Transforming Data
into Analytics-Ready Digital Assets

PHEMI

Unlock the power of your data.

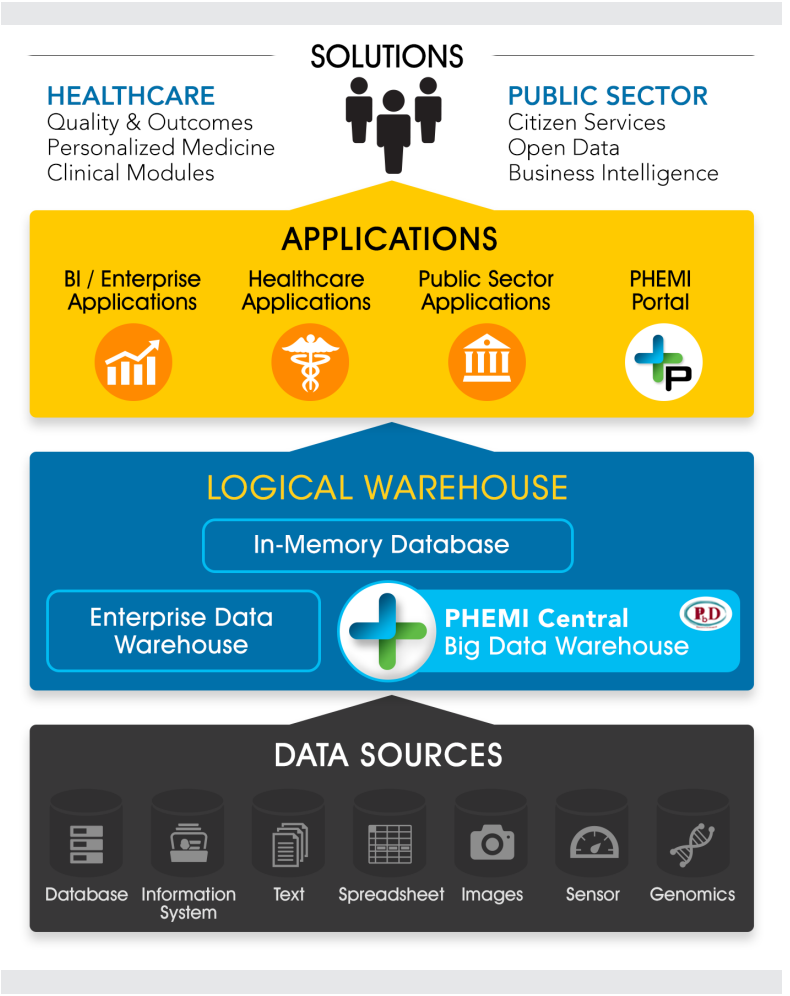
For the first time, organizations that need to protect sensitive information can take advantage of big data technology to access and analyze diverse digital assets.

PHEMI

PHEMI Central transforms structured and unstructured data into analytics-ready digital assets for researchers, analysts and administrators.

PHEMI Central is a next generation data warehouse that combines big data technologies that can store and search petabytes of digital records with fine-grained data management features and an innovative approach to privacy, security and governance.

PHEMI Central gives organizations the agility to aggregate new data sources, conceive new business applications and rapidly build new solutions to exploit their growing datasets and to support strategic objectives.



Add to this section
Available as managed Cloud service or software on premise
Description of where it sits in data ecosystem

PHEMI Central performs a number of functions to ingest and curate data, and to make this data available for secure, privacy-protected consumption.

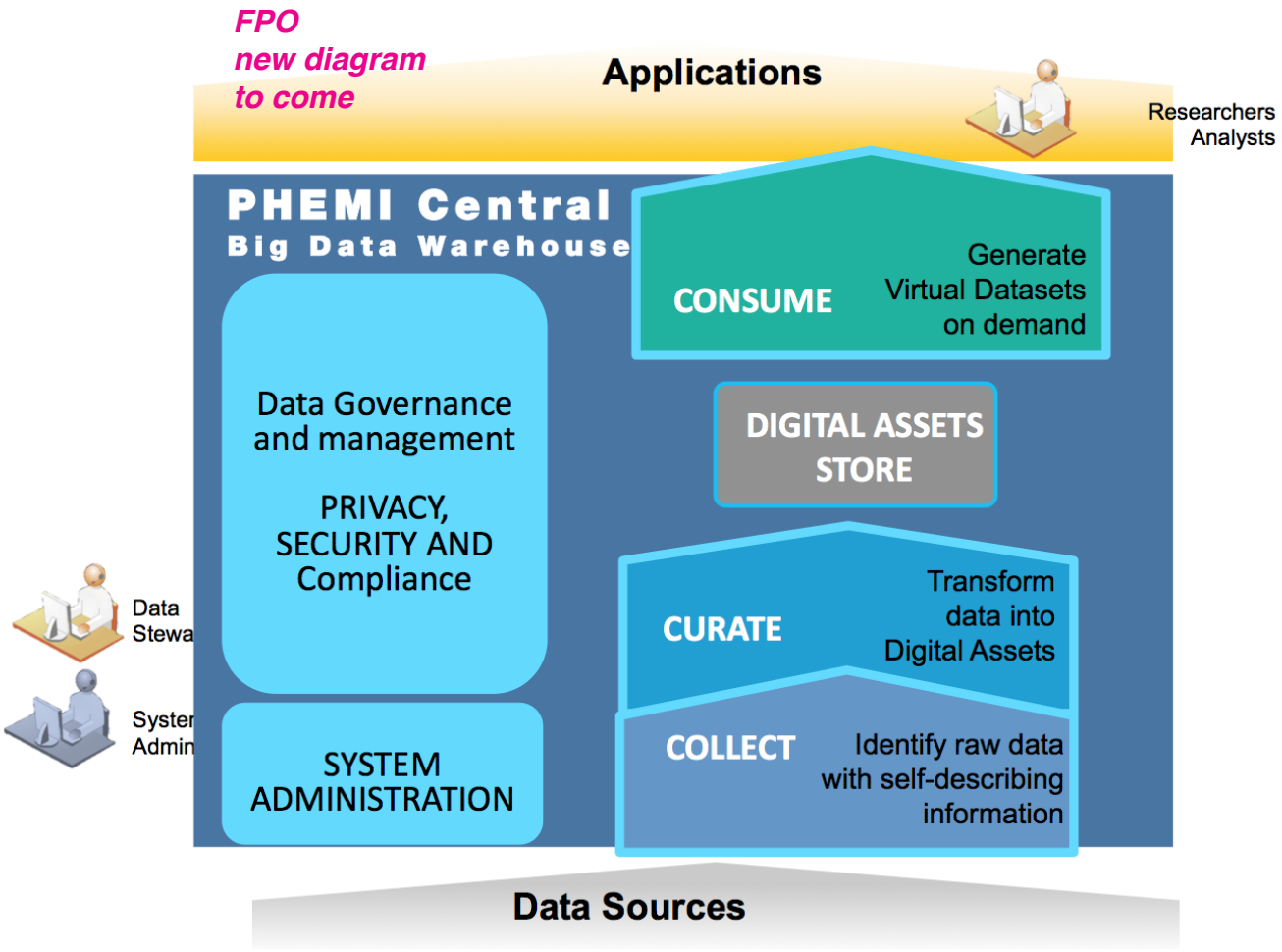
Collect functions extract, load, and tag with metadata virtually any digital record, from structured SQL data and excel files to unstructured data such as images, machine collected data, and documents.

Curate functions cleanse, parse, structure and translate—transform the raw data into cataloged information.

Consume functions create datasets on demand and enforce privacy and security policies to ensure only appropriate access to each individual piece of data.

Privacy, Governance and Security and Data and System management features work with the Collect, Curate and Consume functions to manage crucial aspects such as data encryption, auditing, verification, validation throughout the system, and to control access to the data. […shortened]

Digital Asset Store uses big data technologies, including Hadoop, to store the raw data and the transformed digital assets.



Just about any file format you can imagine—PHEMI Central can take it in and convert it into an analytics-ready digital asset, fully protected and tagged for use.

PHEMI

Collecting Data

PHEMI Central can ingest virtually any digital data record

Data Sources can include:

Databases - examples

Structured Non-Relational Data - XML, JSON, HL7, PHEMI Clinical

Semi-structured – ECGs – add more

Unstructured Data - Images, Genomics, Consult Letters, Reports– add more

User files & datasets - Local data files, spreadsheets

Forms Server -

Metadata—Each data record is imported into PHEMI Central and tagged with self-describing information following rules defined for that Data Source. The Digital Asset illustration shows examples of metadata.

This unique metadata allows PHEMI Central to rapidly find the data, At speed and at scale with psg

Data import options include:

• Existing system publishes to PHEMI's Data listener via http, ftp and our API

• Fetching by PHEMI Central using ODBC/ JDBC

• Manual upload

Data Sharing Agreements- The system tags data according to data sharing agreements, and other governeance rules, such as encryption, access privileges, locality of data, owner, custodian

Consent Management – The system also tags data according to consdent management rules, that define how data is allowed to be used, by whom and for what purpose.

FPO
updated graphic to come

Callout or caption about the benefits of the tagging system that makes the digital asset… advantages, how it works, etc. Are those all the options listed above? can new ones easily be added? …

Curating Data

After collecting the data and storing it in the Digital Asset Store, it goes through one or more Transform processes to cleanse, structure, and extract additional information from the data records.

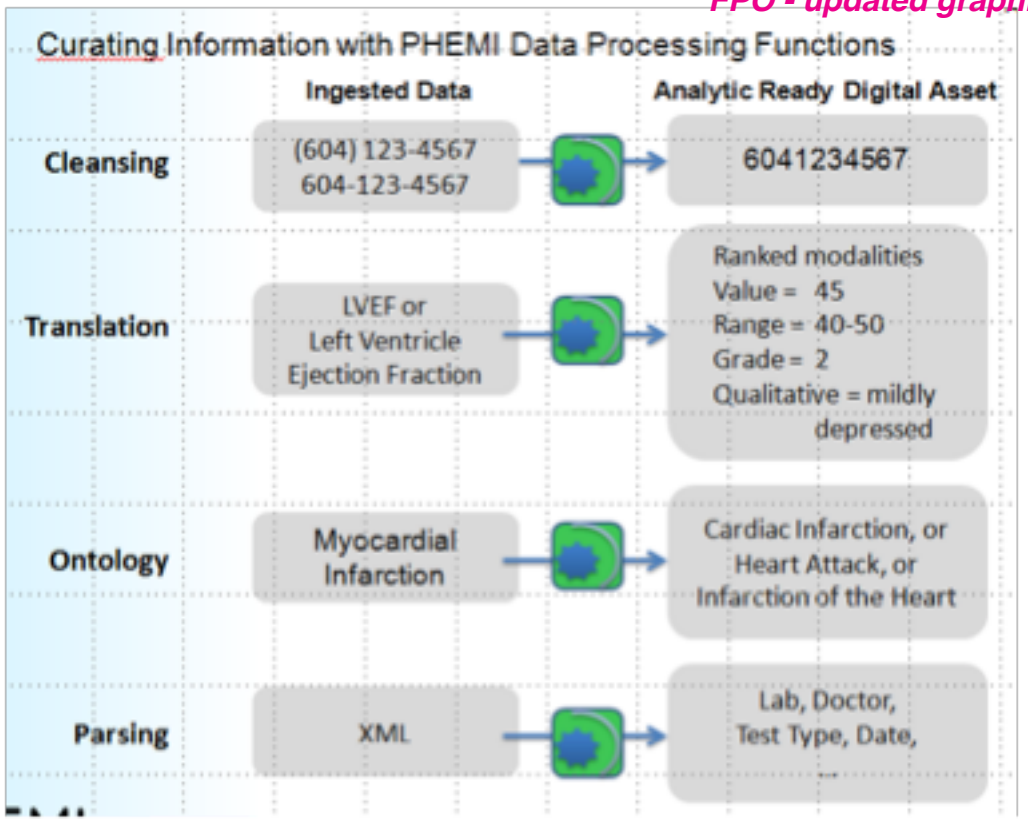
Data Processing Functions are a powerful concept unique to PHEMI Central. Unlike other data warehouse solutions, such as SQL systems which require pre-processed data, or big data platforms, which don't process the data, PHEMI Central can process data during collect, curation and consumption phases.

PHEMI Central includes a library of commonly used DPFs to perform functions such as ?????. New DPFs can be created using common programming tools including Python, Java, C++ coders or contract PHEMI or 3rd parties to develop ti extract additional information or derive data from the raw data.

Data Processing Functions can be used forDPF Examples illustrates some of the available DPFs.

Cataloging and Linking functions make it easier and faster to find and consume the data. Explain cataloging and linking.

FPO - updated graphic to come



Callout or caption about the Data Processing Functions — could provide some more detail about how they work, explain the diagram info …

On-demand queries met with millisecond response times for everyone who needs it, whenever they need it—now that's progress.

PHEMI

Consuming Data

Data Consumption functions support search, reporting, ad-hoc analysis, and pattern discovery. PHEMI Central supports on-demand queries and delivers millisecond response times to user requests with the unique concept of curating analytics-ready digital assets.

PHEMI Central automatically indexes and runs data processing functions in the background, ensuring datasets can be produced on-demand with millisecond response times without needing to wait for MapReduce jobs to complete. Then, on request the Data Consumption function builds datasets and, if access policy rules allow, publishes the datasets to the PHEMI Central Portal or the application. [… more to come here?]

On Demand Datasets- PHEMI Central representations are defined on demand, when the information is needed. With no need to rely on pre-define data representations and schema for queries, PHEMI Central can use the digital assets catalog, linking and metadata to create datasets on demand. PHEMI Central supports new questions, new research and analysis, and new services, without requiring system design or structure changes. .

Anonymization / De-identification— Depending on a user's attributes and the defined policy, the system can call upon a Data Processing Function to de-identify or anonymize information for a given query. This may include disallowing access to personally identifiable information; masking the information, redacting an image or more sophisticated data dependency algorithms to reduce the risk of re-identification. This ensures each data element is properly governed by its associated data sharing agreement(s), provides much greater flexibility for linking data across various sources where appropriate, reduces data sprawl and reduces the risk of data consistency errors.

Data Analysis PHEMI Central supports a range of third party or custom-developed applications and services For example, an enterprise may standardize on 3 different analytics packages, as well as other operational tools, and all can run queries on the PHEMI Central data warehouse. Custom-developed add-ons can equally well leverage the high quality, complete data stored in the Big Data Warehouse

Data Validation— PHEMI Central cleanses and validates data, logging all transformation operations to properly track data governance. It also supports external Extract, Transform, Load (ETL) tools.

Data Verification—checksum capabilities protect digital assets against data corruption or tampering, automatically alerting the data steward if a checksum error occurs.

Add information on using REST for applications, ODBC for analytics and Excel for dataset export (

maybe a GRAPHIC FOR 'CONSUME' ?

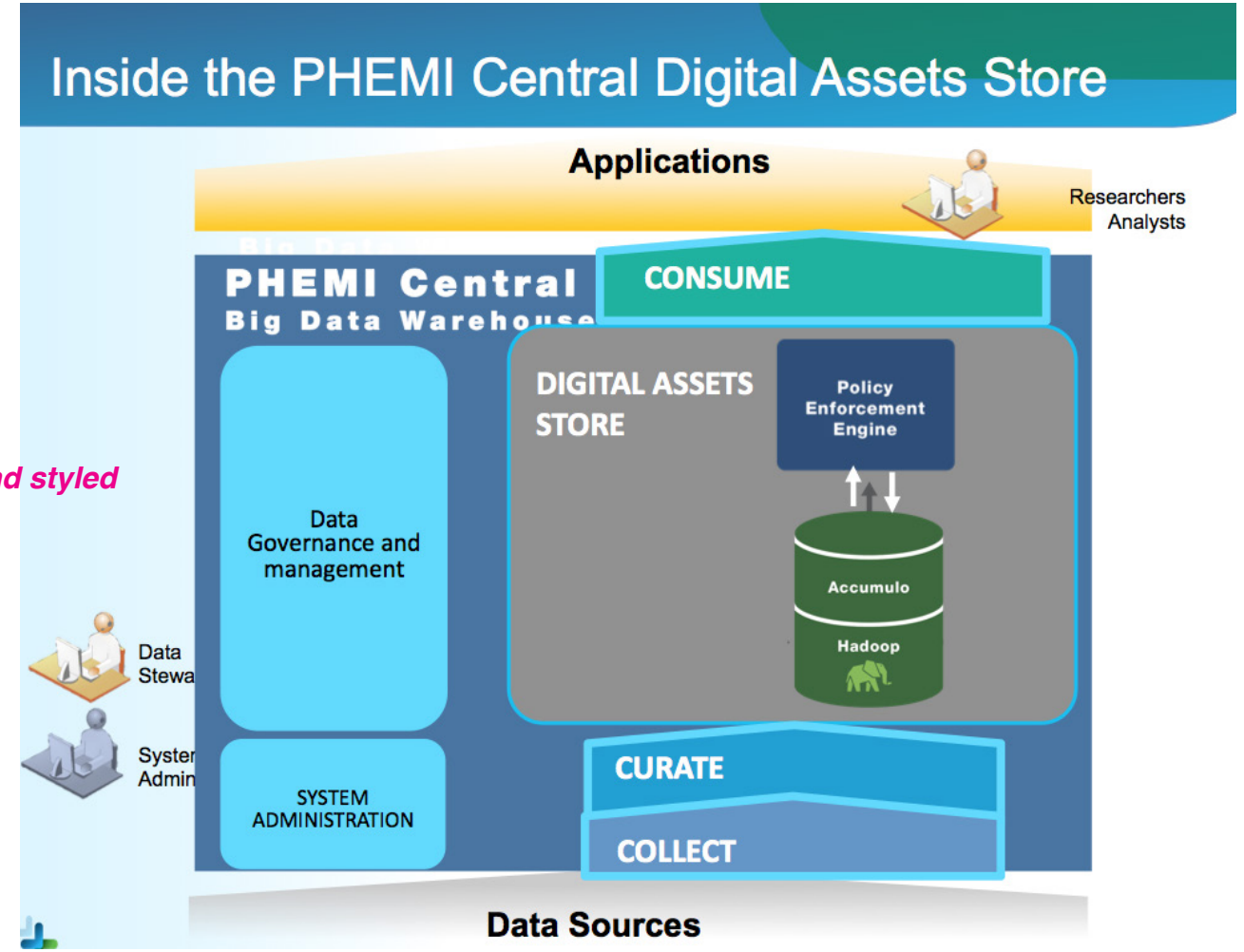
Digital Asset Store

Once tagged, the data element is stored as a digital asset in the Hadoop Data Store, ready for data processing to extract additional information, or derive new data. This schema-less approach to ingesting and storing data allows new data sources to be added as needed, without requiring any rework to

Policy Enforcement—Data Consumption enforces data access and usage policies, defined and managed in the underlying Privacy, Security and Governance and Data and System Management functions

This section needs to talk about Hadoop, Accumlo and the policy enforcement engine. [See more notes in Word doc]

FPO - updated and styled graphic to come



The *Privacy by Design* framework is built into the product to make it easy to define, manage, and enforce data sharing agreements, privacy policies, and patient consent.

PHEMI

Data Governance and Management — Privacy, Security and Compliance

The Privacy by Design framework is built into the product from the ground up to define, manage, and enforce data sharing agreements, privacy policies, and patient consent. PHEMI Central stores fully identified data, strictly controlling the rightful use of all digital assets.

Might need to split out onto two pages… see Word doc for full copy

PHEMI Central stores fully identified data, strictly controlling the rightful use of all digital assets.

PHEMI Central helps organization achieve compliance targets by providing a powerful set of capabilities to manage the privacy, security and governance of data.

Governance rules for privacy and security are enforced at the data repository, rather than application layer, ensuring data custodians control privacy and security, not application developers.

Roles Based Access (RBAC) and Attribute Based Access Control (ABAC) act at the data element level, and support sharing of de-identified, or anonymized data

• Role Based Access Control provides the first level of access control by assigning access privileges to user roles.

• Attribute-based Access Control applies additional policy rules that match user and content attributes and allows access privileges to be managed on a very large scale without arbitrarily complex rules.

Data sharing agreement & consent enforcement —PHEMI Central automatically manages and enforces data sharing agreements that govern data use through Access Policies that control how users access data. Each data sharing agreement can define what data sensitivities exist, how long data is retained (e.g. delete after 5 years), and for what purpose. Changes in data sharing agreements can be done at the click of a button. Automating this function at scale is critical to maintaining privacy, security and governance.

PHEMI Data sharing agreements manage digital assets through their entire life cycle. This includes capabilities such as managing data retention intervals, rules around version control, de-identification, encryption, and data access permissions

Data Steward Access - Using PHEMI Central's Policy Manager Console, a data steward can impose policies on digital assets, which are then enforced through the PHEMI Central policy enforcement engine. Access control can be provided right to the data element level

with th
eleme
data e
syster
of ass
delive
the rig

Privacy by Design framework **establishes**

- ***Proper*** **management and control to open enables positive, secure, use of private data**
- **Internal data firewalls to protect data internally, not just externally**
- **Integrated operational policies to ensure compliance, removing reliance on manual procedures**

FPO - graphic TBD

Security at Rest—PHEMI Central can encrypt data at rest within the data repository. For performance reasons, it is usually unnecessary to encrypt all data. Instead, encryption of only personally identifiable information is advised. PHEMI Central allows for the selection of what data must be encrypted at rest.

Security in Motion— PHEMI Central ensures that communications between data sources, data consumers and the big data warehouse are encrypted using either Secure Sockets Layer (SSL) or Transport Layer Security (TLS).

Audit logs —Audit logging PHEMI Central Users—PHEMI Central maintains complete audit logs of system and user operations including account creation/modification/deletion, policy creation/modification/deletion, and dataset creation/modification/deletion. These log files are

System Administration

EXACT WORDING TBD… Responsible for the creation and configuration of the PHEMI Central Application, such as the Users, Data Protections, Access Controls, Data Management, and Governance capabilities.

Scalability and Performance— Based on technology developed by Google and Yahoo to operate at enormous scale economically, PHEMI Central uses commercially available, commodity hardware components to lower the cost of ownership compared to traditional enterprise data warehouse systems.

Reliability and Availability—PHEMI Central eliminates the cost and performance bottlenecks introduced by expensive Storage Area Network (SAN) or Network Attached Storage (NAS) architectures. Hard drives are distributed across the cluster and can be hot swapped. Faster/larger capacity drives and nodes can easily be absorbed.

PHEMI Central scales linearly with each additional hard drive and node added to the system. Additionally, PHEMI Central automatically load balances to compensate for component failures and to absorb the addition of new hardware capacity.

To maintain high reliability and availability, PHEMI Central automatically replicates all data three times across the cluster, ensuring data survives hard drive and node-level failures.

Data provenance – governance meta data tracks how data has been transferred

Data deduplication

FPO - NEW graphic TBD

www.phemi.com info@phemi.com twitter.com/PHEMIhealth linkedin.com/company/PHEMI

PHEMI