



Multi-modal Data on AWS

Customer Challenges

Healthcare and life sciences organizations want to derive information from multi-modal datasets while limiting the burden of data preparation by staff members. Applying analytics, artificial intelligence, and machine learning (AI/ML) across multi-modal datasets can provide new insights that inform clinical research. While securely storing health data is key to these initiatives, it can be challenging to parse and classify the wide range of data types.

AWS Health for Data:

AWS makes it easier to ingest, parse, and store a variety of health data. This helps organizations discover new insights from disparate datasets to improve clinician experience and patient outcomes..

Top Benefits:



Ingest and catalog diverse data types more easily



Unlock data across the organization



Built-in governance of data



Faster insights from analytics and machine learning

Featured AWS Capabilities:



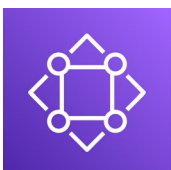
Amazon HealthLake Imaging

Store, access, and analyze medical images at petabyte scale



Amazon Omics

Transform genomic, transcriptomic, and other omics data into insights



Amazon DataZone

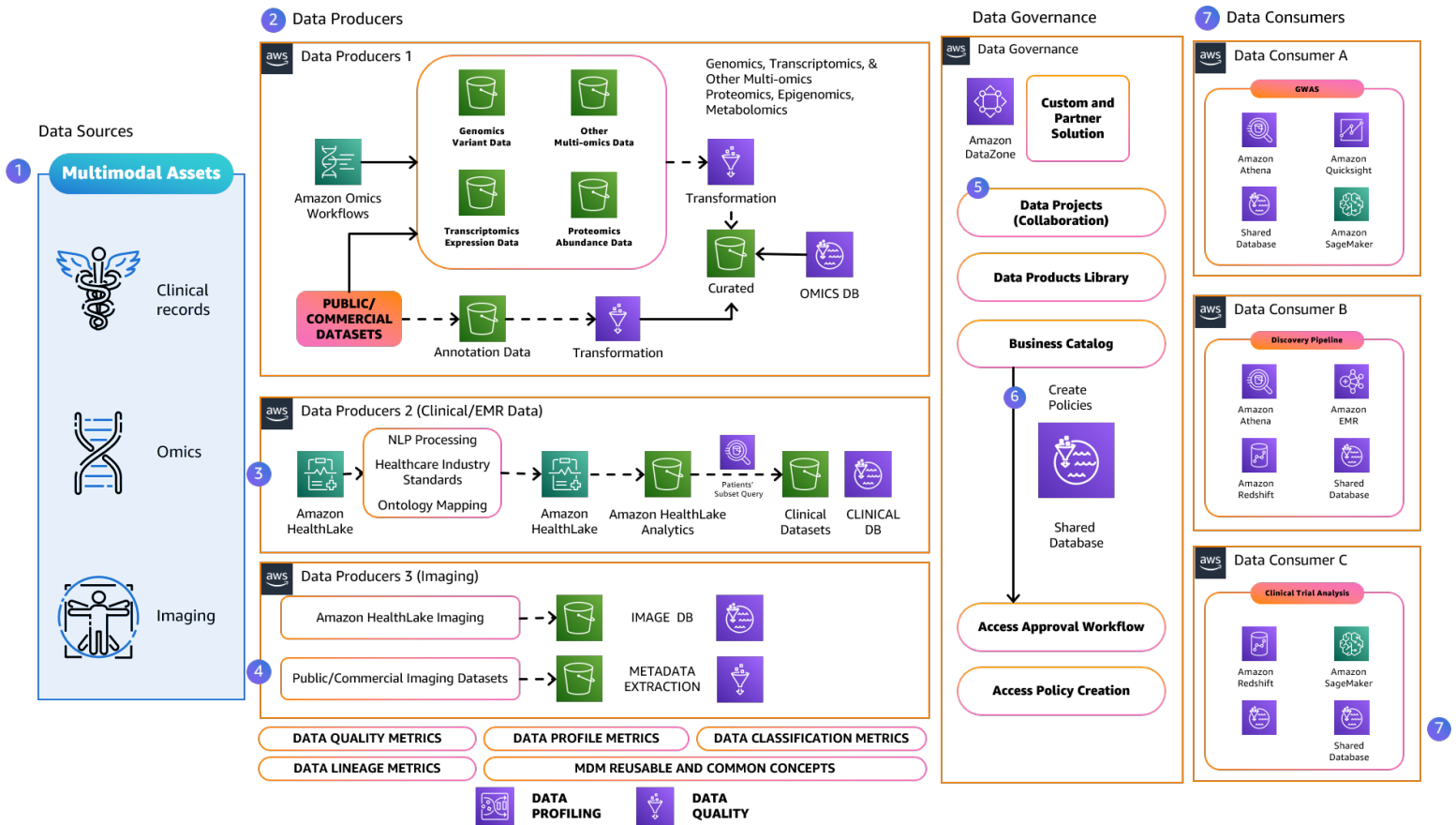
Unlock data across organizational boundaries with built-in governance



Amazon SageMaker

Build, train, and deploy machine learning models with fully managed infrastructure

Multi-Modal Data Architecture for Healthcare & Life Sciences:



- 1. Multi-modal Data Sources:** Data sources includes clinical, omics, and imaging data that needs to be ingested, cataloged and used in building analytics.
- 2. Data Producer 1:** Genomics and transcriptomics data processed using Amazon Omics bioinformatic workflow and it can be stored along with annotation data. Proteomics, epigenomics, and other omics data is processed using a custom data pipeline built in AWS Glue. All omics data schema is built in AWS LakeFormation database
- 3. Data Producer 2:** Clinical data is ingested using Amazon HealthLake and processed using NLP engine and ontology mapping and stored as a FHIR data product. Amazon HealthLake Analytics can then be utilized to build a data extract as a data product. Amazon HealthLake Analytics and a custom clinical data can be shared using AWS LakeFormation database and can be listed in the Amazon DataZone catalog.
- 4. Data Producer 3:** Medical imaging data is ingested and managed using custom solution and/or Amazon HealthLake Imaging solution. A central imaging catalog is built and shared using AWS LakeFormation and listed in the Amazon DataZone catalog.
- 5. Business Catalogs and Metadata:** Publish Data products into the Amazon DataZone central business catalog. Data projects in Amazon DataZone are utilized to publish and subscribe data products and collaborate across team members.
- 6. Self-Serve Workflow:** Amazon DataZone workflow accepts consumer request for data products and route to the data steward of a data product for approval. Once approved , Amazon DataZone and AWS Lakeformation shares the database and build policies for data access.
- 7. Data Consumers:** Consumers can ingest data to AWS via a number of HCLS data transfer best practices and can access data using Amazon Athena and build consumer data products that can be listed in the business catalog as well for discovery and access.