DIGITAL
WHITEPAPER

Peraton

HEALTHCONCOURSE

While the broad industry shift to certified electronic health record (EHR) technology delivered legible digital records and more portable data, it came up short in quantifiably improving the cost, effectiveness and satisfaction of health care services. On average, it takes 17 years to integrate health care best practices into the flow of medicine. Assessing care quality in time to make a difference with high-risk patients is simply not feasible due to fragmented and poorly standardized population health data sets.

Information technology boundaries are obstructing the delivery of effective and efficient healthcare. These boundaries make healthcare more expensive and more difficult to navigate for everyone: patients, providers, caregivers, researchers and payers. Each of these stakeholders is negatively affected by legacy 1.0 EHR systems. In response to the Affordable Care Act (ACA) and Meaningful Use (MU) criteria, the health care industry rushed to purchase 1.0 systems before most technology standards were in place. The systems, built with proprietary data models at their core, resulted in long-term vendor lock. They are primarily built on legacy programming technology from the 1980's (see MUMPS or M, programming language) and they rely on insufficient HL7 version 2.x (Health Level Seven, international standard standards for transferring clinical and administrative data) messages to connect data from producers to an evolving number of consumers. The technology is optimized for an era of client-server architecture and does not translate well to the cloud. These tools are built to provide just enough information to submit the patient's bill. They were never intended to effectively help patients on their personal health journey.

The Peraton HealthConcourse team and our coalition of partners are preparing for Health IT 2.0 solutions—ones that provide "care without boundaries". To realize care without boundaries, we strive to enable complete interoperability of all assets that are shared or of mutual interest to multiple stakeholders. Namely, we aim to overcome the boundaries prohibiting data interoperability (the sharing of medical records), knowledge interoperability (the sharing of clinical decision support algorithms and models) and process interoperability (the sharing of workflow and situational context).

Our unique approach is a 2.0 solution suite: an affordable, scalable and highly secure digital integration hub (DIH) that embraces technology standards and makes care without boundaries possible. The HealthConcourse architecture makes a conscious separation between the "core platform", technologies that provide a common foundation, and the "services ecosystem", which is a plug-and-play, service-oriented integration mechanism with standards-based interfaces and a high degree of system modularity. The services layer allows for best-of-breed by leveraging the best technologies available for the job. We designed HealthConcourse this way because we do not need to recreate the wheel. In other words, we do not need to recreate solutions addressing the core ingredients of interoperability—message exchange, data transformation, terminology management, identity matching, etc.—because robust solutions already exist for these. We are facing a cohesion and orchestration problem, not a technology shortage problem. We focused less on specific point solutions, and more on cohesion across established solutions to provide a functional, harmonized ecosystem. As a result, HealthConcourse was developed to be future-focused: intentionally designed to minimize the cost to switch or add new capabilities when better or more cost-effective technology becomes available. In addition, there are no proprietary data models at the core of the HealthConcourse platform, which markedly reduces vendor-lock risk and the total cost of ownership.

HealthConcourse is successful because it is built on proven technologies at the core. Our internal data model is based on the ever-expanding Fast Healthcare Interoperability Resources (FHIR) standard. We leverage the industry-leading Smile CDR FHIR platform to manage data persistence and querying through FHIR application program interfaces (API). To assure maximum flexibility, scalability and efficiency, we selected the best data processing technologies, NiFi and Kafka, to create highly configurable data ingestion and service orchestration flows. Leveraging the leading Kubernetes-based container platform in the market, OpenShift, our platform is portable, cloud native—yet agnostic to proprietary cloud technologies—and maximally scalable to meet the health care industry's evolving data needs. We also implemented a market-proven suite of process automation tools to cut away the many frivolous data tasks that keep health care providers from having time to talk to patients and work on their wellness plans. These core capabilities provide the foundation for the plug-and-play services, give us the tools to integrate into large, complex and heterogeneous technology ecosystems, and ensure that HealthConcourse is able to meet stringent service level agreements and security requirements.

PURPOSE AND STRUCTURE OF THIS WHITE PAPER

In this paper, we describe the architecture and use of our cloud-agnostic, fully standards compliant DIH. Gartner introduced the DIH concept as an architectural pattern to address data and shared asset dispersion and "to enable high-scale access, minimize workload on systems of record and deliver additional value via use cases like analytics"[1].

We start by describing a technology-agnostic reference architecture for any robust healthcare oriented DIH. We describe how a platform-based approach can bring order and structure to the digital health ecosystem where data is currently siloed and care processes are not shared across boundaries. We share this architecture openly to contribute to a collective, community-informed point of view on how to solve complex health IT, computability and interoperability problems with platforms and standards-based approaches. Implementations of this reference architecture are intended to solve health IT's most vexing and costly problem–the lack of high quality and efficient cross-institutional patient care programs.

The second half of this paper provides a technical overview of HealthConcourse, Peraton's cutting-edge implementation of the DIH reference architecture. We outline technology and implementation choices and demonstrate how the reference architecture can be put into action. We offer details on a demo implementation leveraging a variety of technologies that mimics the complexity and heterogeneity of a real-world health care delivery IT environment. We close with recognition and attribution to our partners who share the vision of a multi-technology, best-of-breed, standards-based, service-oriented, platform-enabled ecosystem approach providing cohesion necessary to overcome boundaries.

HEALTHCARE DIGITAL INTEGRATION HUB – SOLUTION COMPONENTS AND REFERENCE ARCHITECTURE

Gartner's concept of a DIH is industry agnostic. They present a proven pattern to solve data and knowledge fragmentation issues that occur in healthcare as well as other industries. Figure 1 shows Gartner's architectural diagram for this pattern.

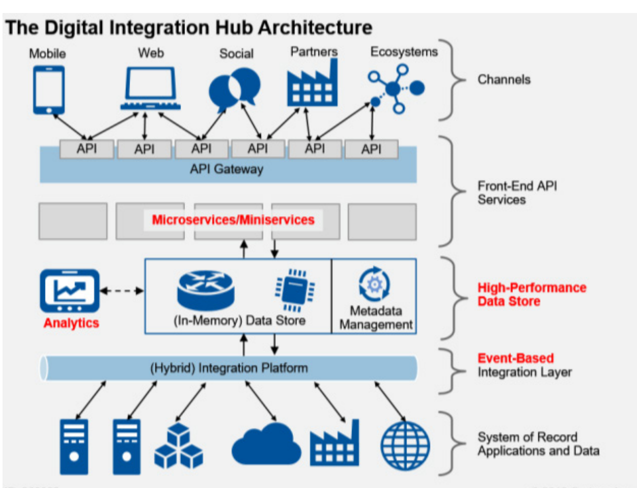


Figure 1: Gartner's Digital Integration Hub (source: Gartner, 2018)

Gartner portrays an integration platform that ingests data from a variable set of systems of record and stores the data, along with metadata, in an internal data store. A collection of microservices (based on the service-oriented architecture concepts) optimize this data for use cases appropriate for the business domain and provide access through APIs controlled through an API Gateway.

We embrace and extend this pattern by applying it specifically to healthcare. We describe the health-specific solution components, as well as the health interoperability standards and patterns, that tailor Gartner's concepts of an "integration platform", "data store" and "microservices" to health and unlock the value of the DIH for the healthcare industry.

We start with a discussion of fundamental solution elements: APIs and orchestration. We then provide a detailed description of our proposed reference architecture (aka, the solution "under the APIs") for an industry-wide healthcare DIH.

APIs—the tip of the iceberg

Accessing and harnessing data across a variety of systems is not a new problem. For decades, engineers have approached this problem using messaging strategies, data file or database sharing strategies and interface strategies. Recently, the use of RESTful (Representational State Transfer) APIs, or RESTful web services, combined with the rapidly emerging FHIR standard is accelerating and standardizing health care interoperability. FHIR-based, RESTful APIs enable health care data sharing across a federated and fragmented environment without necessitating data migration or locking up data in centralized solutions that sometimes reduces an organization's control of the data.

However, the use of APIs does not, by itself, solve the problem of data fragmentation, lack of standardization or sub-optimization across a complex ecosystem. FHIR-based RESTful APIs are just the tip of the iceberg. To fully leverage the benefits of FHIR and APIs, more must be done. Simply adding an API on top of an existing system of record is not enough because the raw data is typically not in a state or format that lends itself to interoperability without manipulation. For example, a recent review of data assets across the Veterans Health Administration concluded that the existing data "is unusable in its current state due to lack of normalization, duplication, being out of date and other noise and clutter. Our landscape analysis revealed a lack of a consistency when it comes to vocabulary, schema, and data formats."[2]

Even if these challenges are overcome, an API strategy ought to do more than wrangle data. Once data has been aggregated, standardized and made computable, new questions can be asked of the data and more can be done to automate routine data gathering and decision making processes. To optimize the health experience, a complete API strategy should:

1. Address the differences between the raw data from the systems of records (SOR) and the data requirements of FHIR and other interoperability standards. Data transformation and translation against industry standards helps to standardize the data into a canonical form.

2. Hide the complexity of the underlying data environment to the data consumer. APIs and the platform behind them provide the opportunity to encapsulate the SOR and shield the data consumers from the underlying data siloes and associated fragmentation. Aggregating the data from similar systems of record behind a single API reduces the burden on the systems of engagement that need this data. A thousand sources of data should not necessarily result in a thousand APIs.
3. Improve and enrich the raw data to maximize its use. Opportunities exist to enrich the data with provenance, security, privacy, tagging, conformance and other types of metadata.
4. Learn from the data. New insights in the form of calculations, inferences, assessments, recommendations, predictions, etc. can and should be derived from the data through analytics and evaluation of the data against evidence-based knowledge services.
5. Drive automation. Manual tasks are often time-consuming, error-prone and a cognitive burden on the end-user. Performing manual tasks requires a clinician to understand the best practice or clinical guidelines for a given health concern, to collect that data and assemble in their mind from a variety of sources, and perform calculations and evaluate logic that aid in clinical decision making. Much of this can be automated. When workflows are expressed in a computable, standard way, the execution of the sequence of activities, delivery of the data the activities require and or generate, and the evidence-based algorithms that support clinical decision-making steps can and should be automated.

What is "under the API" is just as important as the use of APIs themselves. The APIs should be the gateway into a robust DIH capable of leveraging a variety of technologies to enable the APIs and maximize the usefulness of the data behind it. Further, a DIH and what is "under the API" is only as good as the orchestration used to provide cohesion allowing the technologies within a DIH to work together harmoniously.

Orchestration—the glue that holds it all together

A recipe without step-by-step instructions doesn't make for a good meal. A shopping mall without a center aisle connecting the shops doesn't make for a good shopping experience. Likewise, a collection of health IT solutions without orchestration doesn't result in a DIH.

Imagine that you have a risk calculator that can determine the likelihood that a patient has a certain medical condition if provided six different data points in a certain format encoded against pre-defined medical terminologies (e.g., LOINC). For the sake of example, two of the data points are numeric vital signs, one is a confirmed medical condition/diagnosis, one is a lifestyle factor (e.g., smoking or socioeconomic status), one is a family history of a medical condition/diagnosis and one is a diagnostic test result (e.g., imaging study or lab value). Now, assume that these six data points are captured and stored in different systems used by different providers working in different institutions. Specifically, assume that one of the vital signs is a structured numeric value stored in an electronic medical record (EMR), the other vital sign is captured via a medical device; the confirmed

medical condition/diagnosis is recorded as a SNOMED CT or ICD-10 code in a medical claim; the family history of a medical condition is reported verbally by the patient; the imaging study is available as a narrative report written in natural language (i.e., free text or unstructured data); and, the lifestyle factor is inferred based on observation of patient behavior.

In a digital world, the goal is to remove the burden from the clinician to:

- Understand what data is required for the risk calculator
- Identify and extract the necessary data from the different sources
- Transform the data from its native representation into the format required by the calculator—including extracting clinically relevant data from free text
- Infer or compute data where needed
- Feed the data into the calculator so that an automated calculation can be performed

The good news is that tools exist for all of these specific problems. Transformation engines can transform claims data into FHIR resources. Terminology engines can map between code sets (e.g., ICD-10 to SNOMED CT) or assign codes to textual descriptions. Natural language processing engines can extract key data from free text notes. Business rules, clinical decision support and other algorithms can be written to infer or compute data points. In today's health IT environment, there is an abundance of point solutions that do a specific job very well. The problem isn't a lack of technology, the problem is a lack of cohesion across all technologies so they can work together in the correct sequence, have their pre-conditions evaluated and confirmed and be coordinated with the appropriate handoffs. This is what orchestration does. It programmatically combines and sequences the technologies into a flow with clear lines of demarcation, clear handoffs and a clear outcome. It makes the technologies work together harmoniously, even if they were created as independent units of software by different organizations in different programming languages at different times. Robust, configurable, scalable, governable orchestration is one of the most important things a DIH must do well.

Under the APIs—a reference architecture for a healthcare Digital Integration Hub

This reference architecture describes the functions and components that constitute the solution "under the API." The platform-based approach described by this reference architecture emphasizes modularity, interoperability and a common framework for optimizing how different solution components interact within a cohesive ecosystem. The reference architecture details the common elements required to effectively manage health care data across multiple systems of record in a fragmented and federated data landscape to reduce risk, cost and time to market. Implementations of this reference architecture, including HealthConcourse, provide the framework for delivering, managing, orchestrating and optimizing multiple health care services to address the needs of several stakeholders and use cases requiring access to the patient's digital health record and other digital assets supporting health organization services (e.g., care delivery, customer service, operational management, analytics and asset management).

Separation of concerns

The DIH concept is predicated on the idea that health IT ecosystems comprised of multiple EMRs, and other health-related data stores, can no longer rely on a single data source or system of record for managing the complete life cycle of health data. Further, no one system in a complex ecosystem can be solely responsible for managing data processing to drive all business processes or deliver integrated data views to all business users. The DIH reference architecture is an implementation of "separation of concerns"—a design principle for separating solutions into distinct layers, components or sections that individually satisfy a core business requirement or "concern."

At the macro level, a DIH and the ecosystem it operates within are separated into three distinct layers as shown in Figure 2. The proliferation of multiple systems of records (where fragmented data resides) and multiple systems of engagement (where consolidated and standardized data is needed) highlights the business problem. Essentially, multiple systems of record exist, including EMR platforms, non-clinical business systems, data warehouses, patient-generated data stores, medical devices, applications and others that each contribute a portion of data to the overall patient record.

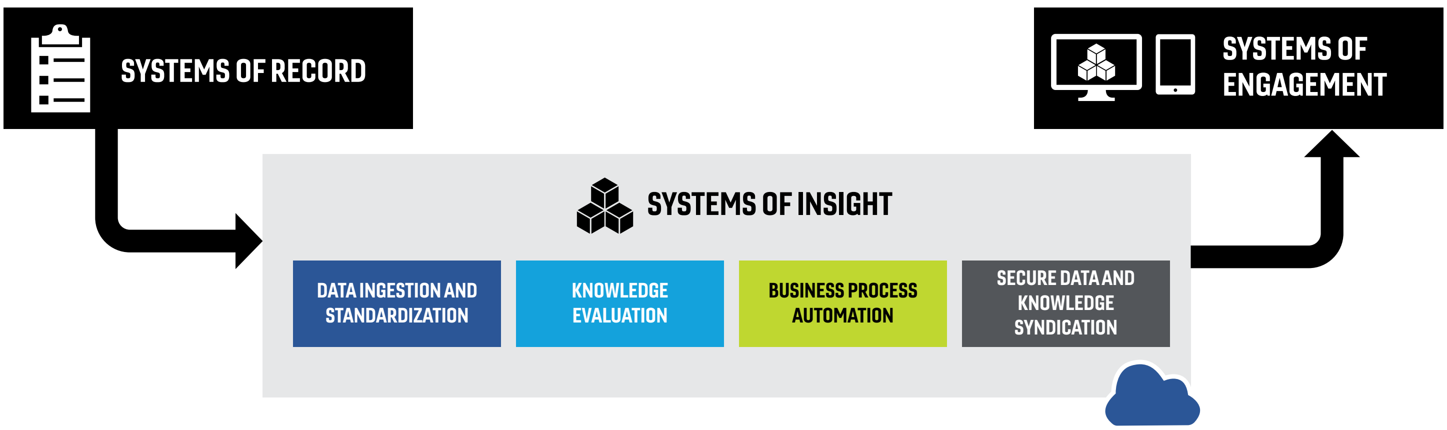


Figure 2: DIH separation of concerns

In today's complex health delivery organizations, the systems of record are a mix of older and newer solutions developed in-house or delivered from a variety of health care solution vendors. In nearly all cases, the data across these systems of record are not aligned to a canonical form and have variable conformance to industry data models, data exchange standards and industry-wide medical terminologies (e.g., SNOMED).

When applying the "separation of concerns" principle, the systems of engagement are decoupled from the systems of record. We see this pattern applied in health care with health information exchanges, SMART on FHIR applications[3], mobile applications, patient portals, analytics dashboards and other systems of engagement where the end-user application is decoupled from the data sources.

The benefits of separating the systems of engagement from the systems of record include:

- Better modularity so that individual components can be changed or replaced without cascading impacts across the ecosystem
- The ability to have a best-of-breed environment with minimal vendor lock-in and greater portability of applications and components
- Longevity and portability of the overall architecture and ecosystem to ensure the systems can continue to provide business value as dependencies and infrastructure change over time
- Greater reusability of common capabilities to increase consistency, governance and time to market

However, separating the systems of engagement from the systems of record can present challenges when each system of engagement has to redundantly resolve underlying data issues. The systems of engagement are encumbered by the fragmentation, lack of standardization and sub-optimization of data access and utility across a heterogeneous collection of systems of record. To address this challenge, the DIH fulfills the role of the system of insight.

A system of insight is an intermediary between the systems of engagement and the systems of record that hides the complexities of the underlying data architecture to ensure that systems of engagement have the information they need to satisfy business requirements. The system of insight is the solution "under the API."

Healthcare DIH core elements

Table 1 and Table 2 define the core elements of a healthcare aligned DIH. These core elements should be implemented as a set of modular services (or microservices) orchestrated to work cohesively as part of a well-managed ecosystem based on a service-oriented architecture (SOA). However, each core element doesn't necessarily correspond to a single unit of software or software component. The actual number, specificity and granularity of software components is open to interpretation provided that the following core elements are not collectively delivered as a single monolith application unable to achieve the benefits of extensibility, modularity, encapsulation and plug-and-play solution component migration over time.

After data is retrieved from the systems of record, a healthcare DIH can do a variety of things to make the data more usable and to then use the data for intelligent things before delivering it and other findings to the systems of engagement. Table 1 organizes DIH capabilities and services into four logical sub-divisions:

1. Data ingestion and standardization–responsible for standardizing and enriching data
2. Knowledge evaluation–responsible for managing and evaluating executable knowledge services
3. Business process automation–responsible for ingesting, running and automating executable process models
4. Secure data and knowledge delivery–responsible for delivering data and knowledge to consumers and managing appropriate access

DHP core elements: capability area	Description
Data ingestion and standardization • Transformation services • Terminology services • Identity management services • Unstructured data and natural language processing services • Data validation services • Labeling and metadata services • Deduplication services • Record linking services	Standardization capabilities take data from the systems of record in whatever format available and standardize the structure and contents. This includes capabilities to map and transform inbound data structures to FHIR resources (or other preferred data exchange standards) and resolve differences in terminology and identities (e.g., patient, provider and organization identifiers). Data structure identifiers should be name-spaced to avoid identity collisions when data is aggregated from multiple sources. A key capability is the extraction of structured data from free text using natural language processing (NLP) to turn unstructured data into more discrete, structured data elements aligned to the canonical model. Enrichment capabilities enhance the base data resources by adding metadata, semantics and context to better index, describe, relate and/or classify the data. Common metadata includes data validation/conformance information, data classifications and information about data relationships. Data classification or segmentation is very powerful for sub-dividing the data to enable functionality like fine-grained consent, attribute-based access control and clinical research. Enrichment also includes capabilities that helps to manage the data relationships that exist once the data is aggregated. Managing data relationships includes identifying and reconciling data duplicates (e.g., duplicate FHIR resources) as well as creating linkages between records (e.g., relationships between different FHIR resources). For example, data reconciliation will weed out duplicate medications in a patient's medication list, whereas record linkages will add information about logical, evidence-based relationships between medications and the medical conditions they treat.
Knowledge evaluation • Real-time analytics services • Clinical decision support (CDS) services • Calculators and computation services • Clinical quality measures (CQM) services	These capabilities are what differentiates a robust DIH from a standard message broker or service bus. Capabilities in this layer focus on management of and access to knowledge services (e.g., CDS, CQM and calculators) and analytical and statistical models (including artificial intelligence and machine learning models (ML and AI)) to extrapolate insights from the data aggregated from the systems of record. This layer is the essence of a learning health system and is what allows the system to inform and guide clinical processes and decision making through evidence-based algorithms from vetted clinical research.
Business process automation • Computable BPM+ Health models • Process automation managers • Model to data bindings • Model to knowledge bindings • Workflow and notification engines	A data and knowledge platform is only useful if humans and other systems use it. Adding business process automation to a DIH provides a mechanism to incorporate, run and automate business processes to help contextualize where data and knowledge are needed. Models must be expressed in a computable format aligned to standards like BPM+ Health. They should indicate where they need to bind data and knowledge to automate menial human tasks and optimize compliance to best practice and user experience. The models should be able to run as software in a run-time environment on top of process automation managers and/or workflow engines. When workflow models are instrumented this way, the DIH gains "situational awareness" by understanding what data is needed when and where, by whom and for what purpose. This transitions a DIH from being a reactive CRUD (create, read, update, delete) service to a proactive enrichment platform.
Secure data and knowledge syndication • Secure APIs • Access control services • Privacy and consent services	This layer includes capabilities to expose the aggregated, standardized and enriched data set and inferred insights to the systems of engagement. This is done through APIs based on standard web protocols (HTTP, REST) and health care standards (FHIR), and through asynchronous message exchanges such as "publish and subscribe". Security controls for authentication, authorization and access control ensure appropriate access. Consent management appropriately filters data based on a variety of variables including the role of the data user, their purpose of use, their organization and other policy based filters.

Table 1: DIH capabilities and services

Another way of depicting the layered approach categorized in Table 1 is to visualize an increasing degree of interoperability and automation capability achieved by systematically adding standards, patterns, controls and related constructs to data as it moves from left to right in Figure 3.

- Starting on the left, data is federated, unstandardized, heterogeneous and often times non-computable. One common practice is to extract and load this data into a central repository, which, will produce an aggregated, yet still heterogeneous data collection. This step alone may be good enough for some analytics and business intelligence platforms, particularly those that operate on the Extract-Load-Transform (ELT) process whereby data is transformed into a more usable state on a use case by use case basis. However, the capabilities of a DIH require more than this.
- The next step is to standardize and optimize this data to produce a computable, normalized data collection. This is the basis for healthcare data interoperability (data interop). A DIH that offers this should provide solutions that normalize the data against a common schema and industry-standard medical terminology. Additionally, data is made computable when identifiers linked to people, organizations, services, etc. are disambiguated and reconciled against a master index. Finally, the application of optical character recognition (OCR) and NLP technologies help extract clinical concepts from unstructured data to increase the computability of this data.
- Once data is computable, you can learn something from it programmatically and algorithmically. Using knowledge standards such as CDS-Hooks and the Clinical Quality Language (CQL) to standardize knowledge expression and delivery democratizes clinical decision support from proprietary EMR systems. This is the basis for shareable knowledge services and knowledge interoperability (knowledge interop). When paired with an event driven architecture, knowledge services can dynamically listen for data and infer or compute new findings in real-time. The algorithms within the knowledge services can be traditional rule-based decision trees or can leverage exciting new advances in probability-based and self-adjusting ML and AI models.
- The last layer of sophistication comes when process standards are applied that implement and deliver clinical workflow as API-enabled, standardized and evidenced-based process services. These process services reflect the processes, decisions, tasks, activities and roles modeled in a standard syntax such as BPM+ Health. A DIH capable of interacting with process services gains situational awareness and an appreciation for when data is needed by who and for what action. Further, process services using BPM+ Health provide a mechanism to democratize control flow from proprietary EMR systems and realize process interoperability (process interop).

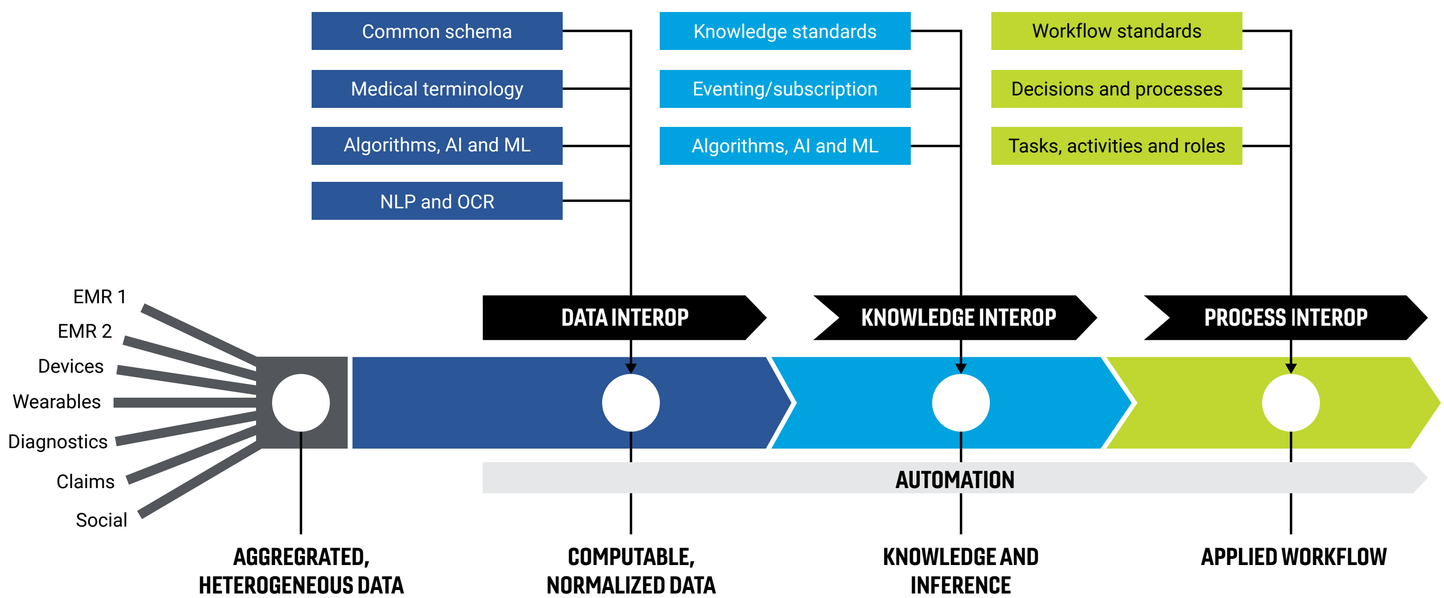


Figure 3: Incremental automation and interoperability enabled by a DIH

The features detailed above need a common chassis in which to operate. They all share requirements related to orchestration, data persistence, auditing, error management, and the like. The additional core elements in Table 2 are infrastructural in nature providing this core and cross-cutting functionality. These elements are not healthcare specific but are nonetheless needed to enable a healthcare DIH. These elements could be implemented as microservices in a SOA or could be native capabilities of the underlying service management platform.

DHP core elements: platform capability/services	Description
Data retrieval	SOR adapters retrieve data from the systems of record using messaging, API and database adapters so that data can be acquired and integrated into the DIH and subsequently transformed from the source format into a defined canonical model.
Orchestration	This area includes capabilities to manage the data processing pipeline by managing and executing configurable internal data processing workflows and calling business logic modules or external services as part of a SOA.
Persistence	This area includes capabilities for managing the data within the DIH as it is acquired, integrated and processed. This foundational capability includes the ability to store the data in long-term databases (e.g., data lake) or temporary caches and index the data for rapid search and retrieval.
Provenance	Data provenance is a critical DIH feature responsible for tracking all changes made to data structures and content as it passes through the DIH data ingestion pipeline. Every new resource or change to an existing resource results in the creation of a new provenance resource.
Auditing	These capabilities keep track of who accesses data and other pertinent details about system behavior.
Error management	Each contributing service or component participating in the DIH ecosystem should manage and raise exceptions in a common way. This capability provides a framework for consistently reporting and managing exceptions.
Surveillance	Surveillance monitors data into and out of a DIH to trigger downstream actions and enable an event-driven architecture. Using the surveillance capability, solution components can trigger, or listen for and respond to, events based on the availability of new information.
Knowledge	Knowledge is overarching solution capability to represent, manage and invoke knowledge services (e.g., algorithms, business rules, decision support, analytics, etc.) on top of the data managed by the DIH.
Workflow	Workflow is the ability to define, manage and execute a sequence of activities relevant to a business problem, process or domain that runs within the DIH.
Automation	Automation refers to the ability to remove unnecessary human tasks and manual intervention in support of surveillance, knowledge execution, workflow or other DIH supported functions. Automation commonly occurs in the form of automated, contextually-correct data or knowledge delivery at the right time and place within a supported business process.
Query management	Query management is the capability to run standards-based queries against the DIH data set and identify, correlate, process, filter and ultimately return requested data.
Secure APIs	Secure APIs are the secure endpoints exposed to data consumers (systems of engagement) providing access to data and functionality provided by the DIH in a secure manner enforcing appropriate authentication, authorization, access control, consent/filtering and data encryption mechanisms.

Table 2: Platform capabilities

Figure 4 provides a conceptual representation (reference architecture) of how the core elements and services—listed in table 1 and table 2—are organized, relative to each other, to create a cohesive system of insight. Any DIH implementation must be cloud aware (preferably cloud native in its design and architecture) and should leverage the computing, network, security and scalability capabilities of modern cloud platforms. However, to achieve a high degree of portability, the solution components of a DIH should be containerized into a common, robust container platform that insulates and abstracts the solution components from the proprietary features of any underlying infrastructure (e.g., specific cloud platforms).

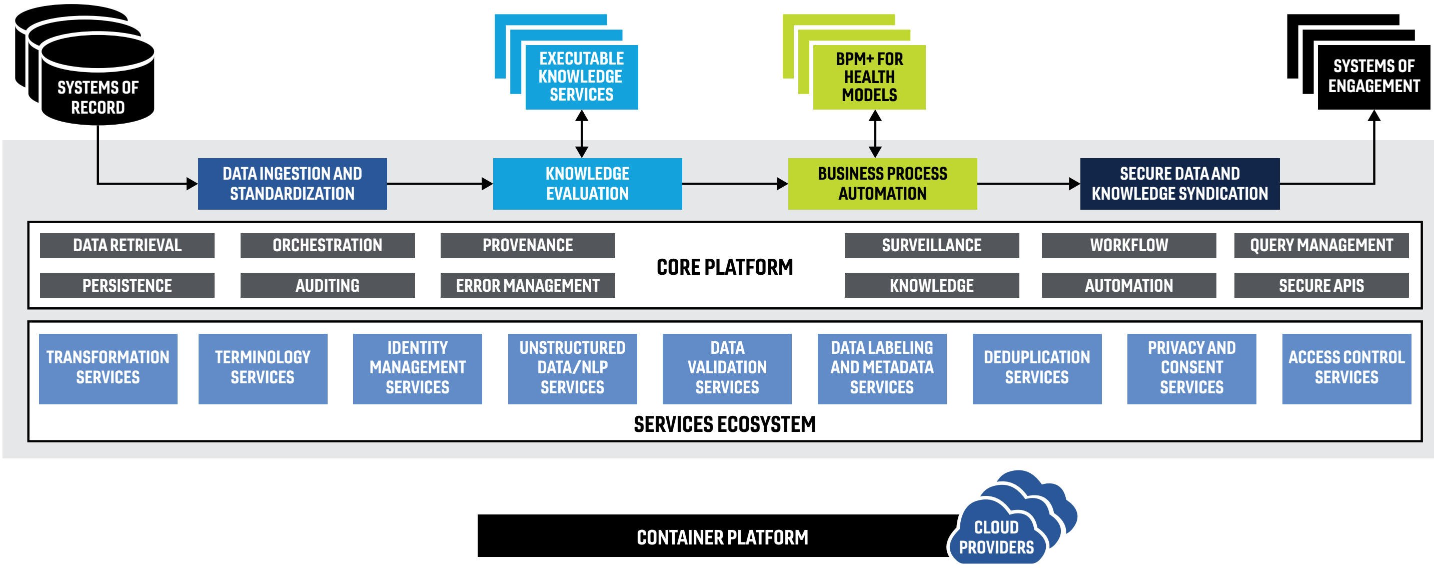


Figure 4: DIH conceptual reference architecture

This DIH reference architecture aligns well to health ecosystem descriptions from leading health care organizations such as Logica Health and HL7. The reference architecture incorporates several of the core services and capabilities described in the 2018 Logica Health Roadmap[4], including the use of the HL7 FHIR specification[5] for data transport and API definitions, Logica Health recognized standard medical terminologies, terminology services, the CDS Hooks specification[6] and data segmentation used for patient right of access and consent[7].

The reference architecture also incorporates several of the core services defined in the HL7/OMG (Object Management Group) Health Services Specification Project (HSSP)[8], including message transport services, terminology services, health services directory services, identity management services, data retrieval services, data mapping and transformation services, event management services, health care security services and clinical decision support services. Implementations of this reference architecture are encouraged to be faithful implementations of the API specifications and related requirements for the services defined by FHIR, Logica Health and HSSP.

HEALTHCONCOURSE—PERATON'S DIH—ENABLING CARE WITHOUT BOUNDARIES

HealthConcourse is Peraton's healthcare DIH based an open architecture and open, standards-based APIs built on the FHIR standard. HealthConcourse is a FHIR-first solution. FHIR resources are used as the canonical data model within HealthConcourse, and FHIR API specifications are used to define the primary set of HealthConcourse APIs. HealthConcourse transforms the raw data retrieved from the systems of record into FHIR resources. The data ingestion pipeline of HealthConcourse includes a configurable collection of services that implement many of the capabilities defined in the DIH reference architecture above (e.g., terminology services, NLP services, identity management services, etc.). Serialized FHIR resources are used as the input and output of most of the microservices, resulting in standards-based internal interoperability that enables modularity, encapsulation, agnosticism and a plug-and-play approach.

Figure 5 extends the DIH reference architecture by identifying the HealthConcourse technologies and solution components used to implement the services and capabilities illustrated in the reference architecture. Technologies in the "core platform" are integral to HealthConcourse and are part of any HealthConcourse implementation. The technologies and solution components in the services ecosystem are highly configurable and flexible. They can be excluded where requirements and use cases do not require them and the technologies actually implementing the capability can be replaced with alternatives. Elaboration of this technology plug-and-play aspect of HealthConcourse is further illustrated in Figure 5.

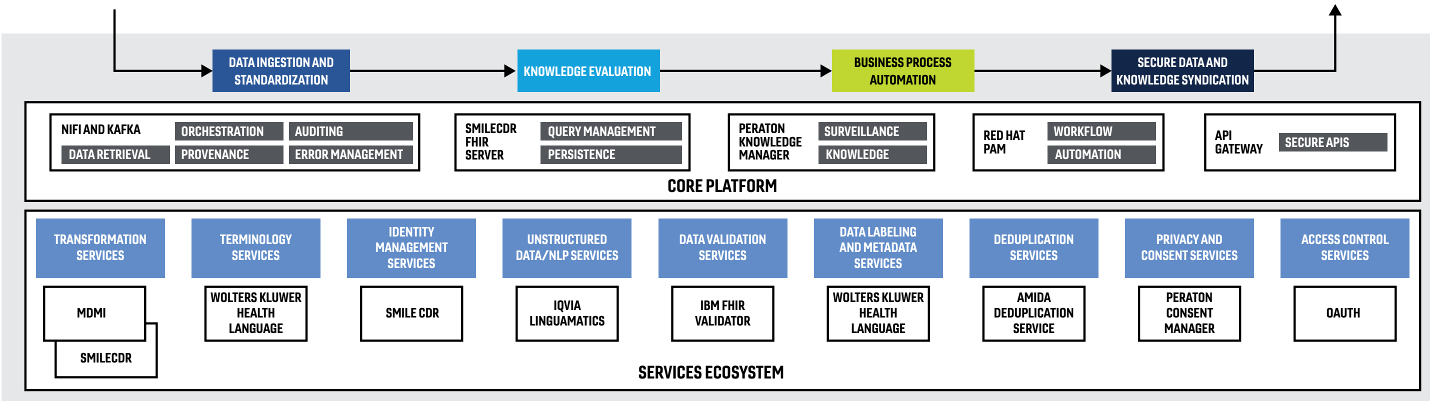


Figure 5: HealthConcourse conceptual solution architecture

HealthConcourse core platform

The HealthConcourse core platform is made up five core components.

Nifi and Kafka: NiFi and Kafka are industry proven, open source technologies. HealthConcourse includes a very robust data orchestration architecture using NiFi and Kafka that handles all data retrieval, ingestion, orchestration, auditing, provenance and error handling. NiFi was originally created by and for the National Security Agency (NSA). It was architected from the ground up to address security, high data throughput and scalability. Likewise, Kafka—initially built by LinkedIn to handle massive amounts of data transactions and stream processing—is built with high data throughput and scalability in mind. Data is acquired and integrated into HealthConcourse through adapters managed by NiFi and/or Kafka (depending on the exchange patterns and data access mechanisms provided by the systems of record). HealthConcourse will actively pull data from systems of record or will deploy listeners that receive data passively from systems of record. Each system of record can provide a different data access mechanism and use a variety of data exchange patterns. As such, HealthConcourse leverages and extends NiFi's extensive collection of adapters and highly configurable, scriptable data ingestion capability and uses Kafka's library of adapters and internal asynchronous communications capabilities. The flexibility afforded by using these two tools allows HealthConcourse implementations to have custom orchestration flows so that each customer or use case can use the best data processing flows, controls and microservices for their requirements.

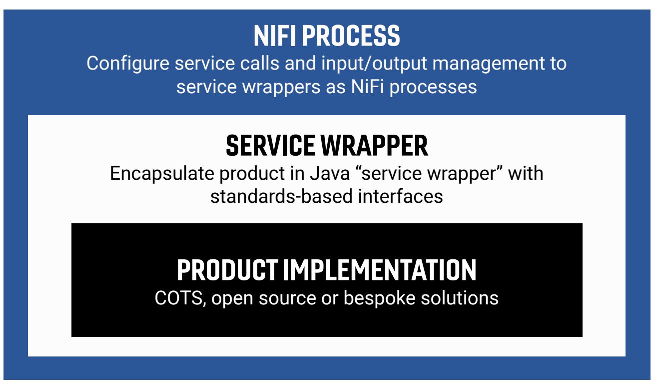
Smile CDR FHIR Server: The Smile CDR FHIR Server is best-in-class. It is the enterprise-strength, commercialized version of the ubiquitous HAPI (HL7 API) open source FHIR Server. Smile CDR (and HAPI) is a complete implementation of the FHIR standard. Smile CDR is part of the core fabric of HealthConcourse providing the internal FHIR server that persists FHIR resources (aka, the HealthConcourse cache: the data store described in the Gartner DIH) and provides the APIs and access mechanisms exposed to internal services as well as external systems of engagement for data retrieval. HealthConcourse includes multiple instances of Smile CDR optimized for different purposes: storing and managing clinical records, storing and managing administrative and product configuration data, and storing and managing master person index (MPI) and master data management (MDM) information.

Peraton knowledge manager: Peraton teamed up with Dynamic Content Group (DCG)–who employs recognized leaders in the emerging field of standards-based, executable knowledge artifacts and knowledge expression languages such as CQL–to build the knowledge manager component of HealthConcourse. The knowledge manager is a revolutionary, first of its kind, knowledge broker that dynamically discovers and evokes knowledge services to evaluate and learn from a patient's clinical records. Knowledge services are computable expressions of knowledge (e.g., clinical decision support rules, quality metrics, machine learning models, statistic computations, medical calculators, algorithms, decision trees, etc.) delivered as microservices with standards-based APIs. The most commonly used API standard for knowledge services is CDS-Hooks. By leveraging standards, the knowledge manager works with any number of knowledge services and dynamically adds or removes knowledge services as needed. The knowledge manager maintains a registry of knowledge services and, through the knowledge service discovery process, learns and maintains information about the data requirements, pre-conditions, usage context/requirements and outputs of each knowledge service. Armed with this information, HealthConcourse determines if new data ingested and persisted into the HealthConcourse cache is appropriate and sufficient to evoke a knowledge service. The knowledge services run their algorithm against the data provided and return new information (e.g., risk assessments, computed results, guidance, recommendations, etc.) that is added as new FHIR resources into the patient's aggregated data set. This results in "closed-loop analytics" that dramatically reduces the time it takes to get knowledge, and the information resulting from knowledge services, into the hands of clinicians, patients and decision makers. The knowledge manager is an important step toward the democratization of knowledge: freeing important business rules, models and algorithms from proprietary and opaque systems and enabling a federated, standards-based knowledge community where knowledge is authored external to record management systems (e.g., EMRs) by numerous people and organizations and ingested and leveraged within robust DIH platforms like HealthConcourse.

Red Hat® Process Automation Manager: The Red Hat Process Automation Manager (PAM) is the run-time BPM+ Health engine. BPM+ Health is an evolution of the popular business process modeling notation (BPMN) standard historically used for workflow and process engineering. BPM+

Health extends BPMN by adding additional standards for decision modeling and case management modeling. When BPM+ Health is applied to health care, clinical protocols, pathways and guidelines traditionally delivered through non-digital means, and not integrated into the tooling used by health care professionals, can be reimagined as computable expressions of process knowledge. By modeling a clinical pathway (e.g., the sequence of activities and decisions recommended to treat a common condition like hypertension) as a BPM+ Health model, the pathway can be implemented as software with automation of tasking and workflow. HealthConcourse takes this one step further. By leveraging and extending the PAM, HealthConcourse uses BPM+ Health models to know exactly when certain data and/or knowledge services could be used within the context of a clinical workflow. This allows us to convert the menial human tasks of finding specific data in a patient's longitudinal health record and performing manual calculations or inferences into an automated delivery of FHIR data and standards-based knowledge services. In essence, the HealthConcourse implementation of the PAM reduces the cognitive burden on clinical end-users while simultaneously aligning clinical activities to the evidence and best practice expressed in the BPM+ Health models.

API Gateway: HealthConcourse is an API based platform. Our API Gateway helps to organize, document and control access to the APIs used by systems of engagement to access patient data and HealthConcourse services. The API Gateway helps enforce security controls and monitor API usage.



HealthConcourse services ecosystem

HealthConcourse is based on a SOA underpinned by the core platform. Specific, discrete functionality is encapsulated in microservices and accessed through predictable, standards-based APIs. HealthConcourse takes a 3-layered encapsulation and abstraction approach to integrating microservices into the platform.

- **Layer 1:** NiFi is used to define the orchestration flow for microservices and control the inputs and outputs. Each time an external microservice is required, a NiFi process is created to manage the interactions with the microservice and incorporate it into the larger orchestration flow.
- **Layer 2:** HealthConcourse aims to enable a plug-and-play approach to technology integration to maximize portability, leverage best-of-breed, and reduce/manage vendor-lock. For each microservice required by HealthConcourse, a service wrapper is created to encapsulate and manage the underlying technology that satisfies the requirements of the microservice. The service wrapper API is exposed to NiFi (and other consumers of the microservice) to abstract the consumers of the microservice from the specific technology "under the hood" for a given build at a specific point in time.
- **Layer 3:** The best technology for the job is selected for each microservice. Preference is given to technologies that meet all functional requirements, have high performance and throughput, are stateless and proven to scale and enforce rigorous security controls. HealthConcourse is built on top of a pre-selected set of technologies (as shown in figure 4), but because each technology is encapsulated behind a service wrapper and NiFi processor, the technologies can change based on specific customer requirements or new advances in technology.

The actual number of microservices needed by HealthConcourse changes over time as HealthConcourse evolves. The number of microservices is also variable based on specific implementation requirements (not all microservices are needed for all customers). That said, there are a common set of microservices that make up the bulk of the HealthConcourse services ecosystem.

Transformation services: Microservice(s) that map and transform data structures and formats provided by systems of record into FHIR resources. The default and preferred technology used for this service is the open source MDMI (model driven message interchange) product. Smile CDR and NiFi are also used for FHIR transforms in some situations.

Terminology services: Microservice(s) that map and translate coded clinical concepts (i.e., clinical concepts that are bound to standard medical terminologies like SNOMED CT) to the preferred text description and code for that concept. Preferred terminology bindings are determined by the FHIR specification or through FHIR profiles. The default and preferred technology used for this service is the Wolters Kluwer Health Language terminology server.

Identity management services: Microservice(s) that manage identities and resolve identity differences for entities including people. Probability-based matching algorithms are commonly used to quantify the likelihood that information from multiple sources does or does not represent the same person (or other physical entity). The default technology used for this service is the Smile CDR MPI. Provider and organization directories also fit into this category of services.

Unstructured data/NLP services: Microservice(s) that use natural language processing to mine free-text data to identify and extract important clinical concepts such as problems, observations, tests, allergies, medications, etc. Extracted concepts are mapped to standard terminologies and turned into new FHIR resources added to the patient's integrated data set. This effectively converts data that was previously only human readable into computable, machine readable data that is useful for analytics- and algorithmic-based computations. The default technology used for this service is the IQVIA Linguamatics NLP engine.

Data validation services: Microservice(s) that validate FHIR resources against the FHIR specification and against FHIR profiles to identify issues and measure the quality of the data. Using FHIR profiles for validation allows us to add, edit or remove profiles on demand to dynamically configure the validation rules. The default and preferred technology used for this service is the open source FHIR Validation Server provided by IBM.

Data labeling and metadata services: Microservice(s) that add metadata to the FHIR to improve the utility and understanding of the data. The data labeling service, for instance, analyzes records against data sensitivity categories and adds metadata tags to the FHIR resource indicating if the contents are sensitive or restricted. Doing this allows us to implement fine-grained consent management whereby patients can decide exactly who can see specific portions of their longitudinal records and for what purpose. The default technology used for this service is the Wolters Kluwer Health Language Code Group Manager.

Deduplication services: Microservice(s) that identify record duplicates that invariably arise when clinical records are aggregated from the same patient across different sites. FHIR Linkage resources are created to capture information related to the duplications between records. The default technology used for this service is the deduplication and reconciliation product built in collaboration between Peraton and Amida.

Privacy and consent services: Microservice(s) that create and manage patient consent provisions and filter FHIR query results based on consent rules. This capability leverages the sensitivity tagging from the data labeling and metadata services. The default technology used for this service is the Peraton consent management and filtering solution.

Access control services: Microservice(s) that enforce security controls. Access control services may not strictly be implemented as microservices the way other services are. The Open Authorization (Oauth) framework is used as part of this service.

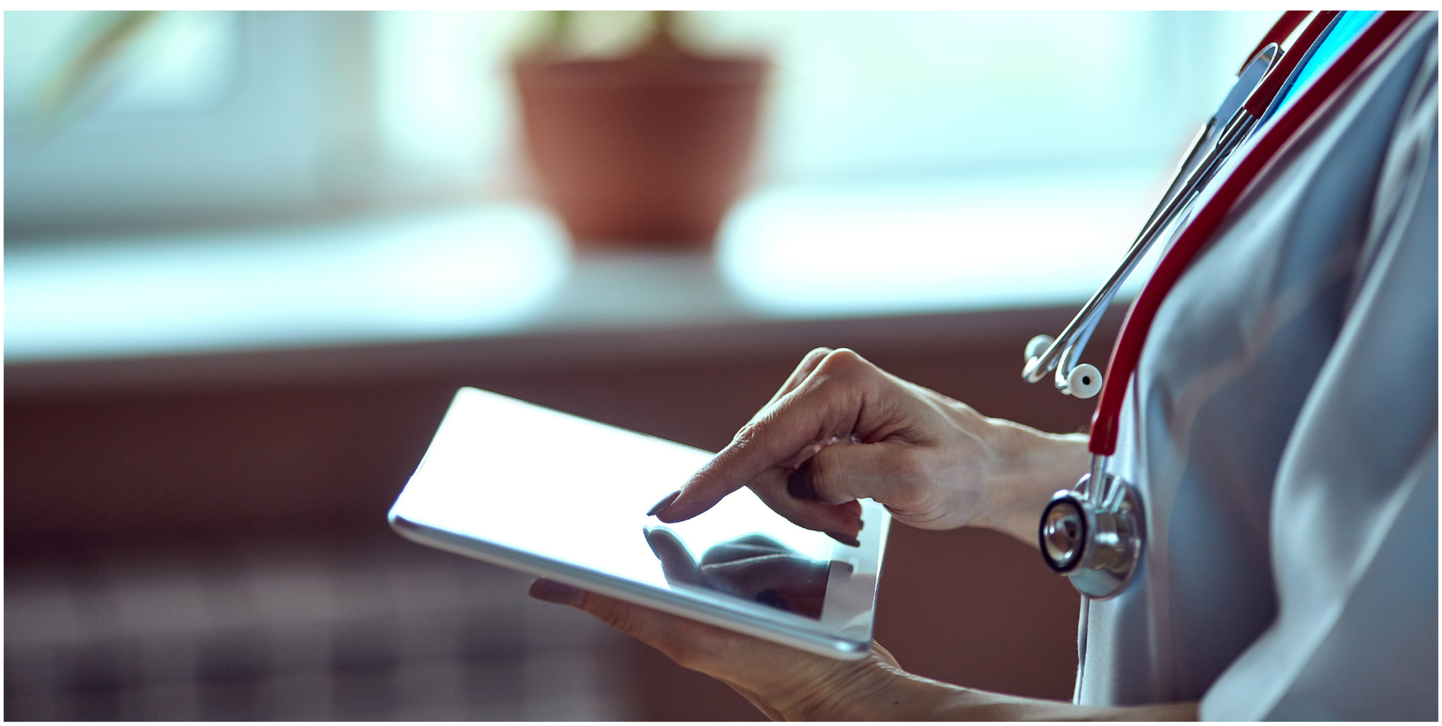
HealthConcourse infrastructure

HealthConcourse infrastructure is organized into three primary categories: cloud hosting, containerization, and security.

Cloud hosting: HealthConcourse is a cloud native platform, but it is intentionally cloud-agnostic. To be cloud-agnostic, HealthConcourse is built on technologies that are self-contained and do not rely on proprietary cloud provider features. HealthConcourse leverages cloud compute, security and network capabilities to assist in scalability, access and cybersecurity. HealthConcourse minimizes reliance and encapsulates cloud capabilities for storage and application functionality to avoid vendor-lock. HealthConcourse has been built, tested and deployed in AWS and Microsoft Azure, the two most prominent cloud platforms on the market today.

Containerization: HealthConcourse is completely containerized using Docker and orchestrated using Docker Enterprise Kubernetes supplied by the Red Hat's OpenShift. This provides cloud portability and enables monitoring, security and built-in scalability.

Security: Our DevSecOps process starts with our container platform. OpenShift affords HealthConcourse a container platform that provides enterprise Kubernetes and enforces that all containers run as a non-privileged user (not ROOT). Furthermore, each container is backed by an enterprise operating system that provides Red Hat's full support and attention to tracking and fixing critical security vulnerabilities (CVEs) in cooperation with the integration with National Institute of Standards and Technology (NIST) security checklists via OpenSCAP (Compliance as Code). HealthConcourse also leverages inherent network and infrastructure security controls provided by the cloud provider. HealthConcourse can handle personally identifiable information (PII) and personal health information (PHI) securely and be in compliance with all relative privacy and security policies.



HEALTHCONCOURSE DEMO ENVIRONMENT

For demonstration and validation purposes, we built a robust demo environment to mimic the complexity and heterogeneity of a real-world health IT computing environment. Figure 6 shows the additional solutions that make up the HealthConcourse demo environment.

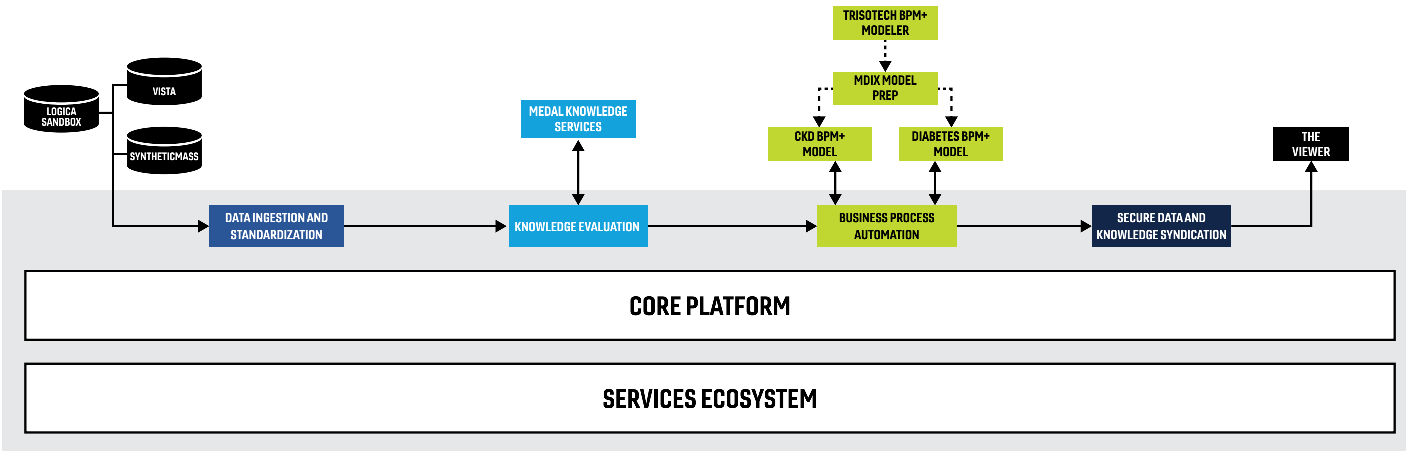


Figure 6: HealthConcourse demo environment

Systems of record: The HealthConcourse demo environment includes data from a variety of data sources. Specific EMR systems were chosen for integration because they are representative of those used by the Veterans Health Administration (VHA) and the Defense Health Agency (DHA). The Veterans Health Information Systems and Technology Architecture (VistA) system is the VHA's legacy EMR. The demo environment includes the open-source version of VistA. VistA is accessed through VistA-internal remote procedure calls that return a data feed that HealthConcourse converts to FHIR. In addition, our demo environment connects to the FHIR server in the Logica Sandbox and the SyntheticMass[9,10] database provided by The Mitre Corporation containing 1 million synthetic patients with complete, synthetized medical histories. We will also ingest data from demo-specific sources as opportunities arise. For instance, we implemented an instance of the Indian Health Service (IHS) Resource and Patient Management System (RPMS) for a demo at the Healthcare Information and Management Systems Society (HIMSS) Symposium.

Integrating with multiple EMRs allows us to recreate an important business challenge facing health care delivery today: a patient is seen by multiple providers in different organizations (e.g., VHA, DHA and private practitioners) each using different EMR products. Each EMR is essentially a boundary that stores and controls a portions of the patient data. Across these systems, a patient may be known by different identities with variance in demographic traits. HealthConcourse overcomes these challenges when the data is aggregated and standardized.

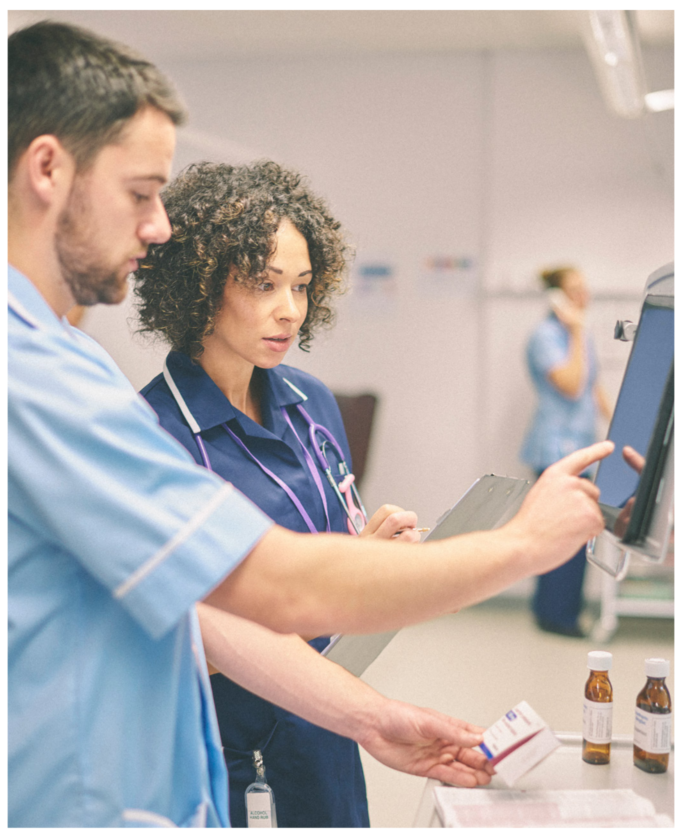
Knowledge services: The HealthConcourse knowledge manager can dynamically incorporate (register, discover and invoke) an unlimited number of knowledge services. This is hugely important for unlocking and democratizing health care knowledge commonly expressed as clinical decision support rules, business rules and/or analytics models historically buried inside proprietary, closed, black-box systems.

The demo environment includes knowledge services provided by the Medical Algorithms Company or Medal. Medal has amassed a collection of more than 20,000 medical calculators based on exhaustive literature review over many years. With emerging standards like CDS Hooks, these calculators can now be delivered as software able to integrate natively through our knowledge manager.

BPM+ Health models and supporting solutions: BPM+ Health is a very exciting development in health informatics. BPM+ Health models allow quality organizations and process models to represent the evidence-based best practice for practicing medicine in machine-readable, automatable clinical workflows and pathways. Instead of producing hundreds of pages of documentation explaining how to treat complex medical conditions, the knowledge baked into the clinical tasking, workflow, data requirements, decision points and end conclusions can be expressed in a BPM+ Health model. HealthConcourse is leading the way in implementing and automating BPM+ Health models as running software. In essence, a BPM+ Health model is an expression of four related things: tasking/activities, sequence, decisions and required data (inputs and outputs into tasks, activities and decisions). When a BPM+ Health model is integrated into HealthConcourse, all four of these items can be automated. Automation provides two important benefits: the ability to enforce the clinical best practices and knowledge expressed in the model without ambiguity or unnecessary or undesired variance; and, the ability to reduce the cognitive burden on clinicians by automating data identification (e.g., chart review to find specific information), data delivery (contextually relevant within the appropriate activity or decision in the model), decision support (through knowledge services) and task/activity sequencing. The demo environment includes the following solution components.

1. **Trisotech's BPM+ Health modeling tool:** Trisotech is a market leader in BPM+ Health tooling, particularly for creating robust, implementable models. For the demo environment, the Trisotech modeling tool was used to reproduce clinical pathways published by the VA and DoD for the treatment of chronic kidney disease and diabetes as computable, well-formed, BPM+ Health models. The modeling environment comes with a graphical editor with drag and drop ability to generate, link and organize activities, decisions, data objects and other relevant modeling widgets into a comprehensive workflow model. Data objects are bound as input and output parameters to modeled activities, conditions and decisions. Knowledge and decision support is also included and bound to the model either through decision modeling notation (DMN) expressions or through service calls to externally defined knowledge services. The models produced in the Trisotech tooling can be platform independent models, in that they do not bind the data and knowledge endpoints of any one particular target platform. While HealthConcourse provides FHIR endpoints for data and dynamically incorporated knowledge services for decision support, HealthConcourse may not be the only platform that runs a given model. To ensure platform portability and maximize uptake and reuse of BPM+ Health models, we used Trisotech to produce platform independent models to the greatest extent possible.
2. **A model operationalization environment powered by MDMI:** The model operationalization step is necessary to take a platform independent model and add the additional metadata and platform specific bindings that allow it to be integrated and run in a target implementation environment. This includes binding the data objects to actual data retrieval queries or APIs. For HealthConcourse, this means mapping the data objects to the HealthConcourse FHIR APIs. An intermediary called the MDMI Semantic Element Exchange Repository (SEER) provides platform independent data objects to the modeler in the Trisotech modeling environment and allows for platform specific binding to HealthConcourse FHIR APIs upon operationalization. By using the SEER, the data objects included in a platform independent model can be mapped to different platform specific implementations. Operationalization also includes mapping knowledge service calls expressed in the model to actual knowledge services running in HealthConcourse or other target platforms. Platform specific knowledge binding is accomplished when the BPM+ Health model is registered and run through the model discovery process within HealthConcourse. The model registration user interface (UI) provides the model registrant with a list of available knowledge services. When the registrant binds a decision node in the BPM+ Health model to an actual knowledge service, transformation logic can be added in the form of CQL scripts that transforms the model input and output parameters to the parameters used by the knowledge service. In the future, as CDS Hooks and other knowledge standards evolve, we anticipate that this manual step can be automated by unambiguously expressing the knowledge and decision support semantics required by the BPM+ Health model such that knowledge service binding can occur through an automated discovery and matching process.
3. **The BPM+ Health models registered and implemented as microservices:** After operationalization, BPM+ Health models are compiled as knowledge services with model metadata (including data and knowledge bindings) stored as FHIR resources in HealthConcourse. The BPM+ Health model microservices are automated when they are invoked by HealthConcourse-mediated processes or through the HealthConcourse viewer UI. Automation includes the delivery of data and knowledge endpoints as well as the execution of the workflow and orchestration of modeled activities

Systems of engagement: The demo environment comes with a clinical record viewer (referred to as the Viewer) user interface. The Viewer enables users to see the aggregated, standardized and enriched data for a select patient via the HealthConcourse FHIR APIs. The Viewer shows the core clinical and demographic data obtained from the systems of record as well as the inferred data derived from evaluating the source data with knowledge bases added into the data processing pipeline. The patient viewer includes a BPM+ Health enabled workflow visualizer. The BPM+ Health user interface depicts a clinical workflow as an interactive process model. This embedded UI highlights current activities and navigation across activities defined by the business process. It demonstrates the power of BPM+ Health models and event-driven architecture as many activities that currently require manual intervention and cognitive burden are now automated.



THE HEALTHCONCOURSE COALITION OF PARTNERS

Developing the many services and capabilities within HealthConcourse requires a coalition of partners with a variety of products aligned to the needs of the DIH reference architecture. Our partners and collaborators represent a thriving and evolving community of organizations that are collectively innovating within the Peraton development lab. Each of these companies offers cutting edge technologies in their own area of expertise. The ecosystem is greater than the sum of its parts when technologies from these companies are brought together with cohesion and orchestrated through a standards-based platform.

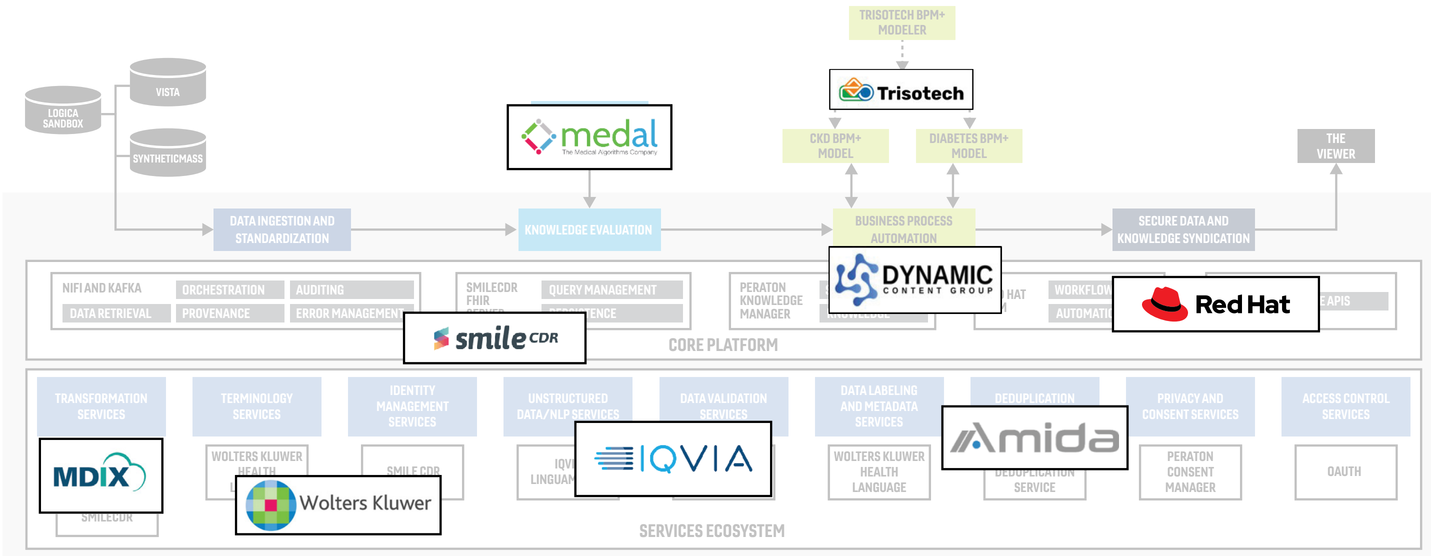


Figure 6: Our primary partners

DO THE CAN'T BE DONE.

Resources

1. Gartner, Digital Integration Hub: https://www.gartner.com/en/documents/3880263/innovation-insight-the-digital-integration-hub-turbochar. Accessed Feb 11, 2022.
2. Kin Lane. (2018). VA API Landscape Analysis and Roadmapping Project Report: https://apievangelist.com/2018/06/18/va-lighthouse-landscape-analysis-and-roadmapping-project-report/. Accessed June 22, 2018.
3. SMART-on-FHIR: https://smarthealthit.org/, accessed Feb. 12, 2018.
4. HSPC Roadmap 2018: https://healthservices.atlassian.net/wiki/spaces/SC/pages/104062021/Roadmap?preview=/104062021/149848065/2018%20HSPC%20Roadmap%20Ballot%20Candidate.pdf, accessed Feb. 19, 2018.
5. FHIR: https://www.hl7.org/fhir/, accessed Feb. 19, 2018.
6. CDS Hooks: https://cds-hooks.org/, accessed Feb. 19, 2018.
7. Omar Bouhaddou, PhD, Mike Davis, MS, Margaret Donahue, MD, Anthony Mallia, Stephania Griffin, JD, Jennifer Teal, RHIA, and Jonathan Nebeker, MD. "Automated Detection of Privacy Sensitive Conditions in C-CDAs: Security Labeling Services at the Department of Veterans Affairs." AMIA; 2017.
8. HSSP: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3002132/, accessed Feb. 11, 2022.
9. Tsung-Ting Kuo, Hyeon-Eui Kim and Lucila Ohno-Machado. "Blockchain distributed ledger technologies for biomedical and health care applications". JAMIA: Volume 24, Issue 6, 1 November 2017, Pages 1211–1220; 2017.
10. Jason Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher and Scott McLachlan. "Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record." AMIA; 2017.

ABOUT PERATON

Peraton drives missions of consequence spanning the globe and extending to the farthest reaches of the galaxy. As the world's leading mission capability integrator and transformative enterprise IT provider, we deliver trusted and highly differentiated national security solutions and technologies that keep people safe and secure. Peraton serves as a valued partner to essential government agencies across the intelligence, space, cyber, defense, citizen security, health, and state and local markets. Every day, our employees do the can't be done, solving the most daunting challenges facing our customers.

Scan to learn more at
https://www.peraton.com/markets/health/