Informatica

White Paper

Recommendations on How to Tackle the 'D' in GDPR

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

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 impacts organizations and how they manage data pertaining to customers, consumers, partners, staff, and other "data subjects" where a "data subject" is an identified 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 personal information assets to incorporate these data privacy requirements and enhance citizens' privacy rights with protection and transparency. 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 located?	Sensitive Data Discovery and Risk Analysis	Detect and Protect
How is personal data being used?	Policy Interpretation	Enterprise Data Governance
How do we manage subject data for greater transparency?	Personal Data Management	Data Matching and Linking Use Case
How do we secure data and prevent unauthorized access?	Enabling Data Protection 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 expanded 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 protection capabilities to maximize their privacy compliance programs. Informatica® delivers integrated and scalable software solutions to automate, protect, and control data risk exposure, and these solutions can quickly support organizations on their GDPR regional initiatives, and beyond globally, through a consistent approach.

2.Background

2.1 General Background and Potential Implications

The digitization of personal and sensitive data is proceeding at a rapid rate, with today's organizations leveraging the power of user and consumer insights to improve business decisions, engage customers and partners better with targeted products and services, and drive transformational business processes to improve efficiencies. The European Commission has recognized that much of the data being created, collected, processed, and stored is personal data, which can reveal extensive information of EU data subjects that must remain private.

Pre-existing data protection regulations had not reduced concerns regarding the protection and safety assurances of personal data to be handled responsibly and aligned with consumer expectations for trusted use. Diversity and inconsistency of data protection regulations across the EU member states had frustrated 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.[1]

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 with new rights and obligations mandated to data stewards.

As of May 2018, the GDPR has required organizations to more effectively manage and protect personal data on customers, citizens, employees and 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, 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 expects organizations to better understand how they will responsibly utilize their current and future information assets to comply with these new personal privacy principles. This will have an impact on the people, the processes, the technology—and the data management practices and policies that apply to them.

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. Generally, fines can be more lenient if an effort has been made to apply data privacy governance best practices and demonstrate controls in place.

[1] http://ec.europa.eu/justice/data-protection/reform/index_en.htm

However, it's not all doom and gloom. Organizations that focus on data privacy governance can benefit from improved data quality and trust that drives digital transformation with ROI from improving protection and transparency. Safer data democratization, better customer loyalty, and new product and services are just a few examples of revenue-generating benefits.

2.2 Who is the GDPR Relevant To?

GDPR compliance has multiple dimensions and is not limited purely by physical geography; organizations based in North America, Asia, and others must comply if they store and process EU subject data. 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 Steward 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 with the visibility and insights needed. Significant trends, like an increase in data diversity and a move to cloud-based computing, create a dynamic IT landscape, and add to the data management and protection complexity. To demonstrate these challenges, many organizations continue to struggle to answer these types of GDPR-related questions:

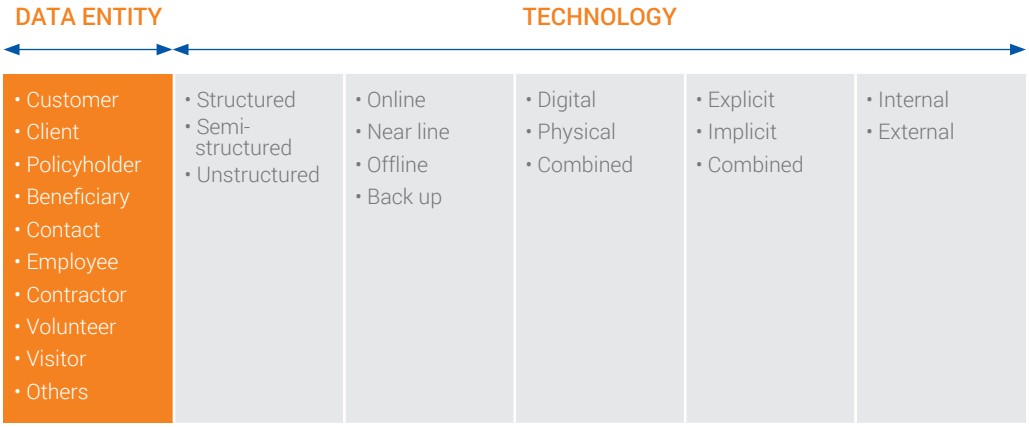
- Where in any organization, and its ecosystem, is all the relevant and in-scope data located to which GDPR principles would apply? Is that data at risk of improper exposure?
- How do organizations keep track of data movement across their operational ecosystems?
- How does an organization classify and manage its relevant personal 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 to data subjects?
- 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 with transparency?
- How can an organization efficiently and effectively respond to subject access requests, right of erasure, and portability requests within the mandated time frames?
- How does the organization control access to protect 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 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

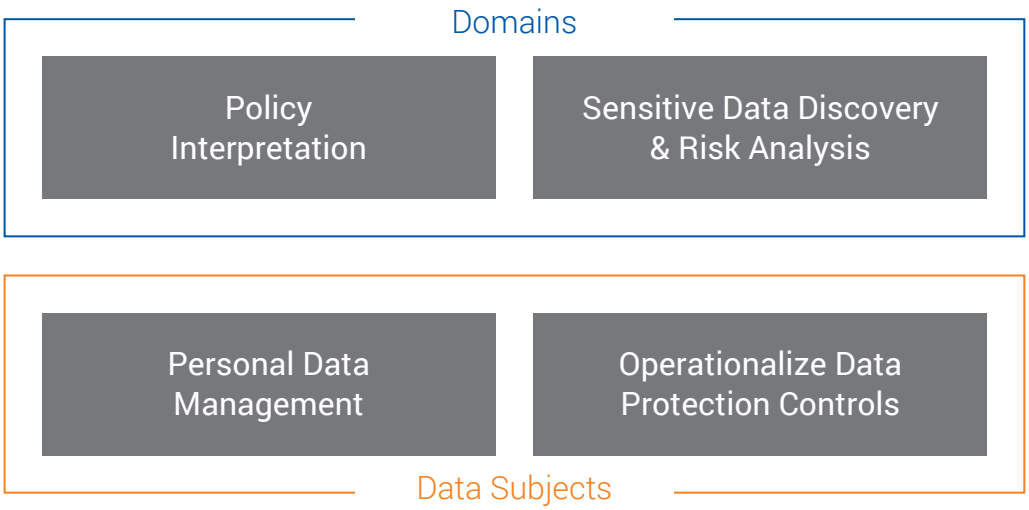
Data subject information generally fits into one or more data entity types, as well as one or more technology types. The diagram below shows some examples of potential data and technology types that may apply to in-scope GDPR data:



These different types require organizations to consider different approaches, methods, and technologies for the capture and management of in-scope GDPR data assets to ensure the risk management solution is operationally efficient and effective for the environment.

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 critical GDPR data governance and privacy challenges to resolve. These entry points are driven by simple questions that 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.



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

Domains relate to the defined information space of data subjects' 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 relate to the identity data at a transactional level. It helps provide insights into personal data management to provide subject-level responses and subject-level insights.

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

Background: Data is usually scattered across disparate systems, applications, and platforms 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 isolated to one department or system. Organizations with more diverse IT systems should not only consider data in core applications but also spreadsheets, local databases, cloud-hosted applications, and data lake repositories.

Capabilities required: Sensitive data discovery and risk analysis is a capability to classify data across a wide range of IT infrastructure and, using this detail along with other sources of insights such as data volume and proliferation, create a risk exposure score for data. The risk score helps organizations understand where most vulnerable data is stored so that potential remediation or similar protection control requirements (monitoring, reporting, etc.) can be prioritized based upon severity impact. Tracking the risk score over time helps demonstrate whether remediation or other control activities have mitigated the data risk position. In support of the lawful purposes, consent may be required so capabilities, such as data lineage (movement), help organizations identify new stores of personal data to aid their understanding of potential unexpected or riskier changes in use.

Technology use case: Sensitive data discovery and risk analysis can 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 located and where it proliferates with analytical insights into data risk. Typical capabilities that 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? What new sources are identified?
- 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 Data Privacy Management is used to help discover the locations of in-scope personal data, classify the data, monitor data proliferation, and assign risk scores for prioritization and remediation. 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 to focus remediation efforts according to risk exposure.

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

Background: Personal data is undergoing a digital transformation that is affecting all sectors. Growth in data generated, collected, and analyzed is a global trend, and this data can be attributed to individuals under privacy mandates such as the GDPR. 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 that offers consistency with a repeatable model.

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 can be best characterized as 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 protection?
- 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 collaborate on the common goal of policy definition. Informatica Axon™ Data Governance is a solution designed to unite business and IT views of data and create the link between logical and physical data assets.

Benefit: Quick and easy collaboration across 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 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 different systems using different mechanisms to store and index data. Without a complete view of an individual subject's data and how this is stored, managed, or processed within an organization, GDPR compliance will be challenging, especially around fulfilling individual data subjects' rights with the transparency required to respond in a timely, cost-effective manner.

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 subject registry enabling a response to inquiries. This registry provides a source of accurate information 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 registry could act as the authoritative source 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 provide transparency into this consent: when and was it given/withdrawn, through which channel, and which specific terms were agreed to by the data subject?

Technology use case: Personal data management can 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 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. Both Informatica Data Privacy Management and MDM Customer 360 leverage advanced algorithms to identify data associated with a data subject, and provide the framework to maintain and manage a common view of data on data subjects.

Benefits: A single view of individuals and their personal data has business benefits beyond GDPR compliance. This is especially true if the individual is a customer, who is increasingly expecting tailored personal experiences. The ability to link all data for each individual data subject will ease the burden of enabling the fulfillment of the individual's GDPR rights while also enabling trust that builds loyalty. This includes the right to understand data usage, right to be forgotten, and ensuring consent is correctly applied for products and services they require.

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

Background: Data protection controls are an approach to enforcing GDPR consent and data access requirements by helping limit personal data exposure. GDPR compliance includes concepts such as data pseudonymization and anonymization for production data and for data used for testing purposes. Data access control for personal data used in applications should be reviewed for compliance scope based on risk exposure prioritization discussed previously.

Capabilities required: The ability to "detect and protect" data can also enable identity-based access controls for information on data subjects. Data subject information is often exposed to many different individuals and applications across an organization and its ecosystem. Data protection controls are used to remove or mask data subject information from those who shouldn't have access to it, while making the information available with conditional discretion when authorized to handle portions of it.

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

- **Risk analysis input:** Use risk scoring to automate data access and protection methods
- **Orchestration:** The ability to schedule and coordinate data protection tasks based on identified risks, and monitoring of unsafe access or conditions
- **Data protection 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 to minimize risk exposure. Informatica Persistent Data Masking and Informatica Dynamic Data Masking can be used to help automatically limit the entitlements of people and systems that otherwise would have unrestricted or unconditional access to personal data. Informatica Data Privacy Management provides data discovery, classification, risk heat mapping, and identity mapping along with automation for data protection remediation by orchestrating updates to security controls.

Benefits: Introduce metadata-driven intelligence and automation into provisioning of data masking to reduce risk of security breaches to personal data. Visibility into personal data is restricted to those conditionally authorized to view it, enabling personal data to avoid unnecessary risk exposure with reliable and appropriate levels of protection.

4. Partners

As with many forms of regulation such as data privacy compliance, technology alone will not ensure compliance. Organizations need best practices for deploying and optimizing resources for their GDPR journey and beyond, as well as traditional service and technology solution delivery to accelerate time-to-value. Informatica is working together with highly-trained and skilled partners to support you on your wider GDPR and globally-mandated initiatives. These partners are specifically chosen due to their deep understanding of data management and privacy governance skills, and a 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 is 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 complex data management challenges for organizations; however, Informatica and its partner ecosystem are ideally placed to help 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

Worldwide Headquarters 2100 Seaport Blvd., Redwood City, CA 94063, USA Phone: 650.385.5000, Toll-free in the US: 1.800.653.3871

IN09_0420_03337