SmartData Fabric® Virtual AI for Healthcare

March 2024

Two primary value-adds to AI by SmartData Fabric®

1. Access to distributed data for AI inferencing:
   a. Assemble high quality, complete and relevant datasets from a wide variety of sources and provide in ready-to-use form for inferencing
   b. Support creation, storage, indexing, search and retrieval of embeddings for enterprise/proprietary data to enable Retrieval Augmented Generation (RAG)
   c. Optimize prompting costs through reuse of saved embeddings
   d. Retrieve referenced source records in generated responses/summaries for human verification and Reinforcement Learning from Human Feedback (RLHF)
   e. Enhance generative LLM inferences by feeding/combining with enterprise/proprietary ML model results
   f. Support heterogenous, distributed LLMs and ML models through hierarchy and federation
2. Access to distributed data for AI model creation, training and fine-tuning

AI is great, but there are factors to be aware of

Cost, time, accuracy/reliability, data availability/access and updates can be impacted by the following:

- Data quality – cleansing, transformation and standardization
- Data is generally not all available in one place – ideally, want access to multiple data sources in near real-time with no latency and accommodate updates
- Data deduplication and existing relationships can be hidden, both of which can greatly improve AI accuracy and potentially reduce tokens and therefore processing
- Data in external third-party data sources and AI models could be of great use, allowing access to better/more data and broader knowledge bases for AI data submission
- Data preparation to create views of data more amenable to AI model creation and use
- Data security and privacy
- Data use after AI is performed and incorporation in operational workflows

Main ways SmartData Fabric® can support AI in healthcare

1. Improve general AI data issues – avoid garbage in, garbage out through data cleansing, transformation and standardization to improve AI results, particularly, with structured data (demo)
2. Access data in multiple data sources to access the same type of data, e.g., multiple EHRs or a distributed data source, and/or different type of data, e.g., individual specialty network providers in healthcare, to:
   a. Assemble complete high quality datasets, perhaps fine-tuned to specific problems, to submit to external foundational AI models (demo)
   b. Enable high quality data extracts for subsequent AI modelling and training
   c. Support hierarchical AI models to avoid the need to move data and promote intelligence sharing in federation for faster, economical and reliable results
3. Incorporate MDM/MPI into data access and subsequent AI – connect data within and across data source through same entities identified within and across data sources (demo)
4. Interact with, and eventually, more seamlessly incorporate, external third-party data sources and AI models with own data sources and AI (OpenAI and ChatGPT demo)
5. Create better knowledge bases for AI models through enabling FHIR HL7-based virtual triple stores/graph databases, which standardizes entities, relationships and the reason for the relationships, greatly improving AI and almost anything else to do with data
6. Address security and privacy issues, including PHI and HIPAA, through intelligent anonymization – perhaps, use two-way masking, tokenization or encryption instead of one-way, to allow data lineage and drilldown (demo)
7. Enable non-AI access to data using SQL and/or other query languages to enable operations, reporting, BI and other analytics (demo)

SmartData Fabric® demo use cases

1. Nurse Assistant - Patient Problem Assessment upon a visit to a doctor appointment:

   Generate initial questionnaire to assess the patient problem based on the reason for the visit along with the longitudinal patient view (history). Generate a summary report based on the patient's answers for use by the doctor during the appointment.

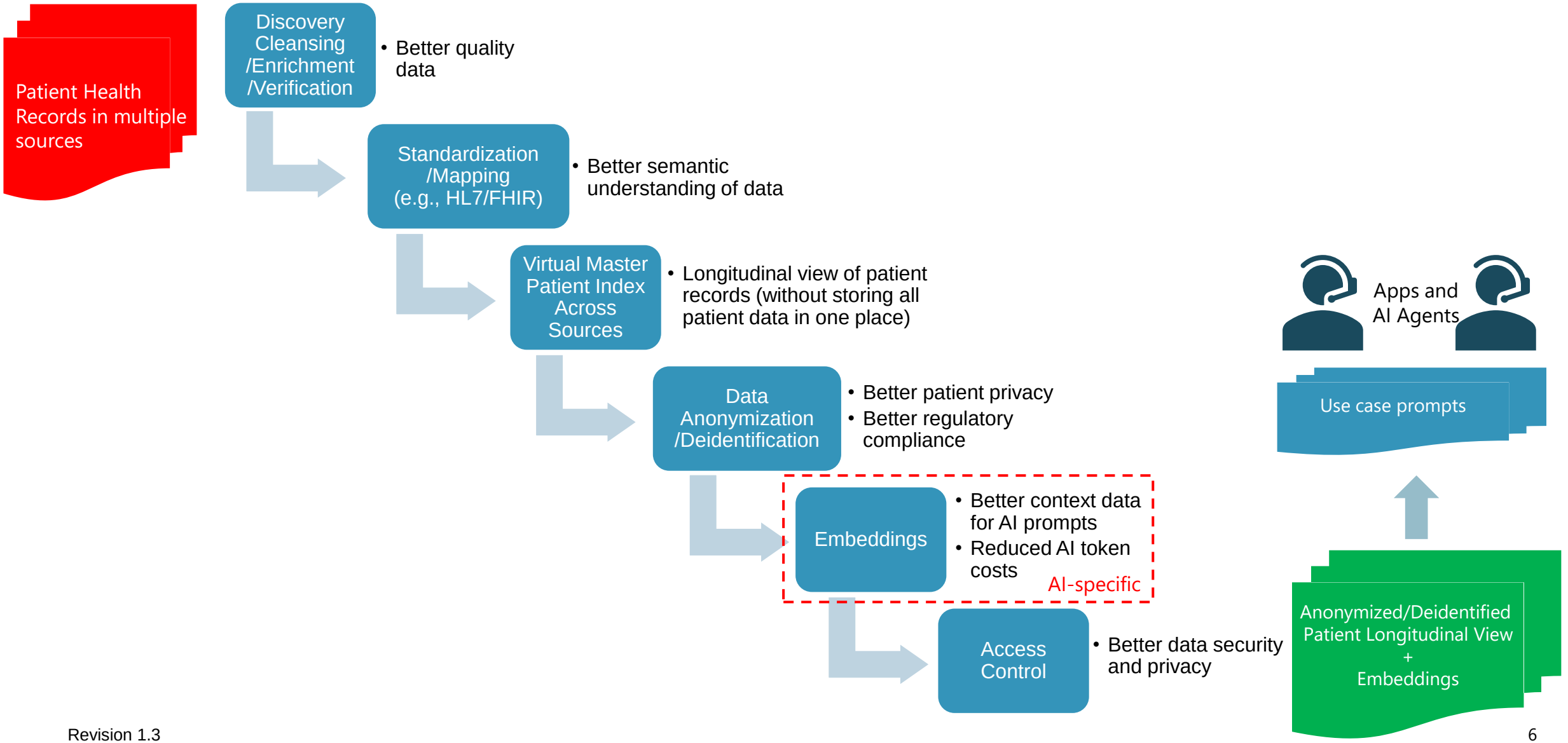
2. Physician Assistant - Patient Diagnostics:

   Generate a list of possible diagnostics for the patient based on patient's answers to the above questionnaire, patient history and further input from doctor.

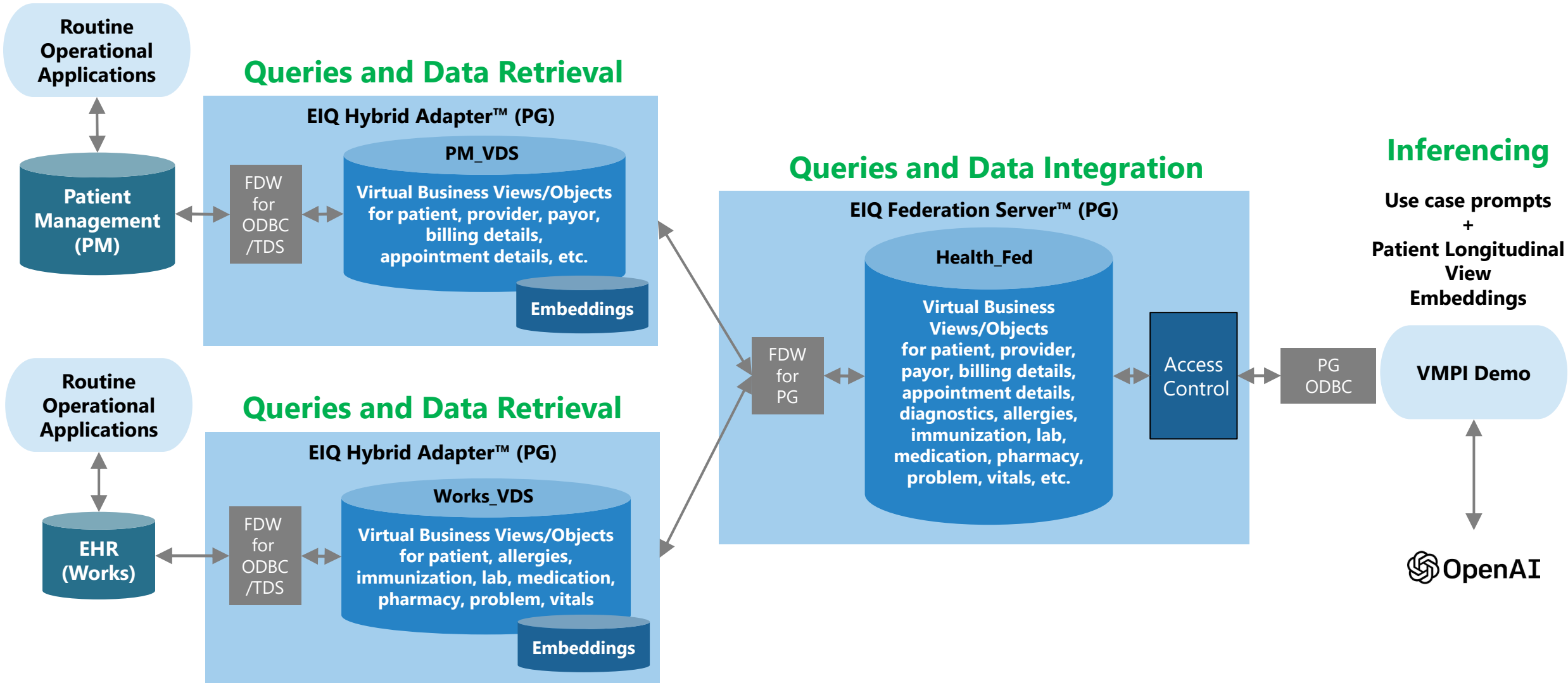
3. Discharge Assistant - Notes/Reports:

   Generate summary notes/reports for the patient based on patient visit information for subsequent doctor review.

Preparing Data for any app and AI



VMPI Demo with OpenAI using Federation of Patient Management (PM) and EHR (Works), with distributed MPI



SmartData Fabric® and AI, in general

- SmartData Fabric® data source adapters and federation servers are based on PostgreSQL
- PostgreSQL has many useful extensions, including AI-related that support embeddings, vector data and search on that data
- Once a complete MPI-based longitudinal patient view has been assembled from multiple adapters and federations servers across the organization, the subsequent result set can be submitted to AI models in JSON format
  - Subsequent data for fine-tuning can also be assembled and submitted in JSON format
- As PostgreSQL supports a number of query languages, including SQL, Python, R and GraphQL, SmartData Fabric® can be used for high quality data extracts for subsequent AI modelling and training, and any other analytics
- In addition to AI, SmartData Fabric® supports ideal data fabric operations, reporting, BI, analytics and other applications

The End