Enterprise Business Data Analytics (EBDA) Platform – FDA's Multifaceted Shared Federated Analytical Ecosystem

Swati Kulkarni, Will Stevenson

Office of Information Management and Technology
Office of Operations, FDA

Why are we Embarking on This Journey?

❑ Why Enterprise Business Data Analytics Platform is needed?

- **Data Sharing** across organization is **not easy** with legacy security models.
- There is **lack of integration** among various systems across FDA leading to **undiscovered insights.**
- **Inability to share** pre-trained **Machine Learning (ML) algorithms and Artificial Intelligence (AI)** services.
- Inability to integrate diverse data sources to produce enterprise wide forecast.
- There is a **need to support data governance, standards, cataloguing and lineage** across FDA.

To address these challenges at the enterprise level, FDA needs to offer common shared services, security and orchestration across Centers

At the same time, each Center will continue to retain the flexibility within the data platforms to be able to meet their own operational needs.

❑ The multifaceted shared, federated analytical solution designed to address the needs of users.

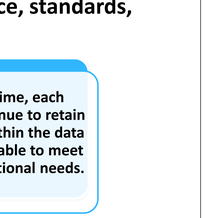
- Platform enables **highly secure** cross-center and external **data sharing**.
- Platform builds foundation for **federated data governance** and **tool agnostic environment minimizing duplicate IT budget spending.**
- Platform leverages **AI and pre-built trained ML** accelerators.

EBDA Platform Design goals

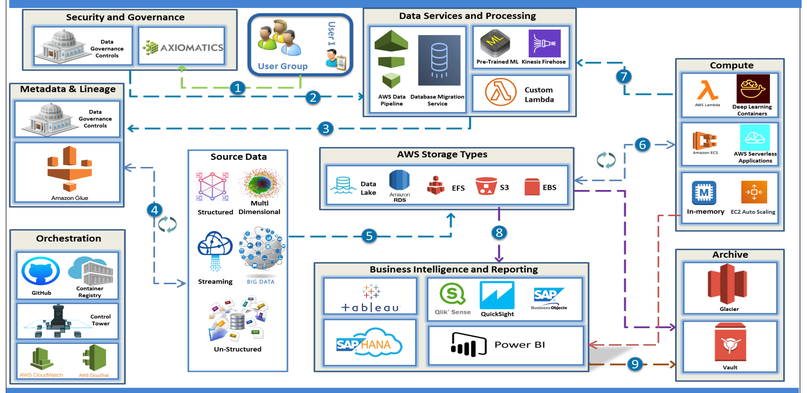
- Provide **secure way of sharing data** without exposing data to internet-gateway through Enterprise Security.
- Allow **cross Center data sharing** while maintaining separation of data where applicable.
- Provide **complete autonomy to Centers** to govern services, access and policies within the centers.
- Address **the goals of the TMAP** in modernizing the agency.
- Introduce and streamline **traceability and auditability** of data.
- Automate **mundane and manual processes** related to data gathering and preparation.
- Create a common enterprise wide **sandbox** to test enterprise wide analytical solutions.

The platform design would be geared to offer key features and services.

- Data-as-a-Service (DaaS)
- Data Catalog Service
- Tool Portfolio Service
- AI/ML Workspace



Detailed Platform Solution Architecture



Results

The platform technology is focused to provide business value and accomplish goals.

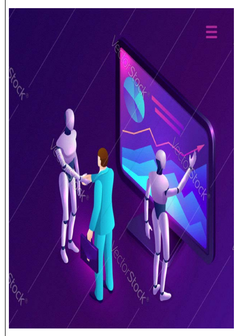
- **Reduced replication of data** across Centers since a data lake can hold multiple center's data and integrate seamlessly with security restrictions.
- **Modern Application Program Interfaces (APIs)** are created once and used multiple times using RestAPI services**.**
- The platform architecture design is based on **containerized microservices**, allowing integration with on-premise, hybrid and centers' cloud environments and components.
- Ability to create **Business Intelligence and Analytics** for structured, unstructured, Semi-Structured and real time streaming data.
- Providing **shared enterprise platform for accessing and managing methods/ algorithms** that **cross Center domains.** Example: Likelihood Ratio Test (LRT) methodology (Huang el al 2011) (https://openfda.shinyapps.io/LRTest)

EBDA platform is designed to –

- Enforce **standardized data Ingestion** for every category of data**.**
- Enable **data search at enterprise level** using Elasticsearch.
- Support a **central metadata catalogue creation** of centralized semantic-based definition.
- Offer **data profiling** capability for Data Quality.
- Identify the **technology centric generic activities that have potential to cross center domains** and perform these activities through **Enterprise Services.**
- Consider **Centers data platforms as another data source** to integrate with this enterprise platform.
- Provide **access across centers to use/reuse pre-trained machine learning accelerators.**
- Enable **central metadata catalogue** creation of centralized semantic-based definition.
- Capture **metadata and reference data at FDA level**.

Solutioning | Enable FDA's Data-as-a-Service and Putting Data to Use

Provide an **onboarding strategy** where Centers and/or offices can integrate their business processes and data onto the Analytical Platform.

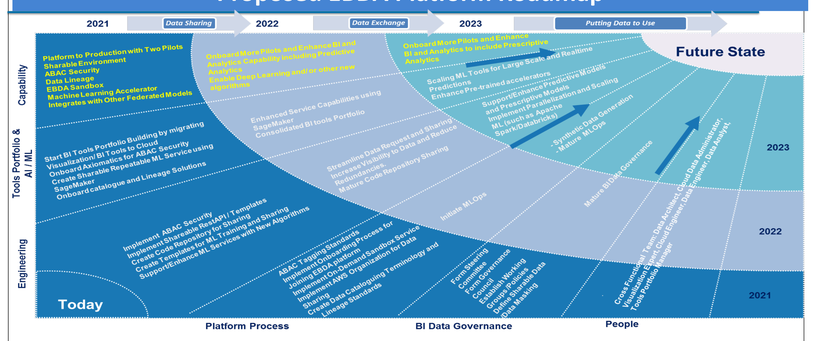


- Provide a questionnaire to **understand centers' goals** and the most suitable areas of the platform for their business processes.
- Develop a **integration plan** for existing operational environments, tools and data source connections.
- Develop a **security strategy** tailored for each new tenant, that is applicable to their new space within the cloud environment.
- Initiate **post integration monitoring** and evaluation to measure success metrics and provide post integration support as needed.

Few examples of the benefits which are realized through EBDA Platform

- Access to **predictive and prescriptive analysis** through AI and ML
- Lower Total Cost of Ownership (TCO)
- Elastic scalability
- Multi-tier data sharing
- Cost savings through **Multi-tenancy** and through centralized BI and Analytical tools governance and management

Proposed EBDA Platform Roadmap



Acknowledgment

The EBDA platform presented was endorsed by OIMT and CDER executive leadership. The platform design was supported by Kartik Murugesan and Rajesh Sripada.

References

Huang, L., J. Zalkikar, and R.C. Tiwari, *A likelihood ratio test based method for signal detection with application to FDA's drug safety data.* Journal of the American Statistical Association, 2011. **106**(496): p. 1230-1241.

Disclaimer