

Information Management and Big Data: A Framework For Sharing Sensitive Data

Oracle and PHEMI provide a robust data protection and compliance solution for sensitive data based on the Rittman Mead big data reference model

According to the Ponemon Institute, more than 60% of breaches come from insiders - due in part to the fact that 62% of end users say they have access to company data they probably shouldn't see.

It's a privacy and security obligation that organizations protect their sensitive data. Yet at the same time, it's a competitive imperative that information is put in the hands of customers and decision-makers. **How do organizations manage the sharing and protection of information?** This Solution Brief describes a robust data protection and compliance solution for sensitive data based on the Oracle-Rittman Mead reference model.

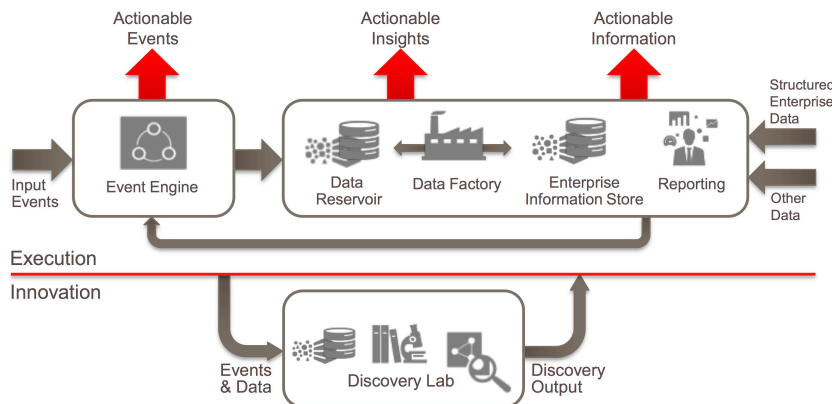
Solution Summary

- Manage and enforce data sharing agreements from multiple data owners across organizational boundaries
- Quickly on-board a wide variety of dirty and heterogeneous data sources
- Maintain provenance of data throughout the workflow from ingest-to-cleansing-to-consumption
- Enforce policy-based access controls with risk-based de-identification and audit controls
- Act as the secure library of truth and trust for distributed, cross-organizational stakeholders
- Support compliance with industry and jurisdictional regulatory requirements such as HIPAA, GDPR, PIPEDA, FIPPA and more

Enterprise data architects face complex data management problems as organizations strive for more insights from more data to outmaneuver competitors and serve the insatiable demands of customers, partners and internal stakeholders. The competitive imperative to share data is offset by the quality, security and compliance obligations that organizations face when working with sensitive data.

Oracle and Rittman Mead collaborated to create the Information Management Reference Architecture, introducing the concept of a Data Lab and Data Reservoir. This framework has become a de facto reference model for architecting the big data workflow.

Oracle and PHEMI provide a robust data protection and compliance solution for sensitive data based on the Rittman Mead model. The combined PHEMI-Oracle solution helps organizations reconcile the tension between sharing and protecting sensitive data.



The Solution Architecture

Organizations need an agile information management framework that is able to share data responsibly for faster time-to-insight and lower cost of ownership. Nowhere is this more critical than in healthcare, insurance, financial services, public sector and jurisdictions with strict compliance regulations.

With PHEMI and Oracle, data engineers can reliably ingest, cleanse, organize and protect data while maintaining data lineage and enforcing strict privacy and security throughout the workflow.

Supporting Oracle Public Cloud, Oracle Big Data Cloud Machine, Oracle Big Data Cloud Service, Oracle Exastack and major Hadoop distributions, PHEMI brings user-centric, policy-based access controls to enforce rightful access to complex and sensitive data at speed and scale.

In compliance with HIPAA, GDPR and Safe Harbor rules, PHEMI offers automated data protection capabilities to identify, tag and protect access to sensitive data at scale and speed across heterogeneous and complex data types. Sophisticated de-identification and pseudo-identification rules such as truncation, generalization, masking, perturbation and pseudo-identification can be automatically applied both on ingest and when data is accessed.

Data Sharing and Usage Agreements can be defined and enforced including provenance, retention policies and access controls to enable legal compliance with licensed, internal and public datasets.

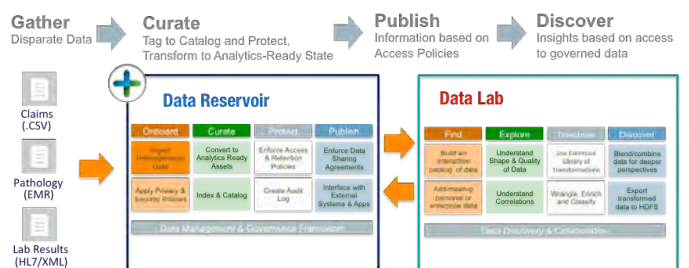
Compliance reports and audit logs track user access requests, dataset proliferation and data usage history so you can detect suspicious or anomalous activities and conduct forensic audits.

The Oracle-PHEMI solution includes Oracle Big Data Cloud, a pay-as-you-grow high-performance Hadoop/Spark environment for advancing businesses' analytical capabilities. Oracle Analytics Cloud provides comprehensive capabilities from self-service data discovery to powerful predictive analytics, enterprise-class reporting, and what-if analysis. Other key Oracle tools include Oracle Data Integrator and Oracle Golden Gate for rapid, low-friction data ingestion.

The customer workflow receives new datasets and updates from data stewards under internal or external data sharing agreements. Data may be XML, HL7, CSV, text files or any other type. The data is immediately transacted in the Data Lab for provenance, security and governance purposes. Data engineers investigate and cleanse the data, pushing results to the Data Reservoir where the information is de-identified and published for consumption by key stakeholders according to policy-based access rules defined by the Privacy Officer.

Transforming the Data Lake into a Data Reservoir

From disparate data to governed, enriched, and protected information for discovery



Ensure Trust: deliver the right information to the right person at the right time.

Data Reservoir

- Provide reliable and scalable storage of raw and curated datasets
- Implement and ensure data security and protection of privacy
- Leverage high performance hardware optimized for Big Data
- Facilitate and simplify the management of Big Data via automation and rich tooling

Data Lab

- Enable agile experimentation to assess the quality and value of the data
- Provide governed access to data for collaboration between business and IT
- Provide rich tooling to wrangle, improve, enrich and blend the data
- Simplify the use of modern data analytics technologies and concepts

For more information, please contact sales@phemi.com or visit www.phemi.com