EU GDPR as a Catalyst for Effective Data Governance and Monetizing Data Assets

WHITE PAPER / JULY 19, 2018

ORACLE

DISCLAIMER

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Table of Contents

INTRODUCTION

The European Union (EU) General Data Protection Regulation (GDPR) was adopted on the 27th of April 2016 and comes into force on the 25th of May 2018. Although many of the principles of GDPR have been present in country-specific legislation for some time, there are a number of new requirements which impact any organization operating within the EU.

As organizations implement changes to processes, organization and technology as part of their GDPR compliance, they should consider how a broader Data Governance strategy can leverage their regulatory investment to offer opportunities to drive business value.

This paper reviews some of the Data Governance challenges associated with GDPR and considers how investment in GDPR Data Governance can be used for broader business benefit. It also reviews the part that Oracle's data governance technologies can play in helping organizations address GDPR. The following Oracle products are discussed in this paper:

- Oracle Enterprise Metadata Manager (OEMM) – metadata harvesting and data lineage
- Oracle Enterprise Data Quality (EDQ) – for operational data policies and data cleansing
- Oracle Data Integration Platform Cloud – Governance Edition (DIPC-GE) – for data movement, cloud-based data cleansing and subscription-based data governance

ABOUT GDPR

GDPR governs the processing of personal information (PI) – any data that could potentially identify a specific individual for example data about customers, employees and contractors – and applies to any organization operating in an EU member state.

GDPR harmonizes the regulatory data processing requirements across the European Union, and introduces new elements, especially in the realm of data privacy. Much greater emphasis is placed on the documentation that data controllers must maintain to demonstrate their compliance.

The GDPR requirements provide strong drivers for adoption of data management and governance tools. With the potentially high level of recurring requests from data protection authorities ("DPAs") and from individuals, Data Governance systems and processes must be robust, scalable and cost-effective to operate.

Organizations must be able to show the purpose for which they collected PI about individuals and prove that the individual has given their consent. Individuals can request organizations to show them all data that they have about them ('subject access right' - Article 15) and they can also request to have all data about them to be deleted ('right to be forgotten' - Article 17) or rectified ('right of rectification' - Article 5).

GDPR also allows individuals to request their data profile or the data held on them by a data processor to be passed on to another processor ('data portability right' - Article 18); demands privacy to be embedded into the design specifications of technologies not just at the point of delivery ('privacy by design' - Article 25); requires organizations to be able to demonstrate to DPAs compliance with the data protection of personal data ('accountability principle' - Article 24); calls for assessments where there might be higher risks of security breaches ('data protection impact assessments' - Article 35); and requires notification of individuals and DPAs about data breaches within 72 hours) ('notification of personal data breach – Article 33).

Special processes must be put in place for any PI held about children. Ages must be verified and parental or guardian consents must be obtained for any data processing activity.

Failure to comply with GDPR could trigger substantial financial penalties (up to 20M EUR or up to 4% of the annual worldwide turnover per non-compliant enterprise, whichever is greater) and dramatically affect the reputation of the organization.

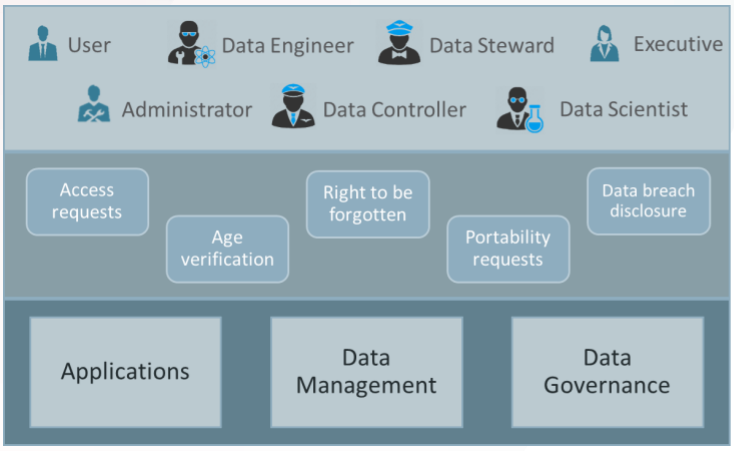
The size of possible penalties has received the attention of company executives and there is general acceptance that this is an enterprise-wide issue that must be dealt with strategically. Stakeholders from many industries recognize this as a potential 'once-in-a-generation' chance to transform their data management practices. The introduction of GDPR provides a compelling business driver to implement what may previously have been seen as merely 'desirable good practice'.

THE IMPORTANCE OF PEOPLE, PROCESS & TECHNOLOGY

Achieving and maintaining compliance with GDPR is a complex and far-reaching exercise that will involve significant changes to the organization, its business processes and many parts of its technology estate.

GDPR imposes fundamental changes to the 3-way relationship between individuals, their data, and the organizations that hold that data. New roles are required, new data ownerships will be assigned and

new processes managed. New data will need to be collected in applications, monitored for currency and correctness, and all personal data traced on its journey through downstream systems.

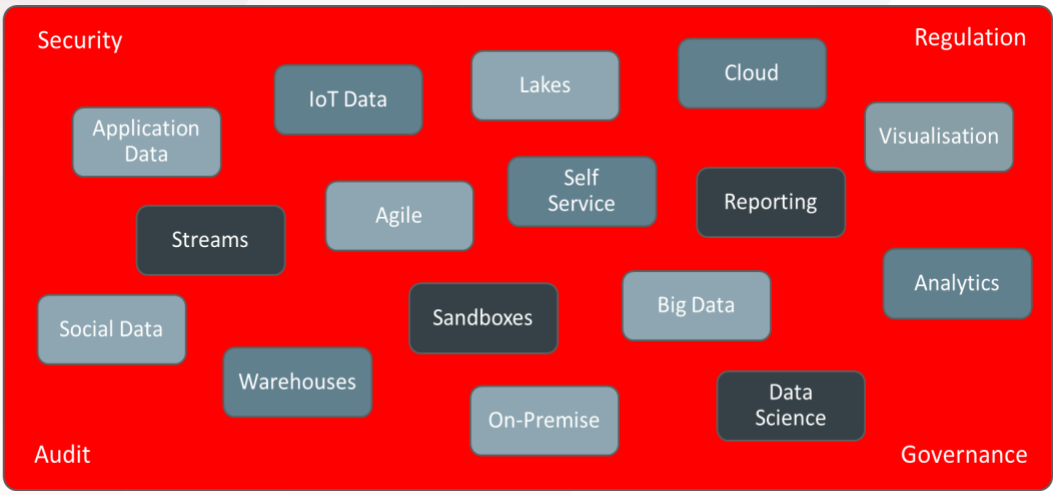


To ease the burden of GDPR compliance, it is essential that the technology solutions are sufficiently flexible to adapt to the new processes and roles within the organization as they evolve and mature.

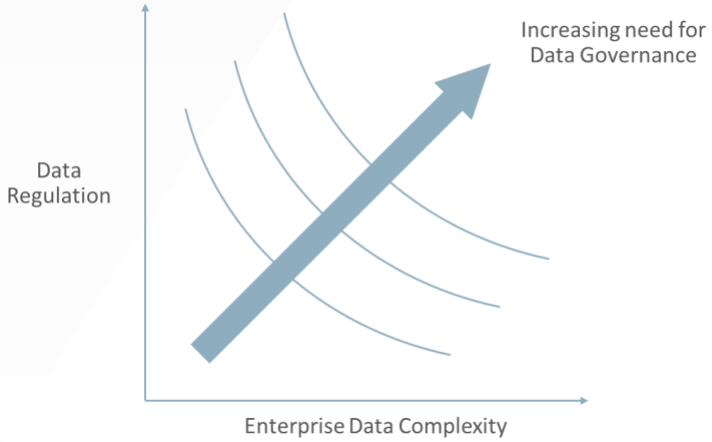
THE ENTERPRISE DATA CONTEXT FOR GDPR

Data is increasingly recognized as a key corporate asset and one which offers the opportunity for competitive advantage if effectively managed and exploited. The last decade has seen a huge increase in the volume of data being captured, accompanied by a dramatic increase in the complexity of the data architectures that are being deployed. New technologies offer the ability to store and analyze data in volumes that would previously have been impossible, while the availability of personal data from third parties is at a level never seen before. Individuals frequently share their personal data with little understanding of the complex terms and conditions they are agreeing to. Under GDPR, consent should be clear and ambiguous with positive opt-in, which will doubtless require the refresh of many existing agreements.

Increasing pressure from the business to innovate and exploit the organization's data has led to self-service initiatives that make it easier for business users to access and analyze data, but potentially at the expense of data security and audit checks. Big data projects have often prioritized flexibility and speed over controls and governance, creating an element of conflict and tension between business agility and regulatory compliance.



Enterprise data architecture is no longer as simple as a number of operational applications with nightly extracts to a data warehouse for reporting. As the complexity of the data estate increases, so does the need for effective Data Governance.



While GDPR is the latest legislative response to an increasingly data-dependent world, it is unlikely to be the last. Effective Data Governance provides the organization with a firm foundation from which it can quickly respond to future data regulation.

THE DATA GOVERNANCE CHALLENGES OF GDPR

Many of the GDPR requirements are about how data may be used by an organization. However, an implicit requirement is that the organization has complete understanding of what personal information is held within its systems, where it is stored, and who has access to it.

According to the UK Information Commissioner's Office:

> "The GDPR requires you to maintain records of your processing activities. … You won't be able to do this unless you know what personal data you hold, where it came from and who you share it with."
>
> 'Preparing for the General Data Protection Regulation (GDPR) – 12 steps to take now'
>
> Information Commissioner's Office

A second important element of GDPR data governance is to ensure that the data held is accurate, up-to-date and being used in accord with the consents given by the individual. The individual has the right to know what information is being held, and the right for it to be corrected if it is wrong.

Although these high-level requirements are easily stated, implementing them in a complex data environment is far from straightforward. Many enterprises struggle to identify where all their source customer data is held, let alone know where that data has been replicated or transformed to during its lifecycle.

EMERGING CHALLENGES WITH THE RISE OF DATA SCIENCE

Data Science has emerged as the latest must-do activity for enterprises seeking to maximize the value of their data assets. Organizations have huge datasets and the role of the Data Scientist is very much on creating and capturing incremental business value, be this by advanced statistical analysis or implementation of machine learning algorithms.

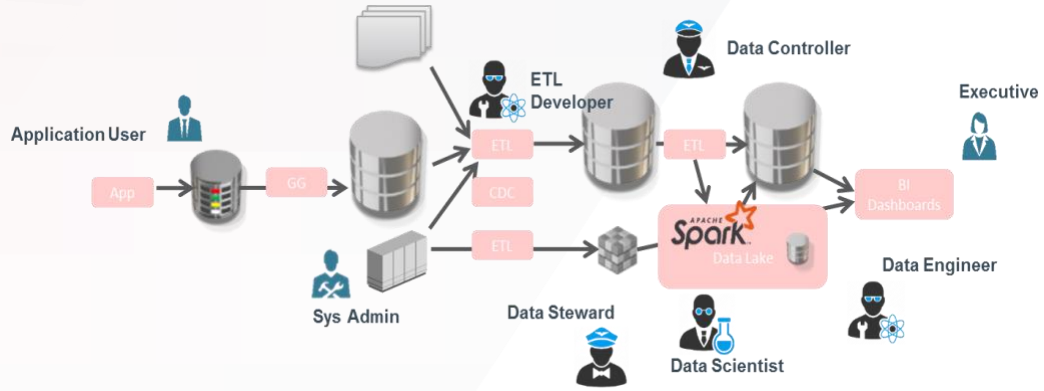
GDPR places strict regulatory obligations on organizations to ensure they have explicit consent from the individual to process their data in a particular way. It is unlikely that existing datasets have consents which would be considered GDPR-compliant and there is clear guidance to refresh those consents as part of GDPR implementation. It is therefore essential that Data Scientists can clearly identify both the datasets that are available to them and what they are allowed to do with each data record. Data with no current consent profile cannot be used and is of no value to the enterprise. It is the responsibility of the organization to be able to demonstrate that they have the necessary consents for the data processing they are undertaking.

GDPR also gives individuals new rights in relation to any decisions that are made based on analysis of their data – the so-called "right to an explanation". This places significant new requirements on the Data Science discipline and how it must be governed to ensure the organization can answer the individual's question "Why?". The provenance of any Data Science work-products and any algorithms used in their generation must be readily available and clearly traceable.

While GDPR could therefore be seen as a negative for the Data Science discipline – flexibility is reduced and costs potentially increased – a strategic investment in holistic data governance can give Data Scientists improved access to higher quality data which can only increase the efficacy of their work.

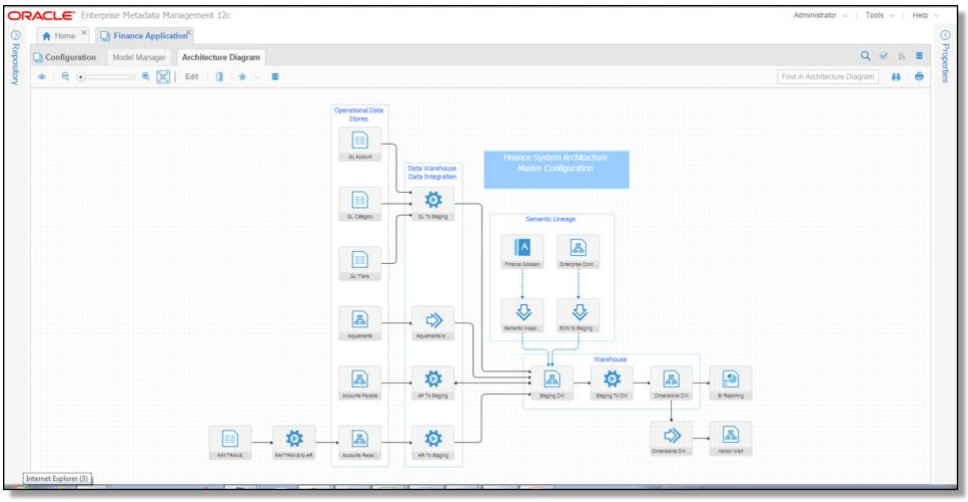
MANAGING YOUR DATA ESTATE WITH ORACLE ENTERPRISE METADATA MANAGER

With the increased complexity of data flows within an organization, keeping track of the propagation of personal information becomes a significant challenge. Data that is captured in an application may end up in a dozen downstream systems or data-stores, via a complex sequence of processes.



For each of these downstream data-stores it is critical that the provenance of the data can be traced back to source. If this cannot be done, it is impossible to meet the 'right to be forgotten' requirement of GDPR or respect any changes to the individual's consent profile.

Oracle Enterprise Metadata Management (OEMM) can harvest and catalog metadata from virtually any metadata provider, including relational databases, Hadoop, ETL, BI, data modeling, and many more. The result is a clear visualization of the lineage of data from sources, through transformation processes, to targets.



Regardless of the complexity of your data estate, OEMM allows you to understand and trace the lineage of data as it flows through the organization's systems. Understanding where personally identifying information flows after its initial capture in an application is critical in the context of GDPR.

Featuring over 150 certified bridges to harvest metadata from enterprise systems into a common model and the ability to map this metadata to centrally defined business terms and standards, OEMM provides the most open and comprehensive platform for the governance of data structures and data flows in an organization. Offering different views of data lineage for different users, OEMM optimizes business users' understanding of analytics reports, as well as technical users' understanding of the impact of data structure and data flow changes, to provide an adaptive and efficient approach to governing data assets.

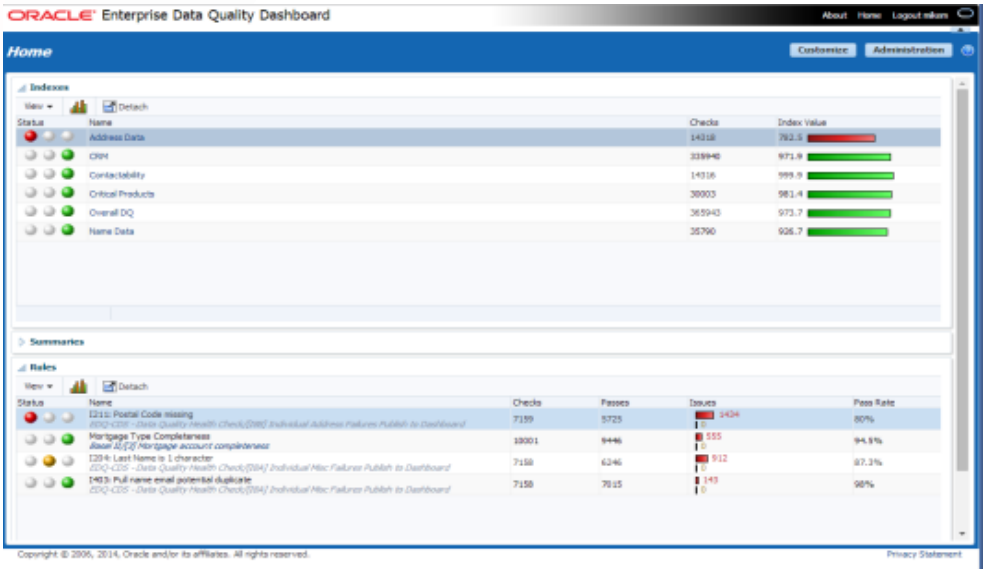
MONITORING POLICY COMPLIANCE WITH ORACLE ENTERPRISE DATA QUALITY

GDPR requires a number of new rules to be implemented around permissions and authority. For example:

- Is all Personal Information correctly age-verified?
- Do we have GDPR-compliant consents to store the information we hold?
- Are those consents up-to-date?

Such rules must be defined based on the data stored, then validated on an ongoing basis to assure the organization continues to comply with policy as data changes.

Oracle Enterprise Data Quality (EDQ) provides a rich environment for the definition and monitoring of business rules associated with data. Data can be profiled and inspected to verify the content is as expected; remediation plans devised if required; rules defined for on-going monitoring and results published to dashboards for highlight any issues.

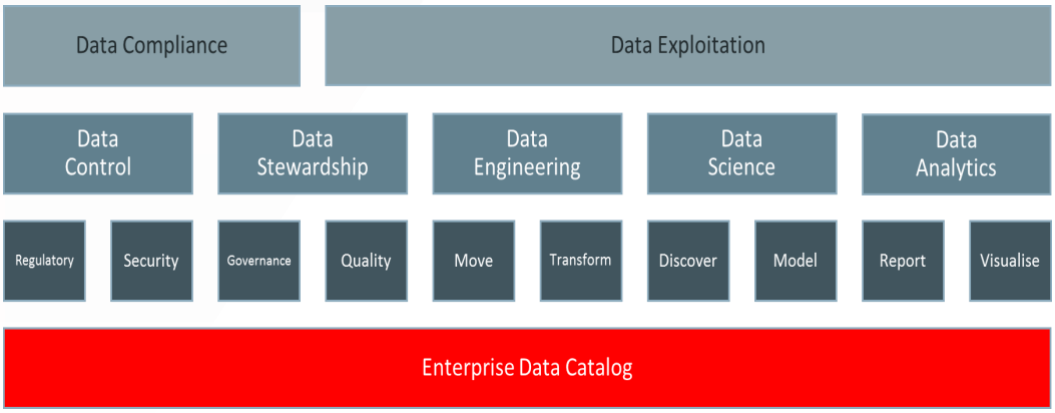


EDQ also provides an integrated case-management capability that allows users to manage any remediation activities that may be required for non-compliant data.

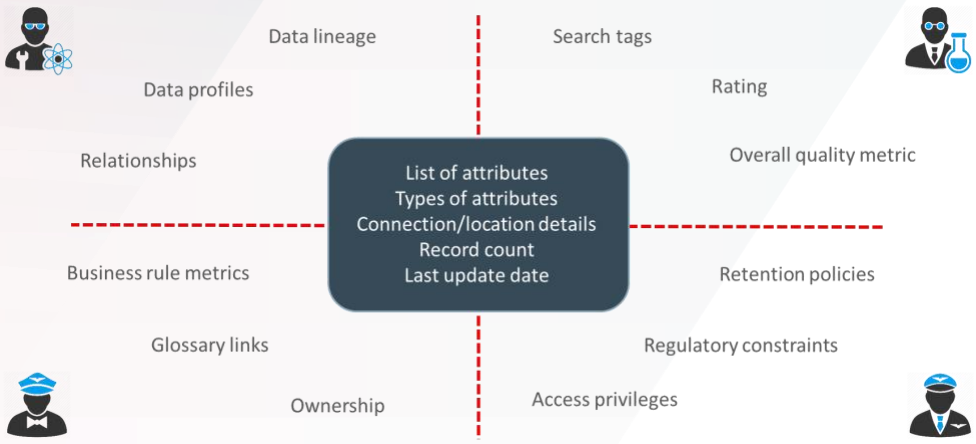
Available either on premise, or in the public cloud as a key component of Oracle Data Integration Platform Cloud (Governance Edition), EDQ offers a fully integrated, collaborative environment to facilitate the discovery, measurement and resolution of all types of data issue, ranging from simple issues such as missing required data values, to more difficult problems, such as the need to reconcile many different records in different systems referring to the same individual. Although it is designed to work with any data in any language, it includes a rich library of out of the box rules and services for working with personal identity data which can accelerate the implementation of data quality rules for critical GDPR data elements.

ACHIEVING ALIGNMENT WITH AN ENTERPRISE DATA CATALOG

Creating and maintaining a catalog of all personal information held by the organization is a significant investment, regardless of the approach taken. However, a catalog of all data assets can have significant value outside of the domain of regulatory compliance. As data exploitation becomes an increasingly important means of competitive advantage and differentiation, the assets used by Data Engineers, Data Scientists and Data Analysts must be traceable, transparent and trusted. A holistic Enterprise Data Catalog provides a foundation for the entire data value chain within an organization.



Although different roles within the organization will have very different uses of the catalog, it is essential that they see the same data assets, perhaps with different information presented base on the user. For example, a Data Controller will be interested in retention policies, access privileges and regulatory constraints whereas a Data Engineer will want to understand attribute-level lineage and data relationships.



If separate catalogs are deployed to serve different roles, compliance becomes even more complex as they will inevitably drift over time. How would the compliance catalog know that a new data lake aggregation of customer data has been created?

It is important that the Enterprise Data Catalog is not seen as simply a documentation exercise. Enterprise data architectures are constantly evolving with new systems being introduced; upgrades taking place; new dataflows being developed; new datasets being added. To be successful, the catalog needs to accommodate the full lifecycle of systems and data from introduction to retirement.

LEVERAGING GDPR INVESTMENT TO DELIVER BUSINESS OPPORTUNITY

As we have seen in this discussion of governance thus far, there is a tremendous opportunity to unlock top-line business opportunities as part of a comprehensive data governance initiative. In other words, the business need not consider GDPR-related data governance a sunk-cost initiative, but rather an opportunity to better monetize data assets across the full breadth of the enterprise. For example, consider the following business initiatives and how they can simultaneously deliver on GDPR requirements as well as prompting the digital transformation of business line functions:

- Data Awareness and Finding Data – Traditional enterprise search tools simply index data for keyword searches, but modern data catalogs, metadata management and data quality tools provide the foundation to find enterprise data based on the underlying semantics, or meaning of the data itself – not just the keywords. From a GDPR standpoint, this can bring a verifiable and auditable record of which customer data is preserved or deleted
- Holistic View of Customer – Classical Master Data Management (MDM) projects have broadly been seen to under-deliver on the initial promise of the technology. Newer 'data lake' approaches have re-energized enterprises to use customer data in innovative ways, such as with Machine Learning (ML) and data science. GDPR investments in the data catalog and metadata management provide a new foundation for understanding a canonical view of customer data attributes that can drive both regulatory as well as sales and marketing initiatives.
- Classification and Linking of Data Flows – One of the key challenges of GDPR is clearly understanding the flow of data through the organization's complex series of systems and processes. Where is data stored? Where did it come from? Where is it distributed to? Understanding these flows for GDPR will also deliver significant benefit to any transformation program by reducing uncertainty and risk, thus reducing costly project overruns.

- Building a Glossary of Critical Data Elements – GDPR imposes greater responsibility on organizations for the accuracy of personal data and associated consents. Building a glossary of data elements gives cross-enterprise clarity of how data should be stored, which can them be monitored for compliance. The increased certainty and confidence in data that results from such investments, improves the organizations analytical agility given an all-important time-advantage to business decisions.
- Establish Operational Controls with Policy-driven Data Quality – Ensuring the accuracy and validity of data for on-going GDPR compliance delivers significant benefits across the organization. Better data allows better decisions, better customer interactions and improved customer satisfaction.

CONCLUSION

GDPR presents significant challenges to any organization in terms of people, process and technology. Many organizations will take a pragmatic, tactical approach to achieving initial compliance, recognizing that the implementation details and guidelines are likely to change based on practical experience. Once the requirements and interpretations are more clearly understood, a strategic approach will provide a more effective, lower cost solution in the long term.

The need to invest in Data Governance to achieve GDPR compliance is unavoidable but if a strategic approach is taken, it can unlock business value through improved agility and ability to better exploit the organization's data. A unified Enterprise Data Catalog allows a single point of control and visibility into all data assets regardless of where the data is stored or how it is managed.

ORACLE CORPORATION

Worldwide Headquarters
500 Oracle Parkway, Redwood Shores, CA 94065 USA

Worldwide Inquiries
TELE + 1.650.506.7000 + 1.800.ORACLE1
FAX + 1.650.506.7200
oracle.com

CONNECT WITH US

Call +1.800.ORACLE1 or visit oracle.com. Outside North America, find your local office at oracle.com/contact.

blogs.oracle.com/oracle facebook.com/oracle twitter.com/oracle

Integrated Cloud Applications & Platform Services

White Paper Title: EU GDPR as a Catalyst for Effective Data Governance and Monetizing Data Assets

July 2018
Author: Oracle

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