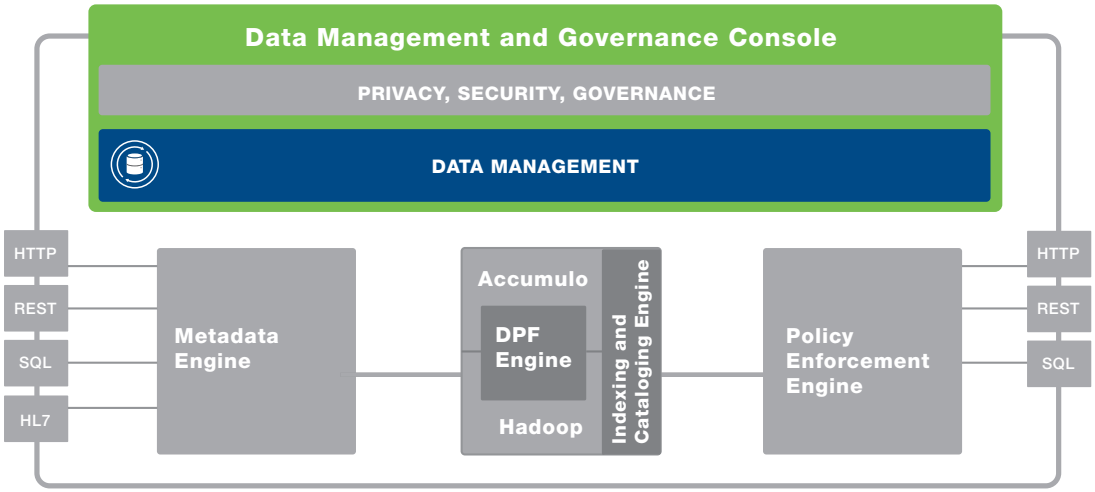
Data Management

Manage your data right at the data element level

Proper management of data is critical as volumes grow and variety increases. The PHEMI Central Data Management and Governance Console provides the primary interface for data management, including data management features for privacy, security, and governance.



Powerful Metadata Framework

The power and sophistication of PHEMI Central's data management capability arises from its powerful metadata framework, which extends from end to end across the system. Metadata is applied on ingestion and enriched by cataloging, indexing, and invoking data processing functions. The result is data description at the element level embedding the rules and policies governing the element, as well as configured properties such as the data source ownership, retention time (time to live), and what visibility the element should have. For example, de-identification, encryption, and masking, along with other privacy restrictions can be enforced per data item, at the data element level.

In a traditional system, a user would have to plan a file-system hierarchy or a database schema. Data would be forced to comply with this rigid hierarchy, and everyone just has to hope that the design scales and that requirements do not change. When PHEMI Central's metadata framework is deployed with its scalable distributed key-value store (based on Accumulo), users no longer need worry about how to structure the system. PHEMI Central structures data automatically, on the fly. Data scales to large volumes at minimal cost while still providing fast access, and changes to requirements do not necessitate changes to design of the data store.

Users and integrated applications benefit from the metadata because they can use simple web-service calls based on the properties of the data, rather than having to navigate complex directories or schemas to find the data they seek.

Governance Rule Enforcement

Most organizations have governance rules and data sharing agreements that stipulate how data may be used and shared. Governance rules are instantiated in PHEMI Central through configuration, using the PHEMI Central Data Management and Governance Console. PHEMI Central manages and enforces data sharing agreements by flagging the sensitivity of individual digital assets, tracking the retention period, recording rules around version control, and specifying de-identification, encryption, and data access permissions. Automating this function at scale across a variety of data sources and types is critical to managing privacy, security, and governance.

Flexible, Attribute-Based Access Policies: Decoupling Data Protection from Data Use

PHEMI Central can automatically de-identify, encrypt, or mask personal information and enforce privacy based on sophisticated user access privileges and fine-grained sharing and consent rules. PHEMI Central stores the fully identified data but strictly controls the rightful use of all digital assets. When the user's access privileges and the recorded data sharing agreements dictate, PHEMI Central can invoke a data processing function to de-identify or anonymize any information information. Anonymization and de-identification may include disallowing access to personally identifiable information, masking certain information, redacting content, or may involve more sophisticated data dependency algorithms to reduce the risk of re-identification. Centralizing anonymization and de-identification helps reduce data sprawl and reduces the risk of data consistency errors.

Data Management continued

Lifecycle Management

PHEMI Central uses organization-specific retention rules to manage digital assets throughout their entire life cycle, from data creation through curation, usage, and end-of-life. Retention rules are captured in the Data Management and Governance Console, and the system calculates a time to time for every data element based on the retention policy and the time of ingestion. It also prevents users from deleting data during a configured retention period and automatically de-identify, delete or otherwise process information when the retention period expires.

Data Immutability

PHEMI Central stores all data in a write-only data system that is never modified. Data may only be deleted at when its predetermined "time to live" expires, as specified by the organization's retention policy. This approach provides assurance of data integrity for audit and compliance requirements.

Version Control

PHEMI Central has robust version control and rollback capabilities to ensure data is never lost, corrupted, or overwritten. The system keeps a history of data revisions and allows administrators to trace changes over time, including the ability to audit who made changes and when, and the ability to roll back changes if necessary. This design provides a complete history for audit and compliance requirements. For example, if your organization develops a DPF and discovers that it has been operating incorrectly, you can revert the data back to the original state without having to revert the DPF (which might impede development).

DPF Framework

Just as the metadata framework manages the tagging of data items throughout the lifetime of data in the system, the PHEMI Central DPF framework manages DPF deployment and execution. The DPF framework is very simple and easy for programmers to learn and use: code libraries are uploaded into the system as simple ZIP files; PHEMI Central manages DPF execution across all datasets and data elements.

The power and sophistication of PHEMI Central's data management capability arises from its powerful metadata framework, which extends from end to end across the entire system.