ACTIVENAV

Managing Data Subject Rights Requests

The Complexity of Unstructured Data Discovery

Professor Geoff Smith
geoff@proprivacy.guru

in association with ActiveNav
www.activenav.com

Abstract

This document provides a comprehensive guide to managing the unstructured data component of Data Subject Access Requests (DSARs)—a critical aspect of regulatory compliance under frameworks such as GDPR and CCPA. It outlines a robust, repeatable workflow encompassing intake, validation, discovery, collation, and secure delivery of data to ensure organizations meet their transparency and accountability obligations.

Particular focus is given to the challenges posed by unstructured data, which remains the most significant hurdle in DSAR fulfilment due to its complexity and the lack of effective tools for discovery, integration, and redaction. The document examines the tools and strategies available to privacy professionals, highlighting their limitations—especially in managing unstructured data, where manual oversight is essential to ensure accuracy, compliance, and the protection of third-party privacy.

Additionally, the growing importance of privacy programs is explored, with an emphasis on the integration of technical and operational solutions as organizations increase their investments to tackle these challenges. Market trends show that many organizations are turning to third-party privacy technology vendors to enhance DSAR automation and compliance. However, significant gaps remain in the discovery and preparation phases, resulting in delays, higher operational costs, and risks of non-compliance.

The document concludes with actionable recommendations for adopting scalable, cost-effective tools and workflows to enhance DSAR fulfilment. ActiveNav's specialised unstructured data discovery solutions are presented as key components for ensuring privacy program success and addressing the complexities of unstructured data management.

Introduction

The introduction of data protection laws, including the General Data Protection Regulation (GDPR), the Data Protection Act 2018, and the California Consumer Privacy Act (CCPA), has led to a significant increase in individuals, known as "Data Subjects," exercising their rights under these regulations. The GDPR greatly expands individuals' right to access personal data. It is crucial that these requests are managed fairly and within specified timeframes, ensuring that the exercise of these rights does not conflict with other data protection responsibilities, such as safeguarding the privacy of third parties and upholding confidentiality obligations. The matter can be extremely challenging where the organization believes that the disclosure sought is effectively a means to start litigation or act against them, their employees, agents or partners.

Data Subject Access Requests DSARs can be inherently complex. Professionals, including those in data protection, often have differing views on handling DSARs, such as determining when redactions are necessary. This makes processing these requests time-intensive and resource-demanding. Research, recently conducted by the DPO Centre , into consumer perceptions of how companies manage personal data revealed that although only 1 in 10 respondents had considered submitting a DSAR, 44% felt companies were mishandling their personal information. Other research from EY Law and Statista have seen 24% increase in DSARs year on year since 2022. This suggests a significant potential for an increase in DSAR volume with experts predicting that DSARs will rise in the coming years. Compound this with data volumes in typical organizations increasing by 23% per annum and it is easy to identify the increasing challenges posed. Much of this data, up to 80%, is unstructured data such as: email, word documents, text files, web content or video/ voice recordings. These are typically one-use items and often referred to as ‘dark data’ due to their lack of discovery or use. These formats are consumed by humans, often unindexed or tagged and can be challenging for computerised systems and tools to discover. Records of Processing Activities (RoPA), under Data Protections laws such as GDPR, requires an organization to maintain detailed records of their processing activities, listing all systems, and databases where personal data is handled. The ROPA serves as a guide to ensure all locations where personal data is processed are reviewed ensuring a thorough response to a DSAR. This in turn cab demonstrate an organizations transparency and accountability, two increasingly prominent factors in brand evaluation and consumer trust.

In this paper we explore the complexities in handling and fulfilling Data Subject Rights Access requests, with particular focus on the discovery and collation of unstructured data. Whilst not an exhaustive list of recommendations it does seek to provide guidance and considerations into the effective handling of requests and things that need to be addressed. To establish a minimum viable product when dealing with the challenges of discovering and collating unstructured data formats. Addressing this challenge involves a two-pronged approach. First, companies should establish procedures and best practices within their data processing operations to make handling complex DSARs more manageable. Second, they should leverage technology and tools that streamline these processes and improve efficiency. By adopting the right procedures and implementing suitable tools, companies can significantly reduce the burden and risks associated with DSARs, leading to time and cost savings.

Data Privacy Regulation & Guidance

DSARs are an important part of the wider data protection framework and are recognised in Article 8 of the EU Charter of Fundamental Rights: “Everyone has the right of access to data which has been collected concerning him or her”. The right helps data subjects to verify the accuracy of any personal data held about them and the lawfulness of the processing of that data.

The substance of the right is set out in Articles 12 and 15 of the GDPR. Data subjects do not need a reason or justification to make a request, and requests are free (unless manifestly unfounded or excessive). They are entitled to specific information about the processing of their personal data, such as details of the purposes of the processing and any retention period. Most importantly, data subjects are also entitled to a copy of the personal data being processed. Organizations are expected to respond without undue delay and within 30 days. Extensions of up a further 60 days can be applied in exceptional circumstances where the requests are complex to the organization.

Recent guidance from the European Data Protection Board (EDPB) confirms that unstructured electronic information such as emails, CCTV and telephone recordings fall within the scope of the DSAR. It also suggests there is no proportionality constraint of the effort needed to search for personal data. This presents significant practical challenges to organizations dealing with undefined requests for content that can be difficult to manage on a case-by-case basis. Indeed, it is often cited as a reasonable expectation that if consumers can make a one-click purchase of goods that will be delivered the following morning, why do they have to wait weeks simply to see their data?

Data however can be a difficult beast to manage, multi formatted, multi versioned and complex. The reality is that in many cases organizations are likely faced with employing manual intervention to manage requests for large undefined, unknown and unstructured content. Whilst 'technically feasible', as described in EDPB guidance is a broad term and open to much interpretation, technologies exist to support organizational structuring and metadata tagging of unstructured data and information assets. The consideration for most organizations is whether they have done enough to manage the risk. Whether they have done enough to govern data sufficiently and in line with expectations of their customers and regulators.

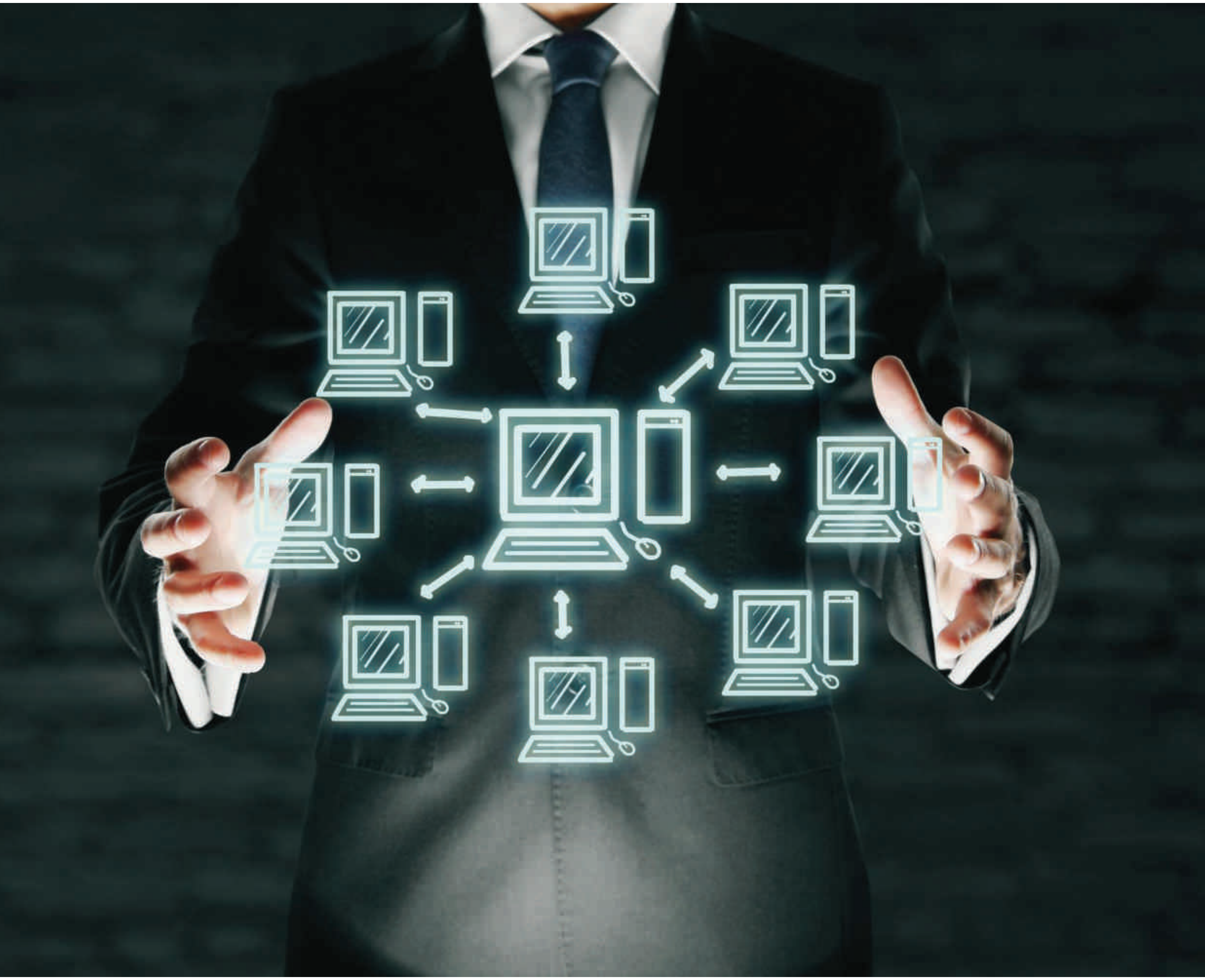
The Significance of Data Curation and Classification

Ideally the method of capturing a request from a data subject should reflect the different channels through which organizations interact with their users. For example, an online portal may limit capture through a web form, but a retail outlet chain would also provide opportunities for data subjects to raise requests in-store. Many emerging digital organizations offer data subject access requests as a feature for consumer experience with self-service portals and privacy centres becoming more common. These services typically provide an API or similar integration to serve data into consumable tables, some even offer personal data vaults or stores for direct consumer management. However, for many others this can prove difficult due technical legacy or other constraints. Access is usually requested through an email directed to the Data Protection Officer. This includes where the data subject may have a complaint against the organization and they are seeking to obtain evidence or where there is legal representation, for example following a cyber incident or data breach.

It is not just consumers of service users that have the right to request a DSAR. Recent research from Privacy Engine indicates that 66% of DPOs stated that they had seen an increase in employee DSARs and that these requests, often received by email, involved searching through diverse data sets including complex employment records, emails and internal messaging systems. This review process typically requires a manual assessment, often using external specialist resources. While keyword searches and deduplication tools can reduce the review burden, it isn't feasible to automate the task of discerning whether a mention of "Mr Smith" refers to the data subject in question or if other individuals' personal data needs redaction.

Even when narrowing search criteria (such as by limiting searches to specific custodians or date ranges), the effort remains costly and labour-intensive. A recent Gartner survey identified that most DSARs take more than 14 days to respond to with an average cost of £1,400. However, some subject access requests require over 1,000 hours of manual review. In one notably complex case, the Nursing and Midwifery Council in the UK spent £240,000 responding to a single access request.

Whilst cases like this are possibly in the extreme, a study conducted on behalf of Truyo , a privacy and compliance AI firm, featuring responses from privacy professionals who worked at companies with more than 1,000 employees identified the level of anxiousness at organizations. With 92% stating concern about honouring data subjects' rights under GDPR and CCPA. A further 51% said data subject right fulfilment was the most difficult part of privacy regulation compliance. It is little wonder that many organizations faced with this level burden and potential cost may seek to delay or even dissuade data subjects' requests through a triage process of re-clarification and identification or simply providing an initial set of easily accessed data that does not necessarily include complete copies of emails or file documentation.



Process

While the right to access personal data may seem straightforward, it can be challenging to fulfil in practice, especially when dealing with unstructured electronic information. Responding to broad requests for "all" personal data held in such formats can be highly complex, even impractical, due to several factors:

Volume of data:

Unstructured data sets in large organizations can be immense, sometimes including hundreds of millions or even billions of emails. This isn't typically due to lax data retention but can simply reflect the scale of the organization and regulatory requirements for record-keeping. Searching these extensive data sets presents significant logistical challenges.

Back-up and Archiving:

Data is often stored across various formats, including live data, backups, and archives. Guidance suggests that archived data should be included in search efforts and, where "technically feasible," organizations responding to a request may need to review backups for any data 'not present' on the live system. Restoring backups is usually costly and time-intensive, rarely making it a proportionate response.

Lack of Indexing:

With unstructured data, quickly and accurately locating information about a specific individual is challenging. Unlike structured databases with unique identifiers for each person, unstructured data may reference individuals in ambiguous ways, such as referring to "John Smith" as "John," "JS," or "Mr Smith." Additionally, not every mention of "Mr Smith" will pertain to the individual requesting access.

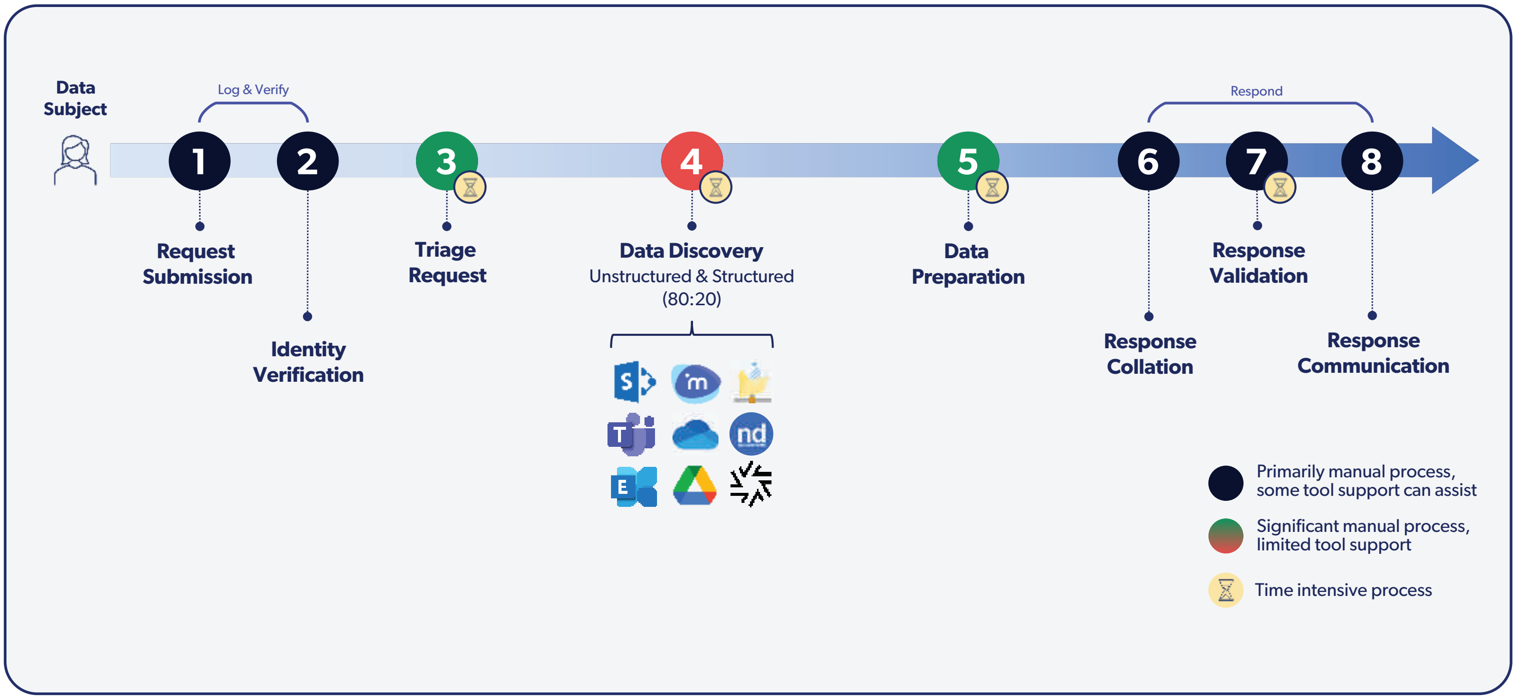
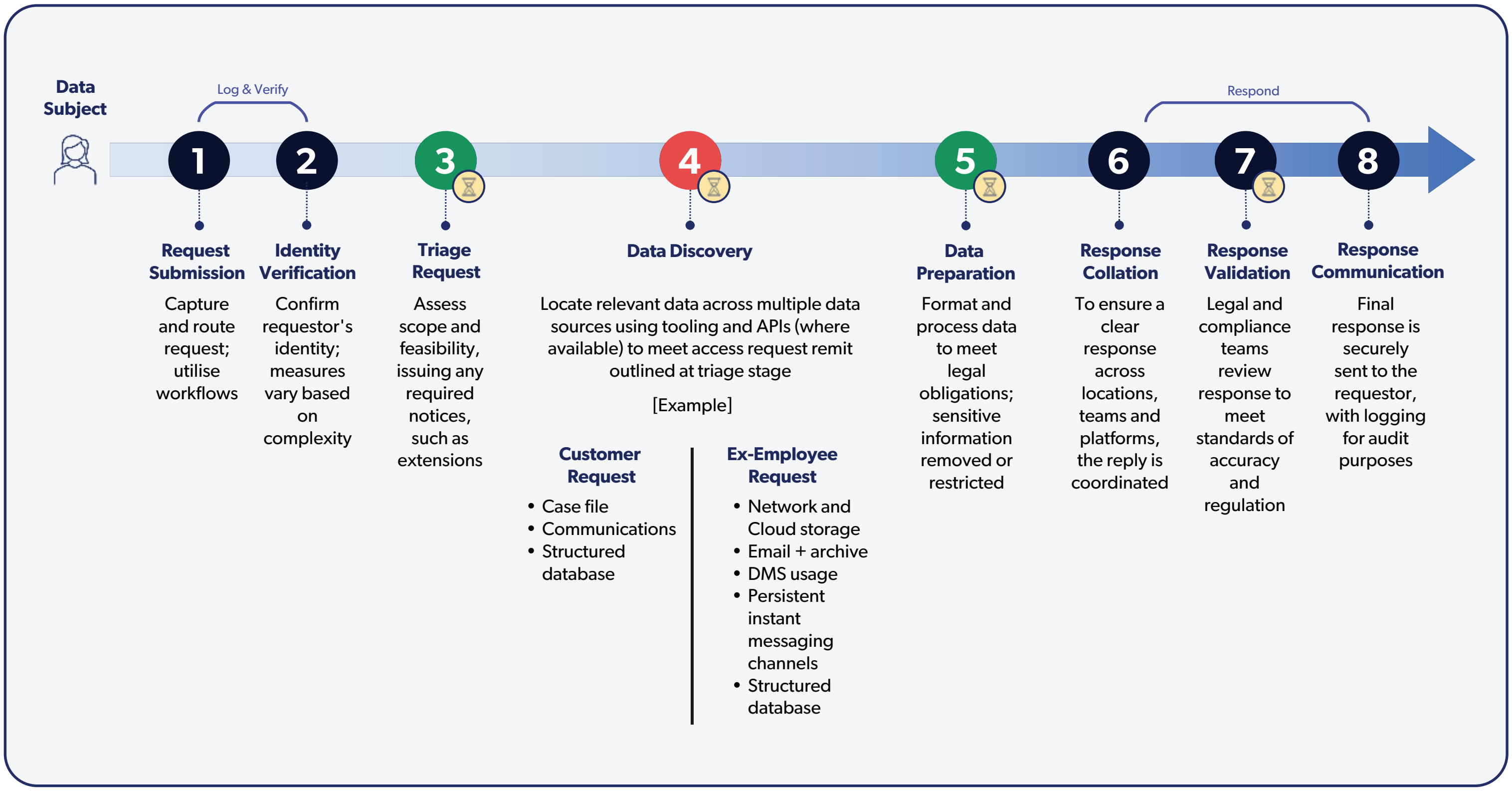


Figure 1. DSAR Fulfilment Process

Organizations face significant challenges when sifting through both structured and unstructured data stores—whether housed on-premises, in the cloud, or with partners and sub-processors. Beyond simply discovering and retrieving data, they must also redact personal information related to other individuals, ensuring one person's rights aren't compromised in responding to another's request. Therefore, to keep the process manageable and efficient, fulfilling these requests requires a scalable, repeatable approach.



Request Submission
Capture and route request; utilise workflows

Identity Verification
Confirm requestor's identity; measures vary based on complexity

Triage Request
Assess scope and feasibility, issuing any required notices, such as extensions

Data Discovery
Locate relevant data across multiple data sources using tooling and APIs (where available) to meet access request remit outlined at triage stage

[Example]

Customer Request

- Case file
- Communications
- Structured database

Ex-Employee Request

- Network and Cloud storage
- Email + archive
- DMS usage
- Persistent instant messaging channels
- Structured database

Data Preparation
Format and process data to meet legal obligations; sensitive information removed or restricted

Response Collation
To ensure a clear response across locations, teams and platforms, the reply is coordinated

Response Validation
Legal and compliance teams review response to meet standards of accuracy and regulation

Response Communication
Final response is securely sent to the requestor, with logging for audit purposes

Figure 2. DSAR Fulfilment Process Detailed

The process of fulfilling a Data Subject Rights (DSR) request involves several key steps, from initial request receipt to final delivery of redacted data. Here's a detailed breakdown of each stage:

1. Receive Submission

- Log Request: Time stamp the validated request to ensure compliance with mandatory time scales are met.

2. Identity Verification

- Verification of Identity: Confirm the identity of the data subject to ensure the request is legitimate and data privacy is maintained.

3. Triage Request

- Request Assessment: Determine the specific rights being requested (e.g., access, deletion, rectification) and confirm if the data subject is eligible based on applicable regulations (e.g., GDPR, CCPA).

- Clarification of Scope: If the request is broad, work with the data subject to narrow the scope where possible, which helps in minimizing unnecessary data processing and speeds up fulfilment.

4. Data Discovery

- Structured Data Search: Begin with structured databases (CRM systems, HR systems, financial databases) where data is more organized and identifiable. Use unique identifiers associated with the data subject, such as employee ID, customer number, or email address.

- Unstructured Data Discovery: This phase is typically the most challenging part of the process:

    - Identify Relevant Repositories: Determine where unstructured data may be stored, including email servers, shared drives, document management systems, collaboration tools, and cloud storage.

    - Keyword and Metadata Searches: Use keyword searches with names, email addresses, and other unique identifiers. Metadata, such as timestamps and custodian information, can help refine the search scope.

    - Natural Language Processing (NLP) and AI Tools: If available, leverage NLP and AI tools to sift through unstructured data more effectively, identifying text patterns and relationships that may relate to the data subject.

    - Automated Deduplication: Use deduplication software to reduce repetitive or redundant data, helping streamline the review process.

5. Data Preparation

- Data Relevance Assessment: Review the collected data to confirm its relevance to the data subject's request. Filter out irrelevant or unnecessary data that falls outside the request's scope.

- Legal and Compliance Review: Verify compliance requirements, identifying any legal constraints on data disclosure. Some data may be exempt from disclosure, such as information that could reveal company trade secrets or the identities of other protected individuals.

- Unstructured Data Redaction: For unstructured data (e.g., emails, documents, PDFs), redact personal information pertaining to other individuals or sensitive company information. This often requires manual review, as automated tools may lack the specificity to distinguish between similar names or ambiguous references.

- Automated Redaction Tools: Where possible, use automated redaction tools for structured datasets or well-organized unstructured data files. AI-driven redaction tools can help identify and mask common identifiers, such as names, addresses, or unique IDs.

6. Response Collation

- Multi-level Review: Conduct multi-level reviews to ensure redactions are accurate and complete, preserving the privacy of unrelated individuals and ensuring compliance.

7. Response Validation

- Final Review for Completeness and Accuracy: Perform a final review to ensure all relevant data has been collected, validated, and correctly redacted. This step is critical for preventing inadvertent exposure of non-requested or sensitive information.

- Compliance Check: Confirm that the response meets all regulatory requirements and internal data protection standards.

- Approval from Data Protection Officer (DPO): For complex or high-risk cases, secure approval from the DPO or another designated authority before delivering the response to the data subject.

8. Response Communication and Recordkeeping

- Data Delivery to the Data Subject: Provide the data to the data subject through a secure channel, ensuring they can only access their requested information. Common methods include secure file transfer protocols (SFTP) or secure portals.

- Confirmation of Receipt: Request confirmation from the data subject that they have received the requested information. This can help close the request and ensure no further actions are necessary.

- Document Retention and Recordkeeping: Maintain a detailed log of the request, including data sources searched, actions taken, redactions made, and the final delivery. This documentation is essential for audit and compliance purposes, should any questions arise about how the request was handled.

By implementing these steps in a repeatable process, organizations can improve efficiency, minimize errors, and meet regulatory requirements.

Without doubt the most challenging and time-consuming aspect of DSAR fulfilment is the discovery and collation of unstructured data which presents several significant obstacles in discovery, integration and redaction. The overwhelming concern for DPOs and Privacy Professionals is meeting compliance expectations of regulations. Recent research on behalf of ActiveNav found that over 80% of Privacy and Legal professionals cited challenges in discovery of unstructured data discovery as a reason for delays in responding to a subject access request. EY Law survey conducted in 2022 also found that 51% of DPOs had received complaints from individuals about their DSAR responses.

◦

Tools & Resources

Most organizations run privacy programs to manage risks associated with compliance to privacy regulations and growing market expectations. Research from International Association of Privacy Professionals in 2024 indicated that organizations were increasing investment into their privacy programs at a level of 12% per annum, with average investment between $1.2 million and $9m depending upon their size and complexity, and that privacy teams were growing. Many privacy functions are also "shifting left" in other words to incorporate and embrace technical operations and engineering. That said many budgets are still shared between legal and technical functions with much of the funding for solutions belonging to the Chief Information Officer of Chief Technology Officer.

After all, Privacy compliance is a 'team sport' and not just the concern of those professionals within the privacy team. Responses to DSARs involves several different team members within its workflow.

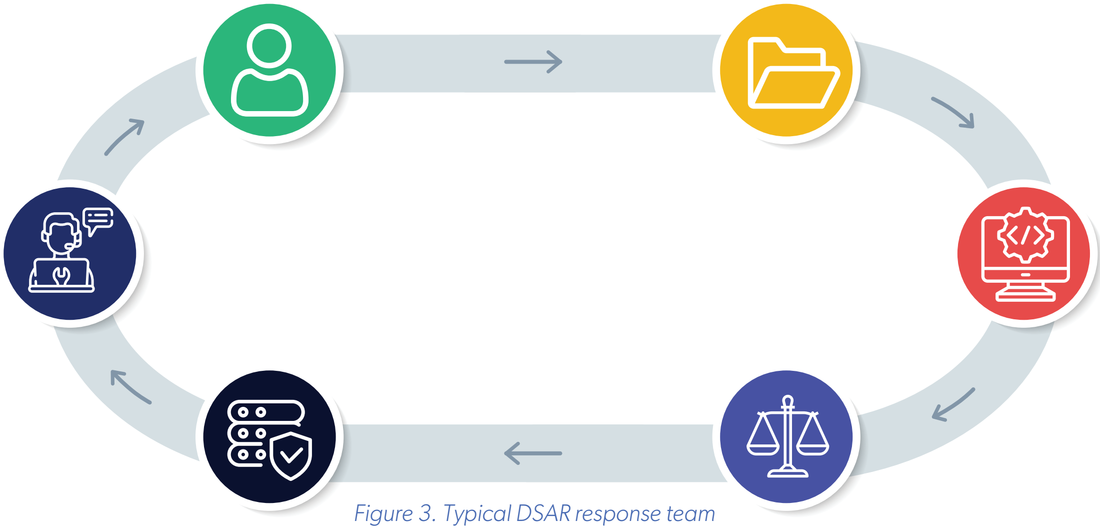


Figure 3. Typical DSAR response team

Data Privacy Officers oversee compliance with data privacy regulations, ensures DSARs are handled accurately, and approves responses before data delivery. The DPO also handles escalations for complex cases, ensuring unstructured data is processed according to privacy policies.

Privacy Analysts are responsible for identifying, collecting, and organizing data relevant to each DSAR. They use discovery and classification tools to locate data across various systems and validate data relevance.

Data Management or IT Specialists who support privacy analysts in accessing various data repositories, especially for unstructured data stored across multiple systems. They handle data extraction and assist with integrating data from diverse formats.

Legal and Compliance Teams review unstructured data for any legal or compliance implications. They assess data for possible exemptions, such as confidential business information or third-party privacy, and ensure legal compliance.

Data Security and Redaction Specialists focus on securely redacting sensitive information in unstructured data. They manage redaction tools and conduct quality assurance checks to verify that all sensitive information has been adequately masked.

Customer Service or Support Teams communicate with data subjects, helping them understand the DSAR process, answer any questions, and clarity requests.

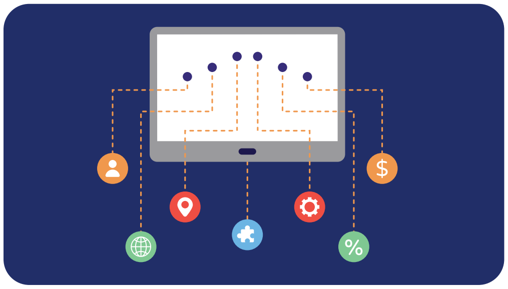
When assessing what solutions best help their organization respond to the requests, organizations are turning to third parties with 56% of DPOs questioned in recent IAPP research saying they have purchased tools from a privacy tech vendor rather than build an in-house solution. The report also found a gap between how legal professionals and IT teams understand privacy technology, particularly regarding automation. Only 13% of IT professionals said the solutions their organizations used were fully automated whereas 55% of legal and compliance teams stated their DSAR capabilities are completely automated. Much of this chasm of understanding could be a result of how privacy platforms have evolved. Whilst case management systems at the front end of an enquiry can appear automated, much of what happens particularly in the discovery phase is subject to a high degree of intervention, especially in consideration to the collation of unstructured data.

At the heart of the DSAR automation market are three key capabilities:

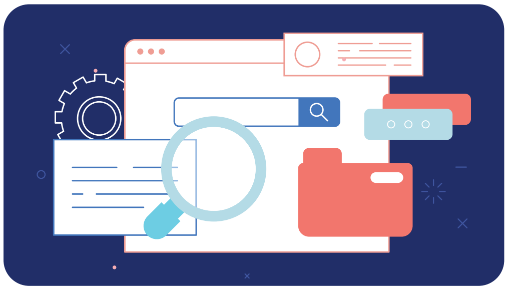
1. Discovery of existing information held on individuals, and continuous monitoring for changes to data stores and new systems that are being onboarded.
2. Maintenance of the capacity to act on that information should the data subject request modification, deletion or restriction of processing.
3. Tracking of request workflows and holding of detailed records to gauge effectiveness and demonstrate compliance.

Expanding this further for Data Subject Access Requests DSARs in workflows that handle unstructured data requires specific tools for discovery, case management, and redaction.

Unstructured Data Discovery Tools



Data Mapping and Classification Software: Tools that help locate unstructured data across various sources (e.g., emails, shared drives, cloud storage) by categorizing and tagging personal data. AI-driven classification software enhances discovery accuracy by identifying patterns associated with personal identifiers.



AI-Powered Search Tools: AI and natural language processing (NLP) tools aid in searching unstructured data sources by identifying keywords, phrases, and context, making it easier to locate relevant information across diverse file types and formats. Whilst viable on small sets of data these solutions can prove costly when operating over large datasets or unknown 'dark' repositories and delivery uncertain results.

Data Discovery Platforms: Platforms like ActiveNav Cloud offer comprehensive unstructured data discovery, enabling privacy teams to locate, inventory, and catalogue personal data across data silos.

Case Management Tools

- DSAR Management Platforms: There are several tools available on the market that can provide centralised case management, allowing teams to track requests from intake to fulfilment. They automate workflows, assign tasks, set deadlines, and track the status of DSARs, helping teams stay compliant with regulatory timelines. Often seen as ‘one stop’ solutions these tools provide useful case and risk management reporting. However, with regards to automation and data discovery they are dependent upon much configuration and/or integration with other third-party systems, typically through APIs. This can lead to expensive set up and operational costs without necessarily fulfilling the challenge of complex data discovery and preparation.

- Ticketing and Workflow Automation: Workflow and ticketing systems integrate with DSAR management platforms to streamline task assignment, automate case routing, and ensure requests progress smoothly through review and approval stages.

Redaction Tools

- Automated Redaction Software: There are tools to automate the redaction of sensitive or irrelevant information in unstructured data. AI-powered redaction tools can identify personal data (e.g., names, emails) and other sensitive information for automated or manual redaction. Working efficiently on assembled documentations or data sets these tools usually provide add-on functionality in preparation of a DSAR submission following the discovery process.

- Manual Redaction Software: When unstructured data requires human oversight, PDF editors or redaction-specific software allow professionals to manually mask specific information within documents and file.

Whilst privacy compliance platforms offer solutions for monitoring subject rights requests and oversight of data inventories, when it comes to specific unstructured data discovery gaps in the market still exist. Handling unstructured data for DSARs remains challenging due to several technology gaps and limitations. These challenges and limitations include limited data discovery capabilities, manual-heavy redaction requirements, fragmented data management systems, insufficient de-duplication or data minimisation, complex context tracking, and lack of real time data processing. These gaps in turn impact the speed, accuracy, and compliance of DSAR processes, making it difficult for organizations to meet regulatory requirements and provide timely, complete responses. Furthermore, a DSAR can lead to further data subject rights requests and actions such as right to be forgotten or rectification of inaccurate data. So just as one spotlight is dimmed others are switched on.

Recommendations

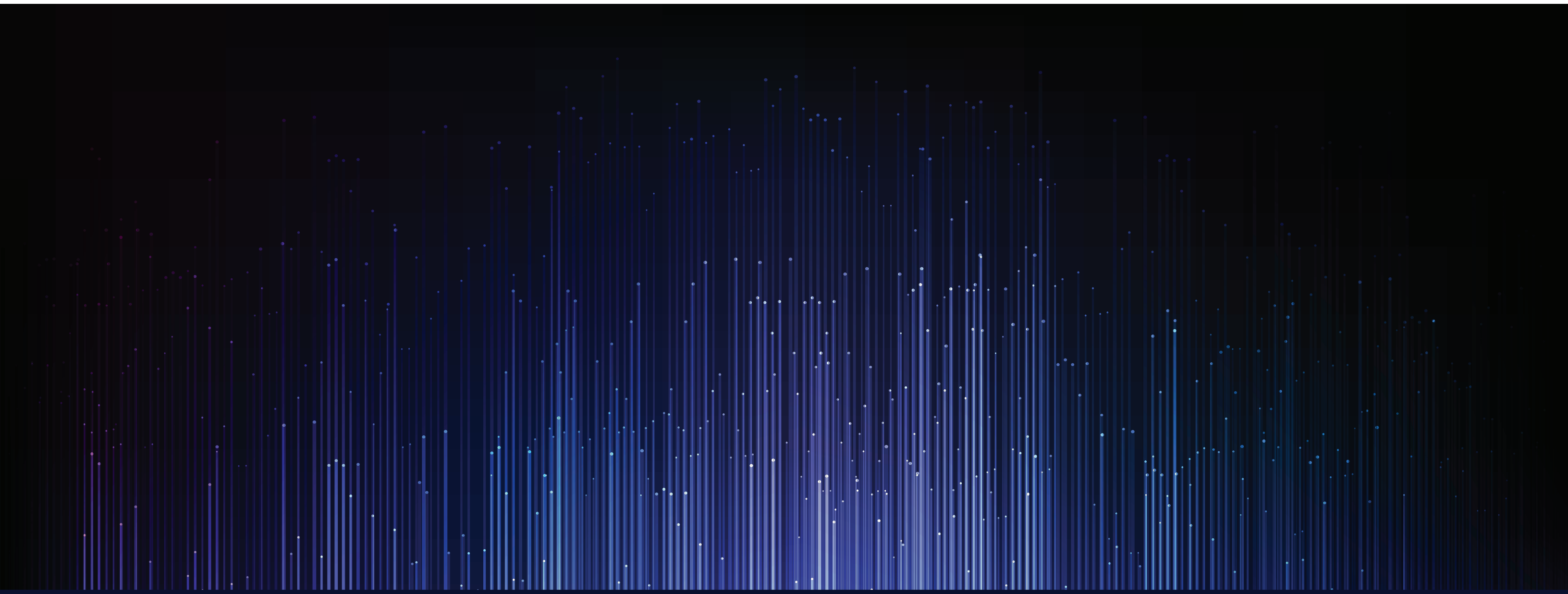
Addressing technology gaps requires investment in advanced data management solutions and a mix of manual and automated approaches for unstructured data. Enhanced AI, NLP, and centralized privacy management platforms are evolving to help bridge these gaps, but organizations must remain vigilant in combining technology with robust workflows and trained personnel. Developing a hybrid strategy of automation, specialised tools, and manual oversight can help organizations better meet DSAR requirements, improving speed, accuracy, and regulatory compliance.

With 74% of respondents, in a recent survey by European Centre for Digital Rights styled as "noyb" from "none of your business", expressing concerns about companies' privacy compliance. increasing drive for operational efficiency and the risks associated with regulatory enforcement increasing, specialised unstructured data discovery solutions remain highly relevant. Their role in data privacy compliance remains threefold, in support of timely and comprehensive data subject rights fulfilment, inventory and cataloguing of personal data in unstructured formats, discovery of data in response to a breach. They form part of every privacy program's essential solutioning fulfilling not only DSARs but also enabling ROPA maintenance, data deletion and comprehensive data breach response.

To meet the growing regulatory challenges and consumer expectations organizations often need a combination of advanced tools and manual processes to overcome these challenges effectively. Solutions such as AI-enhanced data discovery, automated deduplication, and flexible redaction software can streamline parts of the process. However, due to the inherent complexity of unstructured data, manual oversight remains critical for ensuring accuracy, protecting third-party privacy, and ensuring compliance with data protection regulations. This in turn can drive operational costs and financial risks to unpalatable levels.

With questions remaining about the accuracy and completeness of data subject rights fulfilment and increasing dissolution of consumer trust. The need for out of the box, configurable, scalable and low cost of ownership tools is evident. Unstructured data discovery specialist tools are essential to assuring data governance and privacy programs. Activenav provides this solution.

Fulfilling DSARs can be complex, costly and at risk of further legal action, enforcement or simply loss of consumer trust. To ensure your organization meets it commitments to both transparency and accountability contact our experts at ActiveNav.



Americas
Reston, VA, USA
+1 571 375 2780

EMEA
Winchester, UK
+44 01962 454401

APAC
Melbourne, Australia
+61 3 9982 4543

Contact sales@activenav.com
LinkedIn activenav
Visit activenav.com