Recommendations on How to Tackle the 'D' in GDPR

Table of Contents

1. Executive Summary

The European Union began enforcement of the General Data Privacy Regulation (GDPR) as of May 2018, which afforded enhanced protection to personal data. The GDPR applies to any organization established in the EU and to any organization (anywhere in the world) that processes the personal data of EU data subjects when offering them goods or services or when monitoring or tracking their activities. This regulation could have significant impact for many organizations and how they manage data pertaining to customers, consumers, partners, staff, and other "data subjects"; where a "data subject" is an individual. The GDPR impacts the storage, processing, access, transfer, and disclosure of an individual's data records and potentially has some very large penalties for violations.

The GDPR requires organizations to fully understand how they use current and future information assets to incorporate these data privacy requirements and enhance citizens' privacy rights. For many, the associated changes to information management practices will require a thorough evaluation of current and future data capabilities. This paper explores how breaking down these requirements helps aid the understanding of the data challenges and the direction organizations could take around their GDPR initiatives.

To aid understanding, this paper looks at common questions many organizations ask on their GDPR journeys. We call these entry point questions. To help answer each entry point question, we have laid out a set of capability requirements that we consider important and, aligned to each capability, is a technology use case for how each capability can be developed. The table below shows how these items are all related.

Entry Point Question	Capability Requirement	Technology Use Case
Where is all our potential in-scope data?	Sensitive Data Discovery and Risk Analysis	Detect and Protect
How is personal data being used?	Policy Interpretation	Enterprise Data Governance
How do we manage data subject data?	Personal Data Management	Data Matching and Linking Use Case
How do we secure data and prevent unauthorized access?	Enabling Data Security Control	Detect and Protect

There are also examples where requirements, such as consent capture and management, may span multiple capability requirements and technology use cases; so, organizations need a clear understanding of the potential complexities involved.

While the GDPR poses many challenges, it also offers many new opportunities around the safe use of data when lowering risk exposure. This paper outlines potential use case approaches and draws on our depth of experience in data management to help organizations simultaneously address these challenges and introduce innovative data governance and security capabilities to maximize their privacy compliance programs. Informatica delivers integrated and innovative software solutions to automate, secure, and control data, and these solutions can quickly support organizations on their GDPR initiatives and scale beyond through a consistent approach.

2. Background

2.1 General Background and Potential Implications

The digitization of society is proceeding at a rapid rate, with almost every organization leveraging the power of data to improve business decisions, engage customers and partners, and drive transformational business processes. The European Commission has recognized that much of the data being created, collected, processed, and stored is in fact personal data, which can reveal extensive information of EU data subjects.

Pre-existing data protection regulations had not necessarily reduced concerns regarding the protection and safety of personal data. Diversity of data protection regulations across the EU member states frustrates data subjects, with 90 percent indicating that they would like the same data protection regulations across the EU—regardless of where their data is stored or processed.*

Therefore, the GDPR was enacted to better protect citizens' fundamental privacy rights in the digital age and address concerns regarding diversity of data protection laws.

Beginning in May 2018, the GDPR requires organizations to more effectively manage and protect data on customers, citizens, staff, and others. This regulation applies to EU data subjects, regardless of nationality or residence, to provide principles and rules on the protection of personal data.

With GDPR being a "principles" based regulation, this means organizations must consider what obligations they may, or may not, need to meet given the unique circumstances of their business and use of data. Organizations will therefore need to create an interpretation of these principles to help guide and steer their GDPR initiatives.

The GDPR requires organizations to better understand how they will utilize their current and future information assets to comply with these new data privacy principles. This will have an impact on the people, the processes, the technology, and the data management practices and policies.

* http://ec.europa.eu/justice/data-protection/reform/index_en.htm

Violations to the regulation could have significant financial penalties depending upon the type and scale. For example, fines of up to €20M or 4 percent of an organization's total worldwide annual turnover (whichever is larger) may apply—there are several examples of fines in the tens of millions that have already been levied.

2.2 Who is the GDPR Relevant To?

GDPR compliance has multiple dimensions and is not limited by physical geography; organizations in North America, Asia, and others must comply if they store and process EU data subjects. Today personal data is handled by organizations that deal directly with consumers (B2C), organizations that deal with other organizations (B2B), as well as dedicated data processing companies. Organizations that process data on EU data subjects will need to thoroughly understand their compliance requirements, regardless of which country their operations or data centers are physically located in.

2.3 What Makes the GDPR Challenging From a Data Perspective?

For organizations, there are distinct data challenges in relation to the GDPR. Compliance with the GDPR implies control and governance of personal data wherever it is within an organization. However, the proliferation of data throughout organizations and their business ecosystems can make it an ongoing challenge to manage data. Significant trends, like an increase in data diversity and a move to cloud-based computing, create a highly dynamic IT landscape and add to the data management and security challenges. To demonstrate these challenges, many organizations continue to struggle to answer these types of GDPR-related questions:

- Where in any organization, and its ecosystem, is all the relevant and in-scope data that GDPR principles would apply? Is that data at risk?
- How do organizations keep track of their data across their operational ecosystems?
- How does an organization define and manage all its relevant data assets to help ensure all necessary policies and procedures are applied and enforced?
- Where in any organization are all the relevant in-scope data records held that the GDPR principles would apply? How can these be identified and linked?
- How does an organization capture and manage the consent provided by a data subject? How can an organization manage changes to the data subject's choice of consent or manage the definition of consent?
- How can an organization efficiently and effectively respond to subject access requests, right of erasure, and portability requests within the required time frames?
- How does the organization control access to the relevant data? Is personal and sensitive data removed when it is not required for the organization's function or activity?

2.4 Types of Data Potentially in Scope

Another potential challenge is how organizations respond to the types of data they hold. In this context, we define types in two ways:
1. A data entity type
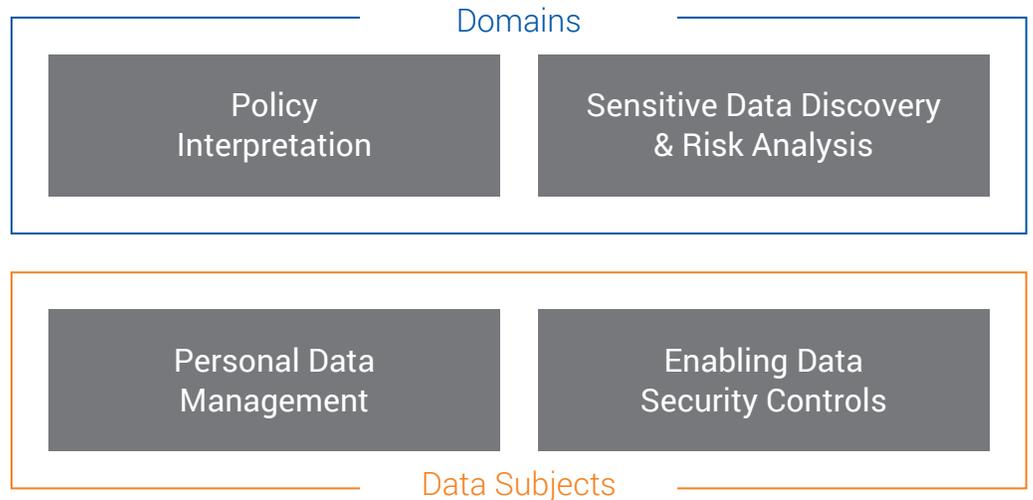2. A technology type that manages the data entity type

Most pieces of data subject information would fit into one or more data entity types, as well as one or more technology types. The diagram below shows some examples of potential data and technology types that may apply to in-scope GDPR data:

DATA ENTITY	TECHNOLOGY				
• Customer • Client • Policyholder • Beneficiary • Contact • Employee • Contractor • Volunteer • Visitor • Others	• Structured • Semi-structured • Unstructured	• Online • Near line • Offline • Back up	• Digital • Physical • Combined	• Explicit • Implicit • Combined	• Internal • External

These different types may require organizations to consider very different approaches, methods, and technologies for the capture and management of in-scope GDPR data assets.

3. Entry Points, Capability Requirements, and Technology Use Cases

To help drive understanding and awareness, as well as aid activity planning, Informatica has identified several key entry point questions that highlight common GDPR data challenges. These entry points are often driven by simple questions that may require organizations to carefully consider the people, processes, and technology they need to produce the answers. To help support answering these questions we have outlined the potential capabilities required, as well as examples of technology use cases that deliver the required capabilities. The capabilities required are structured into groups; the diagram below shows how this grouping works and the relevance of each group.

Domains

Policy Interpretation	Sensitive Data Discovery & Risk Analysis

Personal Data Management	Enabling Data Security Controls

Data Subjects

These capabilities sit within two areas called domains and data subjects.

Domains relates to the domains of data subject data. It helps provide insights into domain discovery and management, which is used in defining scope and providing an organizational view of data.

Data subjects relates to the actual data subject data at a transactional level. It helps provide insights into personal data management, which is used to provide subject-level responses and subject-level insights.

3.1 Entry Point Question: Where is All Our Potential In-Scope Data?

Background: Data is usually scattered across many systems, applications, and sources across an enterprise. This is especially true for larger organizations, and those that have grown by acquisition. Due to the roles EU data subjects could play in an organization (customer, supplier, partner, employee, etc.), it is unlikely that personal data will be restricted to one department or system. Organizations with more diverse IT systems should not only consider data in core applications but also spreadsheets, local databases, and big data solutions.

Capabilities required: Sensitive data discovery and risk analysis is a capability to discover data across a wide range of technology solutions and using this, along with other sources of information such as the amounts of actual data and data proliferation, to create a risk score for data. The risk score helps organizations understand where the highest risk data is stored so that any potential remediation or security control requirements can be prioritized based upon severity or impact. Tracking the risk score over time shows whether remediation or control activities have mitigated the data risk position. In support of the lawful purposes, consent may be required so capabilities, such as data lineage, help organizations identify new stores of personal data to aid understanding of potential changes in use.

Technology use case: Sensitive data discovery and risk analysis could be characterized as being a detect and protect use case, with a focus upon the detect portion. These are core capabilities to provide insights into where in-scope sensitive data is and where it proliferates with analytical insights into data risk. Typical capabilities that could apply to this use case include:

- **Data policy definition:** Business and IT definitions, vague data, policy conflict
- **Automated data discovery:** Find relevant in-scope sensitive data, first pass plus continual monitoring, classification of data, supporting system integration
- **Data proliferation:** Where is the data? Where does it go? New sources?
- **Data risk scoring:** Based upon movement of data + proliferation + access + volume, prioritization plus planning, history, and score monitoring over time
- **Data protection:** Identify where data access needs restrictions, what data should be pseudonymized and anonymized, where should encryption be applied, and the viewing of data based on time, location, and role

Technology solutions: Informatica Secure@Source™ could be used to help discover the locations of in-scope data, classify the data, monitor data proliferation, and assign risk scores. Tracking over time shows how changes are positively or negatively influencing compliance efforts.

Benefit: Provide insights into not just the location of data but also rank data according to risk.

3.2 Entry Point Question: How is Our Personal Data Being Used?

Background: Our world is undergoing a digital transformation that is affecting all sectors. Growth in data generated, collected, and analyzed is a clear global trend, and a significant percentage of this data can be attributed to individuals' personal data. As data proliferates in an organization, the ownership, control, and management of this data becomes more challenging. Like many forms of regulatory compliance, GDPR compliance will be optimally achieved through an enterprise-wide approach to data governance.

Capabilities required: Policy interpretation is a capability to capture both business and technology understanding of policies, responsibilities, processes, data terms, and logical and physical models. Crucially, it is also the location where understanding of the technical environment is linked to the understanding of the business environment. This linkage provides organizations with a holistic view of information about their in-scope data domains and forms an integral part of an approach to managing their data assets.

Technology use case: Policy interpretation could be characterized as being an enterprise data governance use case. These are core capabilities to provide a top-down and bottom-up view of the organizational management of data, with links between the business and IT view of information. Typical requirements that would apply to this use case include:

- **Policy definition:** Business and IT definitions, documentation across all operational levels of the business, logical and physical data and process models
- **Responsibilities:** Who owns the data, who uses the data, and what functions have responsibility for quality and security?
- **Definition of terms and process:** Business processes, key data entities, attributes, systems, quality and controls, standardization, business definitions of consent
- **Change process:** Governed process for definitions, governed process for change process governance
- **Linkage to artifacts:** Logical to physical artifact linkage, technical and business data lineage, data quality incorporation

Technology solutions: Adopt enterprise data governance solutions that enable business and IT functions to work together toward the common goal of data governance. Solutions, such as **Informatica Axon™ Data Governance**, are specifically designed to unite business and IT views of data and create the link between logical and physical data assets.

Benefit: Quick and easy contribution from all subject matter experts to define the processes, policies, and data entities the organization has to rapidly build a holistic data governance capability for in-scope data.

3.3 Entry Point Question: How Do We Manage Data Subject Data?

Background: As a direct result of the diverse use of data in complex IT environments, creating a single view of all information for individual data subjects is challenging. This challenge stems from the fact that different systems use very different mechanisms to store and index data. Without a complete view of an individual data subject's data and how this is stored, managed, or processed within an organization, GDPR compliance will be challenging, especially around individual data subjects' rights.

Capabilities required: Personal data management is a capability to identify data subject records within all identified sources, match and link records together for each individual data subject, and create a repository. This repository provides a source of high-quality data on what actual data records are held across the in-scope sources and how each piece of data is linked to an individual data subject. The repository could act as the authoritative source of data when organizations are responding to subject access requests, right of erasure, or right of portability requests. From a business perspective, it can support organizations in managing consent for personal data usage, and then managing this consent: when was it given/withdrawn, through which channel, and which specific terms were agreed to?

Technology use case: Personal data management could be characterized as being a data matching and linking use case. These are core capabilities to identify data subject records across systems and provide a cross-system view of data by matching like records together and creating linkages. Typical capabilities that could apply to this use case include:

- **Access to relevant data:** Profile data subject data, extract relevant data from source systems, apply analytical processes to semi and unstructured content
- **Data quality processing:** Assess data quality levels, apply manual/automatic remediation, process control for manual remediation, metric reporting
- **Single trusted source of data on data subjects including consent, how it is obtained, and how it is managed:** Includes different views and perspectives of the subject depending on their consents
- **Matching and linking:** Define matching rules based upon business process definitions, match records, link like records with scoring, associate consent
- **Data persistence:** Persist linked/unlinked records, analytics, and reports

Technology solutions: Adopt solutions that help discover data subject records from all data domains, using advanced algorithms to match all data related to the same data subject, regardless of where the data is stored. **Reference 360** leverages advanced algorithms to identify data associated with the same data subject, and master data management provides the framework to maintain and manage a common view of data on data subjects.

Benefits: A single view of individuals has shown to have business benefits beyond the GDPR. This is especially true if the individual in question is a customer, who is increasingly expecting tailored personal experiences. From a GDPR viewpoint, the ability to link all data for each individual data subject will ease the burden of enabling the individual's rights. This includes the right to understand data usage, right to be forgotten, and ensuring consent is correctly applied.

3.4 Entry Point Question: How Do I Secure Data and Prevent Unauthorized Access?

Background: Data protection controls are an approach to enacting the GDPR consent requirements and helping protect personal data. There could be a requirement from IT to remove, mask, or pseudonymize production data used for testing purposes or to pseudonymize or anonymize data used for external data transfers. Data access control for personal data at a user level in applications should be reviewed for compliance purposes.

Capabilities required: Detect and protect also provides identity-based access controls and protection to information on data subjects. Data subject information is often exposed to many different individuals across an organization and its ecosystem. Data security controls are used to remove or hide data subject information from those who shouldn't have visibility into it, while making the information available with conditional discretion to those that should.

Technology use case: Enabling consent control could be characterized as being a detect and protect use case. These are core capabilities to protect and secure data access, applying data-centric controls such as masking, encryption, and access controls, and to manage the lifecycle of data including archiving and deleting data and the application. Typical capabilities that could apply to this use case include:

- **Risk analysis input:** Use risk scoring to direct data control methods
- **Orchestration:** The ability to schedule and coordinate data protection tasks based on identified risks and monitoring of unsafe access or conditions
- **Data security controls:** Static or dynamic masking, pseudonymized/anonymized; encryption or tokenization; identity- and role-based access granularity
- **Change/update history:** Application against source systems, record masking or archiving outcomes against consent record, audit trail generation for evidence
- **Archiving:** Archive data out of production systems, log activity to provide evidence, move offline to prevent accidental usage or access

Technology solutions: Adopt solutions that can help manage the lifecycle of data assets and apply access controls over these assets. **Informatica Persistent Data Masking** and **Informatica Dynamic Data Masking** could be used to help to automatically limit the number of people and systems that have unrestricted or unconditional access to personal data. **Informatica Secure@Source** provides data discovery, classification, risk heat mapping, and identity mapping along with data protection remediation by orchestrating updates to security controls.

Benefits: Introduce automation into data masking to reduce risk of breaches of personal data. Visibility of personal data is restricted to those conditionally authorized to view it, and personal data can avoid unnecessary exposure with reliable and appropriate protection.

4. Partners

As with many forms of regulation and data privacy compliance, technology alone will not ensure compliance. Organizations may need the best thought leadership for their GDPR journey and beyond, as well as traditional service and technology solution delivery. Informatica is working together with many highly trained and skilled partners to support you on your wider GDPR and globally mandated initiatives. These partners have been specifically chosen due to their deep understanding of data management and their focus on GDPR compliance.

Find the right partner for you, or contact your local Informatica representative, who can help you find the best partner based upon your needs and requirements.

5. Conclusion

This paper sets out the need for organizations to consider the data implications of the GDPR. This new regulation brings with it both challenges and opportunities for data privacy governance improvements for many organizations. In the time that this regulation has been in force, organizations still need to consider how their interpretation of the GDPR principles impacts current and future data management processes.

Informatica has been the leading data management vendor for over 25 years and has solved complex data management challenges for thousands of organizations around the globe. The GDPR creates many complex data management challenges for many organizations; however, Informatica and its associated partner ecosystem are ideally placed to help these organizations with their GDPR initiatives and the privacy compliance journey ahead.

6. Disclaimer

Compliance with the GDPR will be based on the specific facts of an organization's business operations and use of data. This document provides a set of discussion points that may be useful in the development of an organization's GDPR compliance efforts, and is not intended to be legal advice, guidance, or recommendations. An organization should consult with its own legal counsel about what obligations it may or may not need to meet.

![Informatica logo]

IN09_0719_03337