configured to talk natively to specific sanctioned applications when integrated with Forcepoint CASB. This approach is preferred where organizations are looking to manage the flow of personal data, regardless of whether the user device is managed (i.e. has a DLP client installed), or is not behind a monitored security gateway or service (e.g. a remote or highly mobile worker).

MANAGEMENT, MONITORING, REPORTING AND INCIDENT RESPONSE CAPABILITIES

For a more detail around these capabilities, please refer to Paper 3 in this series, ‘The Need for Preparation to Report Personal Data Breaches in a Timely Manner’, an in-depth analysis on how Forcepoint’s DLP and Insider Threat technologies 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.
    
    » This means that sensitive data can be masked or anonymized so that system administrators and incident responders are not able to see the sensitive data that is the subject of the violation referenced within the report, or identify the users or processors involved in the policy violation.
    
    » Organizations can then put in place comprehensive policies and processes around who can access the non-anonymized data. So, for example, only with the authorisation of HR, legal or the Workers Councils can the responder ‘unlock’ the identity of the user behind the breaches or view the PII data that was the subject of the incident.

» Reporting, logging and alerting:
  
  • The Forcepoint DLP solution centralises the policy management and reporting within the Forcepoint Security Manager. This is the same console where existing Forcepoint customers are able to manage Web and Email security gateway policies, too.
  
  • With over 20 prebuilt report templates specifically for data in motion, 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:
  
  • Forcepoint was first in the industry to release 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 Security 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 monitor PII data flows. This could be useful as part of a broader classification project, or a cloud migration project. For example, data can be marked as “sensitive: internal storage only” for data that must not leave specific local areas, or “sensitive: specified data center only” for data that can be migrated to specific or approved cloud providers only. Forcepoint’s DLP solution can then make policy-based decisions on data in motion or in use, based on the classification criteria.

Uniquely, Forcepoint’s technology partnership with Microsoft means Forcepoint DLP integrates with Microsoft’s Rights Management and Azure Information Protection (data classification). This allows
Forcepoint DLP customers to extend their existing DLP policies to include protected files and make policy-based decisions using classification tags.

Forcepoint DLP can also integrate with third-party security gateways like email, web or CASB. These third-party gateways can then be configured to ask Forcepoint DLP to help them make policy-based decisions around the flow of personal data across these gateways.

Forcepoint DLP can also integrate with other third-party data protection technical measures, such as email encryption providers, to provide deeper visibility or policy-based remediation of specific violations of privacy found within personal data flows.

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

### 6 Detail and guidance around configuring 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 mapping, monitoring, and control of personal data flows.

**CONFIGURING FORCEPOINT DLP FOR GDPR POLICIES**

Prior to mapping out data flows, 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 developed by our experienced, in-house policy research team.

These policies are provided to identify and classify personal information and address data protection regulations and legislations across 88 countries and territories around the world.

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.

![PII policies](image)

*Figure 5 shows many of the pre-defined policies available to an organization.*
Below you can see a breakdown of personal data classifiers that are relevant to the EU 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 all 28 EU countries and 61 countries total 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.
- All 28 EU country-specific set of classifiers.
- A further 12 country-specific set of classifiers for countries in EMEA but outside of the EU.

Refer to this link to see more detail on all of the classifiers available in the [FORCEPOINT DLP](www.forcepoint.com) solution by default.

### 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.
Lastly, it is important to be able to monitor, manage and control data flows with sanctioned cloud applications like OneDrive, natively through application programmable interfaces (APIs). This approach means that data can be protected regardless of who is processing the data in the cloud application, and on any device, managed or not (i.e. without the need for a DLP agent to be installed). Forcepoint Security Manager consolidates the management of DLP policies with the Forcepoint Web and Email Security Gateways to provide streamlined operations. DLP policies can be written once and applied to multiple DLP elements and content gateways, including web and email gateways and third party gateways, including CASB solutions, to monitor cloud applications as well as endpoint devices.

Figure 7 shows how DLP policies monitoring data flows can easily be applied to the integrated web, email, endpoint and cloud channels from within the Forcepoint Security Manager.

Figure 8: An incident where text within an image is detected as personal data.
DLP policies created to look for personal data flows can simply be re-used across any DLP module with minimal configuration. In addition, DLP policies can be applied to specific URL categories or to encrypted HTTPS streams when integrated with the Forcepoint Web Security Gateway.

Forcepoint DLP, when used in conjunction with the Forcepoint DLP agent, can be configured to encrypt sensitive data being copied to a USB device so that the encrypted data can only be decrypted if that USB device is plugged into any managed device, or optionally, through password-based authentication.

**REVIEWING POLICY VIOLATIONS AND ALERTS FROM MONITORING PERSONAL DATA**

The Forcepoint Security Manager displays the results of monitoring data flows, highlighting where personal data is found.

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

- The source of the data flow (IP address, email address, user, etc.)
- The destination of the data flow
- The channel (application file access, clipboard copy and paste, end, endpoint printing, endpoint LAN, HTTP, HTTPS, network email, endpoint email, FTP, File Sync and sharing...)
- The action taken (e.g., permitted, encrypted, blocked, quarantined, file deleted, confirmed)

Figure 9: Personal data can be masked or anonymized.

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 minimise false positives. Forcepoint’s endpoint DLP is able to apply fingerprint matching event when the device is off-network.

Another key capability Forcepoint DLP provides for our customers is to be able to anonymize personal data and user data, so that only those authorised to see this data have access to it.

Figure 9 shows how personal data can be masked or anonymized, preventing the responder from seeing the personal data that is the subject of the incident, or the user behind the data incident, unless authorised to do so. Forcepoint DLP customers are able to use our built-in hierarchical management capabilities in specific policies and processes so that there are multiple business functions involved in the approval process.

Wherever personal data is found in high-risk data flows (e.g., in personal webmail accounts or in public file sync and share services), the incident can be put into the workflow to start the remediation process.
Remediation actions may include:

- Escalation (to manager or another person)
- Endpoint remediation actions may include:
  - Move
  - Delete
  - Encrypt
  - Apply DRM
  - Apply masking
  - Apply categorization
  - Apply pseudonymization (for a test system for example)

More information regarding the incident management and response will be provided in Paper 3, ‘Preparation to report personal data breaches in a timely manner’.

7 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 paper 1 entitled, ‘Inventory of Personal Data’.

To learn which technologies can assist organizations’ response to data breaches in a timely manner (within 72 hours of the controller becoming aware of the data breach), please see Paper 3, ‘Detect & Respond To A Data Incident’.