Addressing GDPR Compliance Using Oracle Data Integration and Data Governance Solutions

ORACLE WHITE PAPER | DECEMBER 2017

Disclaimer

The purpose of this document is to help organizations understand how Oracle data governance and data integration solutions can be utilized to help comply with the EU General Data Protection Regulation (GDPR). Some of the solutions described in this document may or may not be relevant based upon an organization's specific environment and needs. Oracle always recommends testing data governance and data integration solutions within your specific environment to ensure that performance, availability, and integrity are maintained.

Further, the information in this document is not intended and may not be used as legal advice about the content, interpretation or application of laws, regulations and regulatory guidelines. Customers and prospective customers should seek their own legal counsel about the applicability of laws and regulations to their processing of personal data, including through the use of any vendor's products or services.

Introduction

Organizations are scrambling to understand the impact of the new European Union General Data Protection Regulation (GDPR). The EU General Data Protection Regulation, which is set to replace the Data Protection Directive 95/46/EC, was designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens' data privacy, and to reshape the way organizations across the region approach data privacy. With far-ranging oversight, its impact includes but is not limited to:

» Potential fines up to 4% of annual revenue turnover
» Reviewing and modifying organizational processes, applications, and systems
» New and more stringent privacy and security requirements to be addressed by organizations that collect and handle personal information (PI) from EU residents

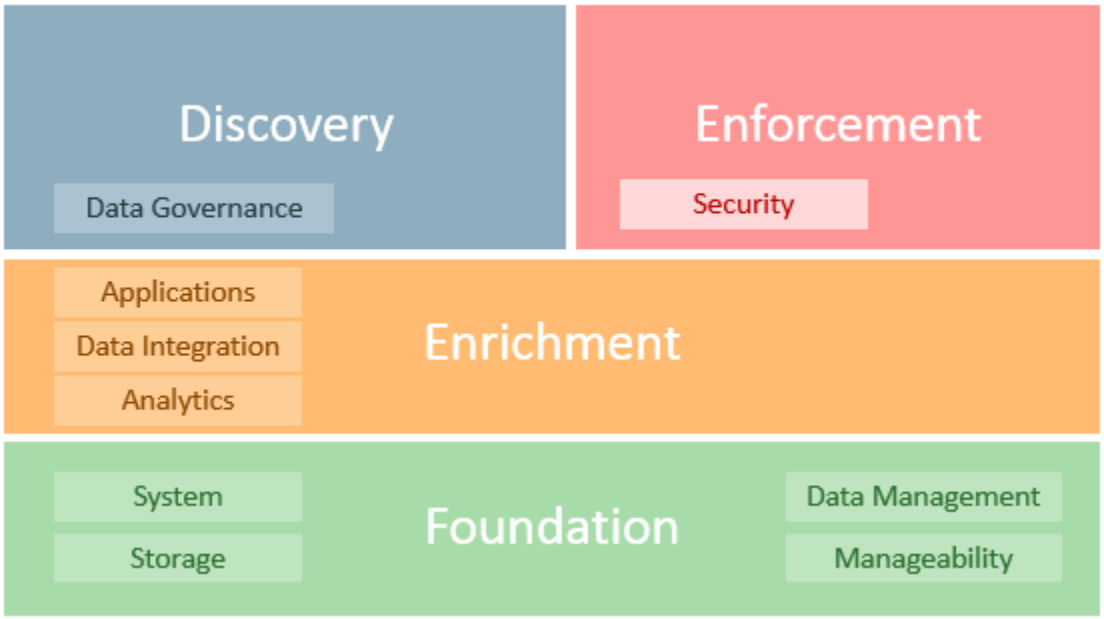
Addressing GDPR compliance requires a coordinated strategy involving different organizational entities including legal, human resources, marketing, security, IT, and other departments. Successful GDPR compliance cannot be achieved without a seamless and secure information strategy across the various entities. In turn, this would require well-coordinated policies and technologies.

Organizations should comply by formulating a clear strategy and action plan to address the GDPR requirements by May 25, 2018.

The Oracle Framework to Help Customers Address their GDPR Needs

Oracle can help customers address their GDPR requirements for data inventory, risk awareness, application modification, and architecture integration. The diagram below provides a high-level overview of Oracle's technology framework enabled by a wide range of products and cloud services.

Figure 1. Oracle solutions framework for GDPR



Discovery. Products and cloud services that can help discover personal data and map data flows. Data governance technologies support auditing and data transparency policies by providing data lineage, asset inventory, and data discovery among other capabilities.

Enrichment. Enrichment includes application modifications that may be helpful to address the rights of the data subject (Art. 15-20). For example, data enrichment is necessary to consolidate customer data to get a single view of the data subjects across the organization.

Foundation. A comprehensive set of mature operational technologies that are a part of Oracle's DNA to enable good IT security with an emphasis on data availability and performance. These solutions can help address the "availability and resilience of processing systems and services; and the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident" (Art. 32). Hybrid cloud solutions ranging from real time data replication solutions to engineered systems to operating systems and processors bolster GDPR's foundational requirements.

Enforcement. Oracle hybrid cloud technologies that enable security policies and controls that are designed to protect people, software, and systems. This encompasses products and services that provide predictive, preventive, detective, and responsive security controls across database security, identity and access management, monitoring, management, and user behavior analytics.

The next sections of this whitepaper go deeper into selected areas of discovery and enrichment as they relate to data integration and data governance. These include the process of finding data, creating a data catalog (classification of data and linking to data processes), building a glossary of critical data, managing risk assessment results metadata, establishing control with policies and rules, monitoring data proliferation, and establishing control with workflow—all are a part of the discovery section of the framework. Eventually, requirements for sharing or deleting data, management of data flows, and creating a single view of customers can fall under the enrichment area.

Data Integration and Governance Solutions

As organizations face the impending EU GDPR compliance deadline – and consider changes in processes, people, and technical controls – it is important to consider how Oracle's products can help accelerate compliance.

The potential hefty penalties promulgated by GDPR have grabbed the attention of company executives and elevated GDPR compliance to high priority within enterprises. Stakeholders from organizations recognize this as a potential 'once-in-a-generation' chance to transform their data management practices. After years of trying to get organizations to recognize the explosive value of data, and the benefits of good data management practices, GDPR can be a compelling business driver.

GDPR provides strong drivers for adoption of disruptive platforms and tools and reimagines business practices that spur innovation. For example, manual ad-hoc and single pass reporting processes were usually enough for most projects. However, GDPR requires higher and more robust reporting and auditing structures to enable organizations to adequately respond to inquiries from Data Protection Authorities ("DPAs") and individuals.

This paper summarizes several key data management activities that can help address GDPR compliance leveraging the Oracle Data Integration and Governance platform.

The Oracle data governance platform includes Oracle Enterprise Metadata Manager (OEMM), Oracle Enterprise Data Quality (OEDQ), Oracle Data Integrator (ODI), and their associated cloud services - Oracle Data Integrator Cloud Service (ODI CS) and Oracle Data Integration Platform Cloud (DIPC). These can be used efficiently to help respond to various GDPR obligations, including by providing transparent controls designed to implement the data governance principles that are at the heart of GDPR.

Discovery

Finding Data

A major challenge for any organization is to create an inventory of all PI across the enterprise. PI can come in many different formats and types (structured or unstructured) and be stored in various locations and held in various forms, such as voice recordings or video. PI is not necessarily stored not in transactional systems only, but can also be hosted in other locations such as a customer engagement system.

» Data held in an application log that captures details about a user session executed via a web application
» Social media feeds when a customer makes a service request
» Web analytical systems that capture customer journeys through the organization's website
» Customers calling the contact center to make an inquiry

GDPR also implies knowledge of the location of all relevant PI (e.g., to provide effective data discovery and to provide all information about an individual upon request, 'subject access right' – or to have all data about them deleted in certain circumstances, the so-called 'right to be forgotten'). This requires that organizations have flexible platforms to dynamically handle, and with a level of automation, potentially large number of these requests. By developing self-service systems that allow appropriate and easy access to online information, organizations can save a great deal on cost.

Any viable solution needs to provide automated discovery of relevant or partial PI.

OEMM harvests metadata from heterogeneous (Oracle and non-Oracle) solutions that span applications, data management, data warehouse, data integration, extract-transform-load, business intelligence, big data, and Hadoop technologies. This allows data identification through the exploration of connected metadata. OEMM is also extremely user friendly with high levels of visualization in metadata analysis featuring text searches, metadata browsers, diagrammatic visualizers, birds-eye views, fast, straight-forward data flows, and lineage that impacts analyzers to trace data provenance and improve data trust.

Oracle's data governance solutions, and their associated cloud services, further link the process of data identification with the OEDQ or DIPC engine.

By using a navigable visualization of data assets, users can explore various IP stores and their associated metadata. They can also see how data flows between these two critical elements.

OEDQ and DIPC provide the ability to browse results, metrics, and reports of data profiling and validation at various granularity (e.g., by system, by table, by domain, by rule). OEDQ and DIPC can also drill down on results in order to view records that are failing data quality checks and perform on-the-spot root cause analysis into the identified data issues.

Classification of Data and Linking to Data Processes

Assuming an organization has mapped all relevant PI and PI data flows, they must then understand what to do with that data. In this phase, metadata about identified data elements relevant for GDPR can be loaded into a data governance platform and classified. This further contributes to an effective discovery phase.

Effective classification is driven from metadata definitions and metadata patterns and rules. A description of classified data sets is achieved with a metadata-driven approach and assignment of additional metadata attributes by answering the following questions:

» Who is accountable and responsible for this data?
» What is the origin of the PI?
» Why was it collected (the purpose)?
» Which data sets and processes does it feed into?
» Who has access to this data set?
» What is the SLA associated with using this data?
» Who are the data subjects involved?

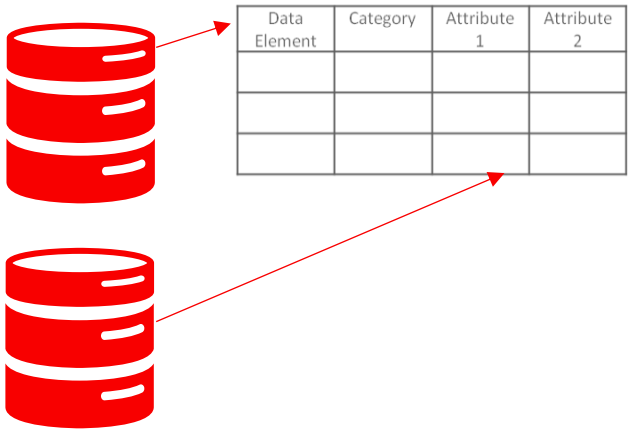


Figure 2. Classification of data and linking to data processes

If data flows are identified and managed and the data elements are identified and governed, they can be linked together and easily traced to the exact data element used in specific business (data) processes.

With OEMM, organizations can document technical metadata, describing the structure of a piece of data, its relationship to other data, its origin, format, and use. This will form a list of data elements with a pool of definitions one may search and report on.

OEMM can organize terms and/or domains by category, making updates and administration easy. Categorization can help with managing stewardship assignments (at the category level), sub-setting by subject matter, or sub-setting by organizational structure.

You can define and document data processes by defining relationships among the semantic and the technical architecture.

Building a Glossary of Critical Data

Examples of critical data elements (CDEs) could be: bank account, IBAN, demographic data (name, gender, DOB, age, nationality), channels (phone number, address, email), government identifiers (passport number, ID number, social security, driver license), digital identifiers (IP address, coordinate), social media (Twitter, FB, LinkedIn), and sensitive personal information (health, sexual orientation, political views, religious affiliations, genetic, ethnicity, etc.).

Each organization should establish a business glossary to determine critical, sensitive, or general personal data. The glossary should align business and IT on the definitions of business terms to add value to the discovery phase. This will help all parts of the business better understand the data context and usage and provide a clear overview of responsibilities including how to document data, but also how to show where it came from and with whom it has been shared.

Business glossaries are also self-contained collections of categories semantically mapped to objects, such as tables and columns in a data model.

OEMM supports business glossaries of terminology, data definitions, domains, business and validation rules with descriptions, inter-relationships, and code sets. Building a business glossary can be as simple as importing in an existing, well-documented data model, or it can be populated directly via the user interface, or in the process of classifying objects in other data store models.

As PI glossaries can be directed to the raw data through vertical lineage, data quality calls can be pushed down through the various technology layers. OEDQ and DIPC can easily be integrated with other components of the environment to track the source of PI attributes and ultimately profile, audit, parse, standardize, case, and transform data.

Establishing Control with Policies and Rules

GDPR's 'accountability principle' recommends organizations to be able to show to DPAs how they comply with data protection principles, including by demonstrating that they have effective policies and procedures in place.

For instance, an individual may invoke the right to have its data deleted, and yet the organization that handles the PI related to that individual might have a requirement to keep transaction data for an additional retention period for auditing purposes.

OEMM can be used to configure and implement or import business rules and link them to the rest of the metadata (i.e. database table/column) using a semantic link. This allows users to see all business rules related to any item and is contributing to an effective discovery phase. A policy can be created through one or more combined rules and through an explanation of the policy by adding new metadata associated with the rules.

Data quality rules can be maintained in OEMM and they are linked to physical entities/attributes in conjunction with OEDQ data quality scores. It is possible to further link the process of policies control with the OEDQ or DIPC engine.

Policies will typically define thresholds for rules, allowing business to define service-level agreements for the data. These SLAs can measure the effectiveness of data quality and measure consumer data trust to generate issues (and notifications) to appropriate users.

Oracle's data governance solutions can deliver a way to link multiple PI assets to policies and rules that define an expected level of data quality based on this 'semantic' understanding of what the data should represent. These solutions can also include built-in issue management with a customizable workflow. Particular events can trigger automatic issue notification by measuring published data quality results against a defined policy. With automatic assignment of issues based on their context, issue management will serve as a way of tracking the work of data stewards formally in a dashboard.

Monitoring Data Proliferation and Establishing Control with Workflows

In the normal course of business, PI is often moved from a known source location to another. This is the case where IT solutions exist to solve specific business challenges involving transferring PI data to new locations for further analysis. Continual monitoring would assist organizations in discovering and capturing these transfers.

In every organization, there is always a great deal of information that gets taken from the core systems and transformed in other tool sets, which makes continual monitoring very important.

OEMM's policy definition, automated discovery, and classification all become important as they bring clarity on whether this newly generated source really contains relevant PI, and if so, what the risk associated with it is.

Oracle's data governance tools provide further visibility on changes through their advanced workflow capabilities.

The workflow is a prepackaged sequence of activities around term proposals, reviews, acceptances publishing, and depreciations. It is a flexible process that can be customized – for example, to require only publishing activity, approval with or without review, approval and review by one or multiple users, and so on.

OEMM's business glossary provides a very flexible workflow and publication process that may be implemented depending upon on an organization's needs. It can be achieved by collaboration features in OEMM.

Oracle's data governance solutions are designed to track changes to business glossaries, policies and rules, and reference data. Reference data and policies can also be managed using Oracle's data governance workflow capabilities to approve changes and to check data quality.

Risk Assessment of Data Elements and Data Processes

After identification and categorization, organizations may want to conduct a risk analysis of their data flows and, based on prioritization, implement and document control measures.

GDPR mandates that organizations implement a risk-based approach to such process. In practice, this means that organizations may want to use risk metrics for both the business processes and the data elements in order to identify, for example, locating higher risk for breaches and adding value to the discovery phase. For higher risk processing operations, GDPR requires a 'data protection impact assessment' to be carried out.

OEMM supports a metadata driven approach to achieve this with attributes assignment.

The GDPR risk assessment results are also generated based upon the level of data governance control understanding, movement of relevant PI, volume of data, data protection availability, data proliferation, and data accessibility. As remediation takes place, the high risk assessments will start to drop, so tracking the risk assessment history highlights how much progress is actually being made over time.

Enrichment

Sharing Data

Individuals have the right to request that an organization provides them with all the information held about them ('subject access right').

If an organization cannot identify all PI that belongs to an individual, that would be an indication that they do not have appropriate control over their PI, which might give pause to regulators.

The right to 'Data portability' allows individuals in some cases to request that their data profile be passed to another organization.

Oracle's data governance solutions can help organizations implement policies and rules to govern these processes within the *enrichment* area of Oracle's solutions framework for GDPR. In particular, OEDQ and DIPC can profile, audit, cleanse, parse, standardize, transform, and de-duplicate all this data.

Deleting Data

GDPR's data deletion or 'right to be forgotten' requirements require organizations to remove all relevant individuals' PI from their systems upon request in some circumstances.

Removing data held in multiple different systems and in multiple different formats is a very time consuming, ineffective, and expensive manual exercise. Data quality issues like duplicates pose a threat to this process.

OEMM supports process of identification of these data and their policy control and OEDQ and DIPC de-duplication.

Creating Single View of Customer

Traditional data discovery techniques will only take an organization so far. The impact of not finding and integrating customers' data for a typical customer centricity program can result in potential lower service levels or smaller up-sell/cross-sell opportunities. The impact of not being GDPR compliant can be much more significant.

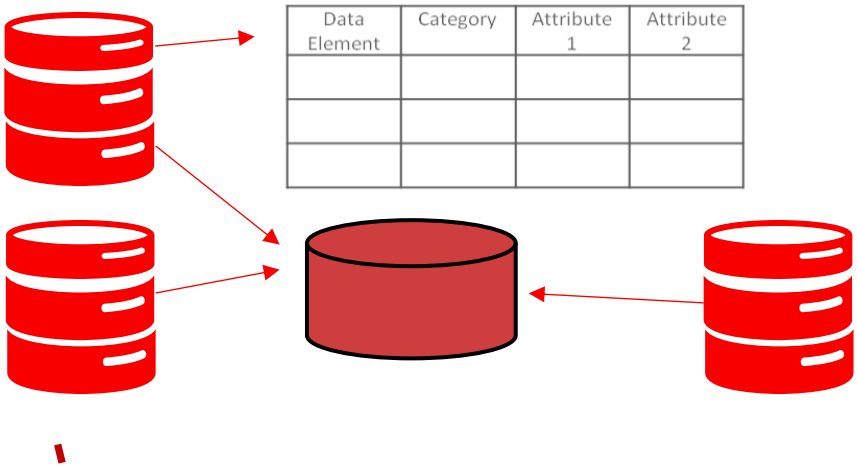


Figure 3: Creating single view of customers

Finding all relevant PI is not always an easy exercise. Creating a single customer view is therefore considered to be a key component of a GDPR compliance framework. Bringing silos of data together in an operational data hub reduces the risks associated with not being able to find the data within the prescribed timeframes. Recording, organizing, and indexing PI, and the systems that contain PI using an ODS (operational data store), can go a long way towards addressing some of the key GDPR requirements efficiently. Besides the disadvantages of creating additional data silos, with its own silo-specific standards, this approach can save time and effort and help achieve commercial long-term benefits from a 360-degree view of all individuals who interact with an organization. There has never been a better time to check levels of data quality on individuals' data.

OEDQ and DIPC are architected to be deployed in many different configurations and have high levels of understanding of data models and processes. OEDQ and DIPC can seamlessly be integrated with the other components of an environment, track the source of customer attributes, profile, audit, parse, and standardize, create a single record from multiple sources using configurable "survivorship" rules, cleanse addresses via open interfaces with address cleansing providers, and identify and eliminate duplicates (using intelligent merge, auto-merge, and unmerge techniques). Eventually, this can provide the "golden record" for selected attributes to all applications and analytical systems, supporting batch processing and maintaining consistent data in real time.

Data integration solutions like ODI and ODI CS, with their easy-to-use user interfaces, combined with rich extensibility frameworks, populate such PI attributes that are derived by the analytical processing in the data warehouse. They also help keep dimension tables in sync with the data warehouse fact tables and support joins across domains. Such information becomes immediately available to the connected applications and business processes. ODI and ODI CS easy-to-use user interface combined with a rich extensibility framework improves productivity and reduces development costs.

Such an Operational Data Store can be complemented with a data lake designed to be leveraged to address an organization's specific GDPR needs. This is a larger data representation with big data based capability which can access a range of types of data including unstructured data to match and link data regardless of their origin and to apply advanced analytical techniques to give answers (e.g. which data to provide for a 'subject access right').

Capabilities can be constructed from existing marketing single customer view ODS and marketing data lakes that many organizations have, or are building with the appropriate infrastructure with changes in the principles: for example, rather than holding marketing preference data, data can be held based on opt-in preferences or based on the source data (lineage) as evidence that the right data records are matched and linked.

Identification and Management of Data Flows

Organizations may be required to demonstrate to individuals regulators for which purpose(s) they are collecting and keeping PI about individuals: from which sources into which targets, what columns are used, which ones have been deleted, and which ones should not be there if they do not fit the purpose. Therefore, organizations should analyze the data processes involving PI and document and maintain these processes within the data governance platform. To that end, a dynamic platform with a level of automation will be particularly useful.

Additionally, identification and management of data flows is also useful to help address the GDPR's 'privacy by design' requirements. 'Privacy by design' embeds privacy into the design specifications of technologies, as opposed to considering privacy only at the point of delivery.

Once the way in which data is being handled is identified by data flows visualizations, organizations will want to verify that it is secured and documented effectively.

OEMM harvests metadata from Data Marts, Data Warehouses, Extract Transform Load, Data Integration, Business Intelligence, and Big Data/Hadoop tools, allowing easy high-level visualization in metadata analysis and fast and straightforward data flow and lineage analyzer. Through semantic model harvesting from existing models when defining relationships among the semantic and the technical architecture, data processes can be defined and/or documented, further contributing to the outcome of the enrichment phase.

Application Services Governance

The more data is being reused without proper data governance, the greater the risk of data-handling mishaps, including data accuracy and data integrity issues. Applications, which necessitate complex data and process integration, contribute to further potential data and process abuse if not approached with proper governance. To help prevent unintended data uses within the organization, Oracle has integrated Data Governance and SOA Governance or Application Services Governance within the Oracle API Platform Cloud Service. Process owners provide the subject matter expertise required to understand the meaning of data within the context of their processes, while data owners will bring the understanding of the processes and metrics using their data sets.

Pre-built compliance of any packaged application or custom-built application should not be mistaken as a one-shot fix for compliance needs. This is because business needs and regulatory requirements may eventually grow out of end-to-end capabilities of even the most comprehensively packaged or custom-built business application. Processes that originally resided within the application may thus spill outside the initial application boundary. Therefore, capabilities of existing applications should be extended by modeling their data and process interactions with other applications or user-channels. This can be addressed by process modeling tool such as Oracle Business Process Management Suite and its associated cloud services such as Oracle Process Cloud Service.

Conclusion

Non-compliance with GDPR can result in hefty fines and increased regulatory actions. But probably even more important: it can also damage an organization's brand, value, and reputation. Organization that collect and handle PI in scope of GDPR will therefore want to ensure they have reviewed their data collection and handling practices.

The path towards GDPR compliance requires a coordinated strategy involving different organizational entities including legal, human resources, marketing, security and IT. Organizations should therefore have a clear strategy and action plan to address the GDPR requirements with an eye towards the May 25th 2018 deadline. Oracle has the experience and technology required to successfully helping customers adopt a strategy designed to achieve GDPR compliance. To learn more about how Oracle can help, please contact your local sales representative and visit https://oracle.com/goto/gdpr.

CONNECT WITH US


blogs.oracle.com/dataintegration


facebook.com/oraclesecurity


twitter.com/oracleDI


oracle.com/dataintegration

Integrated Cloud Applications & Platform Services

CONNECT WITH US


blogs.oracle.com/oracle


facebook.com/oracle


twitter.com/oracle


oracle.com

Helping Address GDPR Compliance Using Oracle Data Integration And Governance Solutions

Integrated Cloud Applications & Platform Services

September 2017 Author: Milomir Vojvodic


Oracle is committed to developing practices and products that help protect the environment