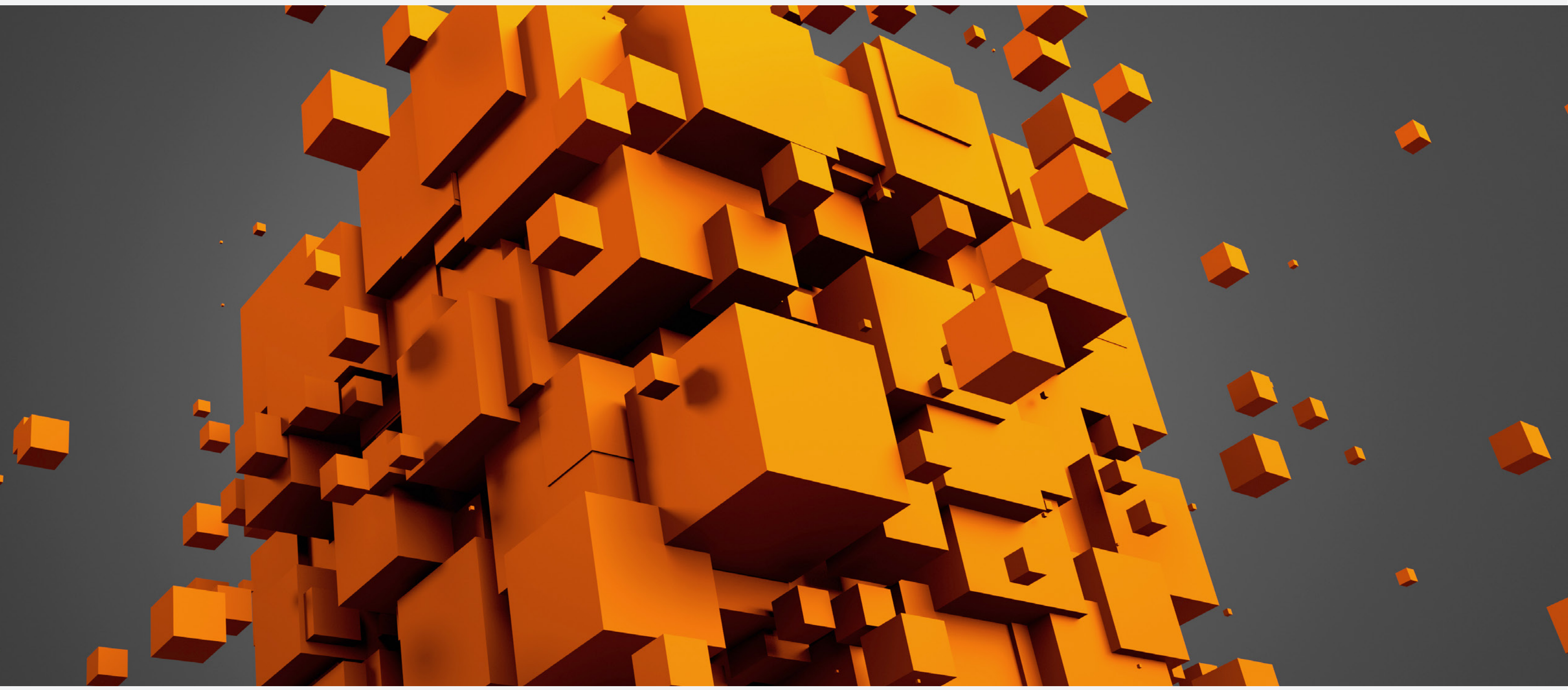
A Roadmap for Achieving a Largely Automated, AI and Analytics Assisted Learning Health System

An adjunct or starting point to today's VNA-based Enterprise Imaging Platform



By Michael J. Gray - Principal, Gray Consulting / November 2020

Sponsored by an unrestricted grant from Vital Images

VITAL®

Canon
CANON GROUP

Table of Contents

Executive Summary 3
Introduction 3
Terminology 5
Technology Requirements 8
Problems and Solutions 11
Conclusion and Recommendations to the Reader.. 15

Executive Summary

This paper will focus on the infrastructure of both software and hardware required to support the collection, normalization, and contextualization of all forms of clinical content. Collectively discrete content and imaging data is subsequently acted upon by Artificial Intelligence algorithms in the creation of predictive/suggestive and prescriptive analytics which can be used by clinicians to help reduce turnaround times and help improve patient care through better patient outcomes. The necessary infrastructure can be thought of as a next generation of Vendor Neutral Archive (VNA), or a standalone adjunct to the existing VNA. This new environment, referred to by some software designers as VNAi Services, is the go between inside human/clinical driven workflows vs machine/data driven workflows. In combination with all of the other application components of an enterprise imaging platform, workflow, analytics, visualization, this is the necessary "next step" for a healthcare organization that seeks to meet the Office of the National Coordinator for Health Information Technology's Interoperability Roadmap for 2025 and thereby achieve a Learning Health System.[1]

Introduction

In 2014, I wrote a white paper that addresses the need for what was then referred to as a "Third Generation" Enterprise Workflow application. In that paper, I discussed the need for a PACS workflow application that would automatically search the Electronic Medical Record (EMR) prior to the interpretation of a new imaging study for the patient's clinical information, automatically identify that information that is clinically relevant to the imaging study about to be interpreted, and present a summary of this relevant information on the PACS navigation screen, so the radiologist could use this additional "contextual" information to render a more complete and accurate interpretation of the imaging study. Most of the installed department PACS at that point in time, could not automatically launch the EMR in patient context, much less identify and contextualize clinical information pertinent to the imaging study.

This automated workflow process described over five years ago would eliminate the radiologist's time and effort required to proactively search the EMR for the information, but not eliminate the effort required to contextualize the EMR data and interpret it in the context of the imaging study and vice versa. The workflow concept promoted over 5 years ago is an early example of data mining of the simplest degree, and it was not conceived of as an entirely automated process. Even though a summary of the relevant EMR data would automatically be delivered to a screen, the physician still had to invest the time to review that data and apply it in the context of the imaging study while dictating the interpretation. Today we not only have to think about accessing all of the patient's clinical data that is relevant to an episode of care, no matter where it is created or where it is being managed, but also make the mining and conceptualization process almost entirely automatic and pro-active. This new workflow concept is an example of machine learning. It is the cornerstone of a Learning Health System, that makes a complete set of contextualized clinical information accessible to all of the caregivers regardless of location or context, thousands of potential users throughout the day, and not just those physicians interpreting imaging exams.

Today, healthcare is facing a new set of challenges, as there needs to be a wholesale change in the way care is delivered. Globally, healthcare is migrating from traditional care methods to methods that have a more consumer-centric approach. These new methods will need to support continuous population management and be more virtual in care provision. These new methods will require the seamless integration of data silos into a more expansive virtual data system that can deliver contextualized information and provide knowledge and insight to care delivery of the patient.

These changes are a shift in not only the way healthcare is to be delivered, they represent a change in the healthcare business model as well. Healthcare must shift from provider-driven to consumer-driven, from face-to-face to virtual, from encounter-based to episodic and continuous. In addition, healthcare will be shifting from procedure-based reimbursement to outcome/value-based reimbursement, healthcare must shift from diagnosis and disease treatment to health management and disease prevention.

Because the patient story is constantly evolving, all the patient's clinical data must be available for connection, interpretation and analysis. Tomorrow's healthcare organization needs a process that brings context and clarity to clinical data, in real time, to help all of the physicians associated with the patient's episode of care. Not just the radiologist interpreting the patient's imaging studies but an application that gives complete data perception allowing a view of the patient's health portfolio in real time.

[1] https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf

Consider now the technical implications required to achieve this shift. The data management systems now in place will need to shift from the relatively simple data-to-disk persistence model to a data interoperability perception model. Instead of a system that simply accepts data, corrects it and stores it to disk, providing centralization without interoperability, the new system will need to connect to, collaborate and orchestrate with all of those external systems where the rest of the clinical data is still stored and then contextualize that data. Instead of a system that is driven by humans, we need to shift to a system that is machine driven and capable of capturing, manipulating and delivering data, using smart automation to produce clean data.

Instead of a system that simply processes pre-determined functions and limits the use of data to routine dashboarding, we need to shift to a system that supports limitless data potential, a system that understands the data, challenges apparent results and, if necessary, changes the conclusions in an autonomous or semi-autonomous way.

This shift in technology would allow us to move from data that is organized and stored in a systematic, non-contextualized way according to predefined rules, patterns and commands in data silos to a technology that produces knowledge, driven by analytics, allowing humans first (and then machines) to drill deep into the data, find the outlier and challenge the status quo in real time. Such unbound, unrestrained databases would offer limitless potential to Real-Time Health Systems that would present the diagnostic picture with such clarity that it would substantially help to improve care and clinical outcomes.

This proposed shift in the way healthcare is to be delivered and the corresponding shift in the underlying data management technology required to achieve this shift is largely due to the US government's efforts to support interoperable, private and secure nationwide health information technology (health IT) systems. As defined on the HealthIT.gov web site, "the Office of the National Coordinator for Health Information Technology (ONC) is at the forefront of the administration's health IT efforts and is a resource to the entire health system to support the adoption of health information technology and the promotion of nationwide health information exchange to improve health care". In short, ONC has developed a Shared Nationwide Interoperability Roadmap that defines interoperability as the foundation of a Learning Health System. The major milestones in the proposed roadmap are:

2017 – Achieve the ability to send, receive, find and use priority data elements to improve health and healthcare quality.

2020 – To expand interoperable health IT and users to improve health and lower costs.

2024 – Achieve nationwide interoperability to enable a Learning Health System.

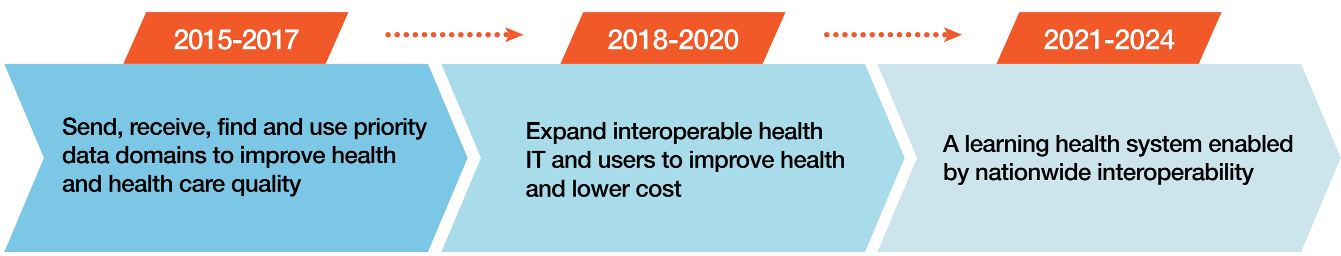


Figure 1. The ONC's Shared Nationwide Interoperability Roadmap.[2]

This should sound familiar—very similar to the government motivation encapsulated in the Meaningful Use Act that resulted in the deployment of Electronic Medical Record systems in the last decade. Like the Meaningful Use guidelines, the desired results are once again what is actually being emphasized, not the actual technology that will be required to achieve those results. To understand the required technology shift, it is necessary to start the learning process.

[2] https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf

Terminology

Readers may find the subject matter in this paper somewhat complicated (as I did for quite some time), and you will probably be confused by some of the new vocabulary.

Some key words already used in the Introduction are immediately understood, providing the reader with a base point from which to leap into whatever subtleties of the subject lay ahead. A Picture Archiving and Communication System or "PACS" is one such word, or acronym, that needs little introduction. The same can be said for the Electronic Medical Record (EMR) system, the Vendor Neutral Archive (VNA), a PACS Worklist and Workflow applications, and the Universal Clinical Viewer. Many other words that will be used in this paper and the concepts that they are associated with are going to require a little education, as an understanding of their meanings is somewhat dependent on the reader's experience.

Therefore, I think it is important to provide the following vocabulary list that will define the new (advanced) terms and concepts used in this paper:

Learning Health Systems (LHS)

Learning Health Systems are healthcare systems in which knowledge generation processes are embedded in daily practice to produce continual improvement in care. LHS entails a clinical lifecycle. Patient data is collected, it is amalgamated across multiple patients and a problem is defined. These are activities largely driven by healthcare professionals. With the support of technology, an analysis is performed, which returns evidence, from which knowledge is generated, which leads to changed clinical practice, and thus to new patient data being collected.[3]

Office of the National Coordinator for Health Information Technology (ONC)

ONC is the principal federal entity charged with coordination of nationwide efforts to implement and use the most advanced health information technology and the electronic exchange of health information. The position of National Coordinator was created in 2004, through an Executive Order, and legislatively mandated in the Health Information Technology for Economic and Clinical Health Act (HITECH Act) of 2009.[4]

Enterprise Content Management (ECM) System

The ECM system is a system solution designed to manage an organization's documents. Unstructured information—including Word documents, Excel spreadsheets, PDFs and scanned images—are stored and made accessible through access to the EMR to the right people at the right time. An ECM takes over the management of unstructured data (not DICOM study data) that might otherwise be managed less efficiently by an EMR.

XML File

An XML file is an XML (Extensible Markup Language) data file. It is formatted much like an HTML document but uses custom tags to define objects and the data within each object. XML files can be thought of as a text-based database.

JSON

JavaScript Object Notation, an open standard file format or data interchange format that uses human-readable text to transmit data objects consisting of attribute–value pairs and array data types.

FHIR

FHIR is an acronym for Fast Healthcare Interoperability Resources. It's a more granular way to exchange data without the rigid workflow of traditional HL7.

Syntactical Persistence

This refers to the process of taking data in and putting it on disc. It's the traditional way of thinking about healthcare data.

Semantical Perception

This refers to the process of taking data from any source, whether hospital-generated or patient-generated, and delivering it to the user, whether the user is a human being or even a machine. It's also about delivering it in the unique way that the user needs to receive it, so he or she can work efficiently and provide optimal patient care.

[3] https://en.wikipedia.org/wiki/Learning_health_systems [4] https://www.healthit.gov/topic/about-onc

Conventional VNA

The current generation of Vendor Neutral Archive is more of a smart “PACS extension” than a real “data sharing” solution. Conventional VNAs have limited data sharing capabilities when different PACS brands are connected. Data can be retrieved when users know what to query, but users have no real-time visibility on what data is available throughout the network.

Next generation VNA

The next-generation VNA (or Enterprise Imaging Repository) manages both DICOM and non-DICOM (HIS/EMPI, RIS and potentially ECM data). Most importantly, it will enable semantical perception and thereby support a Learning Health System, thus allowing message orchestration directly as HL7, DICOM, FHIR, RESTful integrations to act directly together. This VNAi Services allows the storage of all aggregated healthcare content as JSON in an opensource Elasticsearch database model to be used by all applications not just PACS and traditional VNAs or EMRs.

ODBC

Open Database Connectivity (ODBC) is a standard application programming interface (API) for accessing database management systems (DBMS). The designers of ODBC aimed to make it independent of database systems and operating systems.

Enterprise Service Bus (ESB)

An enterprise service bus (ESB) is a middleware tool used to distribute work among connected components of an application. ESBs are designed to provide a uniform means of moving work, offering applications the ability to connect to the bus and subscribe to messages based on simple structural and business policy rules. The next generation VNA will utilize a “service bus” that supports the aggregation of additional information that is contained in other systems that may not be able to be pragmatically retired. That information can be left in place or migrated into the next generation VNA on demand or in a systematically controlled manner. With an environment that is completely federated, the next generation VNA can participate in data sharing and image exchange with an EMPI-integrated service.

Artificial Intelligence (AI)

AI is the ability of a computer program or a machine to think and learn. In medicine, AI algorithms running in a hosting environment (typically standalone servers) mainly perform clinical diagnoses and suggest treatments. AI has the capability of detecting meaningful relationships in a data set and has been widely used in many clinical situations to diagnose, treat, and predict the results. All types of clinical data, stored both in the VNA and in other federated data repositories, must be pushed or pulled traditionally from multiple sources into the AI application.

Canonical Data Models (CDM)

CDM is a type of data model that aims to present data entities and relationships in the simplest possible form in order to integrate processes across various systems and databases. More often than not, the data exchanged across various systems rely on different languages, syntax, and protocols. The purpose of a CDM is to enable an enterprise to create and distribute a common definition of its entire data unit. This allows for smoother integration between systems, which can improve processes, and make data mining easier.

Perhaps it would be useful to provide more detail about the role that CDM will have to play in making the technology shift required to achieve the Learning Health System, and more specifically, how the next generation VNA may utilize CDM to simplify the integration of AI in the existing enterprise data management environments already deployed.

A Canonical Data Model (CDM) would enable the healthcare organization to create and distribute a common definition for all its clinical data; DICOM and non-DICOM image sets, XML, JSON and PDF objects, HL7 messages, etc. This would allow for easier exchange of all types of clinical data between the various repositories in the enterprise, PACS, VNA, ECM, EMR, HIS/RIS, EMPI, etc., facilitate the provision of contextual clinical data to the AI algorithms, and also make data mining for the analytics applications in general that much easier. Without a CDM and specifically data normalization per ONC, the accuracy of the AI algorithms will be skewed given inaccurate meta-data representing the information to be analyzed.

A CDM approach would also make it that much easier to exchange data between multiple applications plugged into the new generation VNA’s service bus (referred to in this paper as VNAi Services). By employing a CDM in the new generation VNA, the healthcare organization would be taking a canonical approach in which every application running on the VNA service bus translates its data into a single, common model that all the other applications on the bus also understand. This level of abstraction allows content to be used directly if necessary without a specific application to access the information (specifically that the AI applications will be able to understand).

A blog from BMC explains the process quite well.[5] "Importantly, a canonical data model is not a merge of all data models. Instead, it is a new way to model data that is different from the connected systems. This model must be able to contain and translate the other types of data. For instance, when one system needs to send data to another system, it first translates its data into the standard syntax (a canonical format or a common format) that are not the same syntax or protocol of the other system. When the second system receives data from the first system, it translates that canonical format into its own data format." All of this "translation" and "normalization" would be accomplished in the new generation VNA making the data much easier to share across its service bus.

The healthcare organizations that will achieve the Learning Health System being promoted by the ONC will be those that have successfully integrated their various data repositories, and the key to that integration will be the deployment of a CDM that will simplify the exchange of all clinical data objects across the new generation VNA's service bus. This breaks the business as usual paradigm of a siloed PACS deployment. This also should suggest to Healthcare Delivery Organizations (HDOs) they can no longer follow the traditional thought process of a consolidated PACS they believe does everything, when in effect it brings them nothing but a bigger problem to come. This problem is due to a lack of scalability and most certainly a lack of accessibility. To access data in a PACS you must go directly through the PACS software stack and cannot go around independently to the files stored on a disk drive. Therefore, it's not neutral, accessible, sharable, and most importantly normalized in any manner.

Thirty-plus years of PACS experience has taught us that it isn't easy to share data across different PACS, especially when the healthcare organization is very large and getting larger with every new merger and acquisition. Sharing and logical/physical centralization of image data between the various PACS used in the individual hospitals and imaging departments was the rational for developing the first generation VNA. These VNAs were/are designed for data persistence, and older technology VNAs often focus on just storing data or syntactical persistence of data. When we expand the scope of data exchange beyond image data to include all the clinical data in the patient's longitudinal medical record, we quickly recognize that there is rarely a single system architecture across the enterprise. In actuality, the IT department oversees a mixed bag of disparate systems, each with its own operating system, data formats, and communication protocols. These are the silos of disconnected information with no normalized form. No CDM exist.

As we look forward to achieving the Learning Health System, the successful healthcare organization will need to share data across all these systems. The problem is how difficult sharing data is when each system has different languages, requirements and protocols – and you consider the many iterations of one system talking to another. Perhaps the following diagram will help.

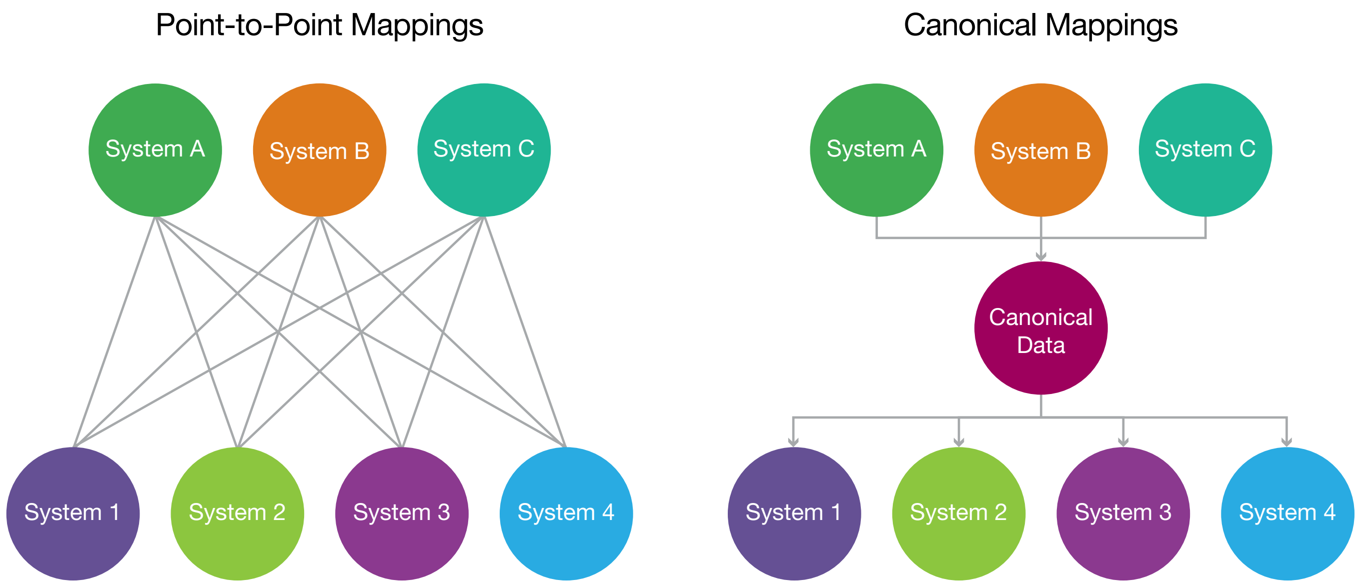


Figure 2. The comparative complexity of Point-to-Point mappings versus Canonical Mappings

[5] https://www.bmc.com/blogs/canonical-data-model/

Technology Requirements

While the broader goal of the healthcare organization is moving towards achieving the Learning Health System, a more immediate goal might be determining how best to deploy Artificial Intelligence algorithms in the existing enterprise imaging and clinical data management environment. While some believe that the concept of AI is immature, and many believe that the individual applications will be difficult to cost-justify without additional reimbursement, AI is real and it's coming, so it's time for healthcare organizations, specifically the planners in IT, to start thinking about how the organization is going to deploy it. As with many technologies, there may be multiple approaches to how the technology is developed and deployed, some better than others. In the case of AI, it is imperative that we determine the best approach to integrating the various AI applications into the organization's existing data management infrastructure.

From the very beginning, the medical imaging equipment market, from imaging modalities to PACS, was said to be "box-oriented" and siloed. Discussions of imaging technology used detailed block diagrams comprised of multiple boxes. If the scope of discussion is broadened to include all clinical data sources and their respective repositories, the box count in our block diagram of the enterprise data management infrastructure increases dramatically, and the interface connections represented in our block diagram overlap and get quite complicated. In keeping with this "box-oriented" practice, it is natural to think that AI will be yet another box or set of boxes in our block diagrams, and we will simply add the additional lines required to connect the various boxes. Also consider that those AI algorithms feed new findings, results, and recommendations back into the set of existing applications and AI algorithms. Consider as well that the exponential increase in future users (thousands of users) of these systems, both machine and human, by which the number of requests and services has not been completely considered.

Today, AI applications are frequently showcased as multiple independent applications hosted by individual physical or virtual servers, each supporting a myriad of physical interfaces to the various data repositories managing the clinical data that the algorithms will act upon. This model probably evolved because there are so many individual companies developing individual algorithms, and the standalone box model makes it easy to conceptualize how individual AI application from multiple vendors will be added to an existing enterprise infrastructure.

With this picture in mind, consider the following key technology-oriented questions related to the deployment of AI in the complex technology environment of a healthcare organization:

1. How will the AI servers access the relevant clinical data (DICOM, non-DICOM, HL7, XML, JSON, etc.)? Will this data be transferred to the AI servers, or will the AI servers query and retrieve the data?
   a. Semi-autonomous AI: Will the data transfer be manually executed when a physician decides there is a need to run the algorithm, or,
   b. Autonomous AI: Will the data be pro-actively and automatically transferred from repository to the AI servers by some intelligent operation executed in the repository itself or in the AI server?

2. Will those AI servers have to physically manage all of that clinical data for some specified period of time online, resulting in the costs associated with storing multiple copies of a large percentage of the organization's clinical data, or will the data be transferred anew each time someone wants to run the algorithm when new data on the patient is generated?

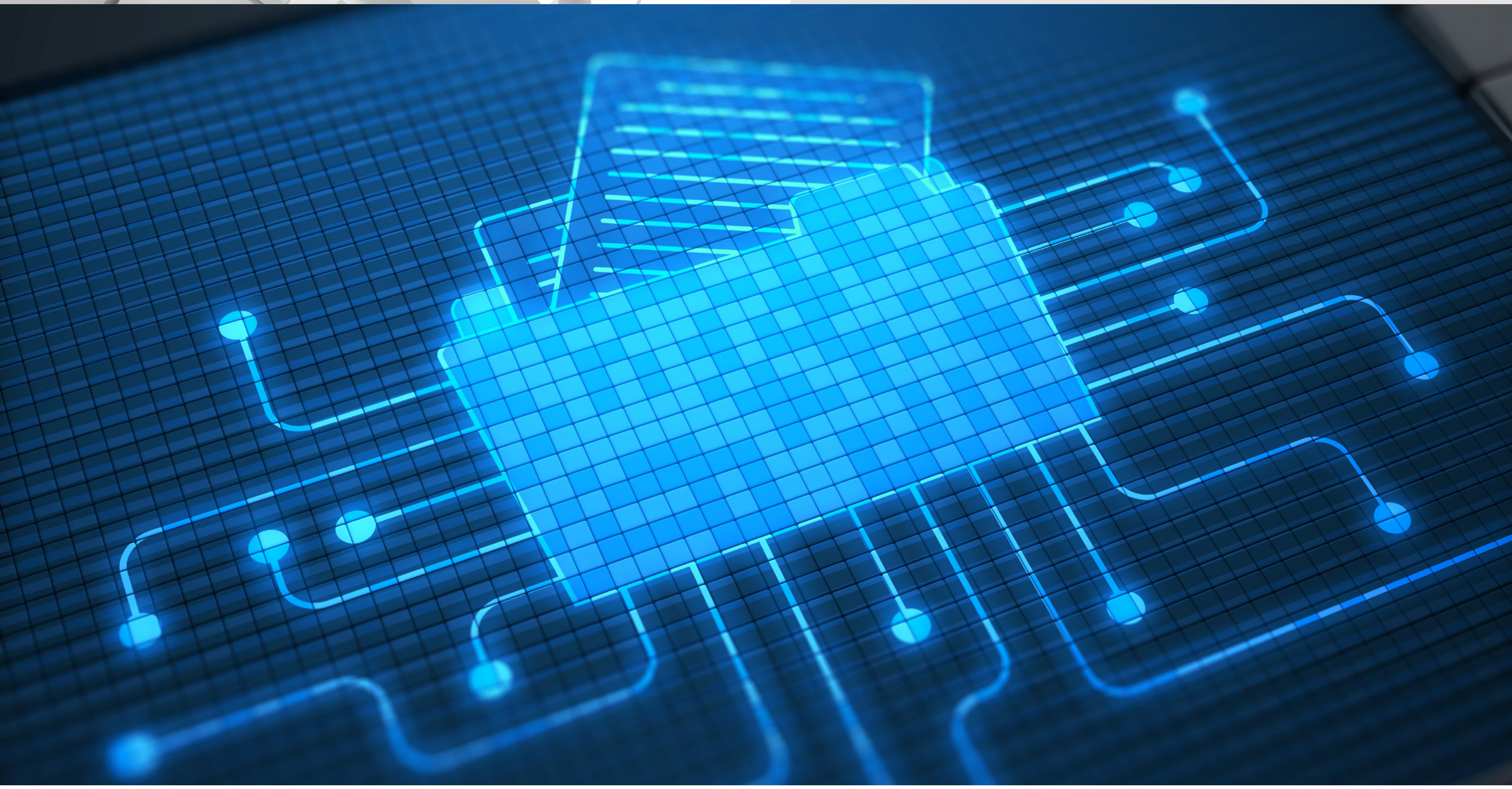
3. How will the person manually transferring the data, or the application that is automatically forwarding the data, determine what clinical data is relevant to and therefore required by the algorithm?

4. What are the best interfacing methodologies to support the data transfers from data repositories to AI servers?

5. How and in what system will the clinical data be translated into the format required by the AI algorithm? The value of the CDM comes to mind in this complex multi-system environment.
   a. Will it be necessary to add an application to each of the various data repositories to translate the data on the way out?
   b. Will the incoming data be translated in the "front end" of the AI server?

6. How will the results of the AI algorithm (findings, predictions, recommendations, etc.) be presented to the various users?
   a. Will users have to search for the results? Using what system, the EMR?
   b. Will results be automatically delivered by the AI servers to other designated systems like the PACS or EMR?
   c. How will users be notified of the existence/availability of the results?

There may be multiple approaches to how AI technology could be deployed in a given healthcare organization, some better than others, but the best answers to the above questions will help lead to the best (fastest, most efficient) solutions.



For the time being, let's set aside the issue of where the AI algorithms will reside. The **first major technology issue** must address the interfacing and data translation.

Let us start by thinking of AI in the context of the existing enterprise data management infrastructure, as needing some kind of "pre-processing" platform…a new box in our Infrastructure block diagram acting as a "front-end" to AI. This new platform will provide the means for accepting new data content formats (DICOM and non-DICOM image data, HL7 as well as XML clinical data objects, etc.) from all of the various sources and repositories, then pre-processing that data for use by the available AI algorithms and analytics applications. As for accepting new data, the front-end platform will use DICOM and XDS to access the image data and XML and JSON per FHIR responses and queries will be used to access the non-image data.

As for pre-processing the new data, the concepts of syntactical persistence and semantical perception play a major role. All of the data management systems currently in place in the healthcare enterprise (PACS, VNA, ECM, EMR, etc.) are simply taking data in and writing it to disk. This is syntactical persistence. It's the traditional way of thinking about healthcare data. In order to achieve a Learning Health System, we need to switch the way we think about this data from syntactical persistence to semantical perception.

Therefore, this new front-end platform has to pre-process the various forms of clinical data so that no matter the source or data repository where it is being managed, whether hospital-generated or patient-generated, it can subsequently be delivered to each of the various AI algorithms in the unique way that the algorithm needs to receive it. Once the AI algorithm has produced its results (findings, predictions, recommendations), the same platform will support certain back-end applications that process those results into a format that can be delivered to the intended human user, in the unique way that the user needs to receive it, so he or she can work efficiently and provide optimal patient care. At the core of these new front-end and back-end applications is the strategy of using a CDM to effectively translate the data back and forth between formats used by the various applications running in the healthcare environment.

The **second major technology issue** is deciding what will be the most efficient way to access, move, and store all of this clinical data that will be used by the AI algorithms. Will the new data be transferred manually when a physician decides there is a need to run the algorithm, or will the data be pro-actively and automatically transferred by the new pre-processing platform introduced above running some new application that somehow decides what data is pertinent to the AI algorithm? Where will all this new data be managed for whatever period of time it needs to be accessible? Will all of this new data be managed on storage associated with the AI applications, storage associated with the pre-processing platform described above, in the VNA?

One somewhat attractive solution due to its simplicity is to simply route all new clinical data, image data and non-image data, to the pre-processing platform described above. As data is received by the pre-processing platform, a set of rules is used to determine the pertinence of the data, and semantical perception is used to provide context and meaning to the data. A CDM is used to standardize and contextualize the data, so it is ready to be consumed by the AI algorithms. The context of the data is then provided to the AI algorithms. Pertinent data is then routed to wherever the algorithms are running. The original raw data then continues on to the appropriate data repository; VNA, ECM, EMR, etc. and is stored there as is the case today. In this sense, the pre-processing platform becomes a "front-end" to the existing data repositories in the enterprise. No new interfaces are needed for these repositories. Existing DICOM, XDS, HL7, XML, FHIR interfaces are already in place. No new medium to long-range storage is needed on the servers hosting the AI algorithms.

The **third major technology issue** is making the results available to the users. In the pre-processing platform scenario described above, the results of the AI algorithms are then collected by one or more "back-end" type of applications potentially running in the same platform as the front-end application(s) that directs them to a new type of dashboard presented in any one of several types of display application (advanced visualization, diagnostic display or enterprise clinical viewer) so the results can be used to help improve patient care. This process goes beyond just interpreting an imaging study and creating a report.

I have intentionally chosen to use the concept of a new box in our block diagram of the future enterprise data management infrastructure, because it makes it easier to imagine the "front-end"/"back-end" pre-processing platform described above working with all of the existing data management and display solutions already in place in the healthcare organization and the multitude of new AI servers that are coming. But I also want to point out that it is also possible to imagine the functional applications of this "front-end"/"back-end" pre-processing platform and the individual AI algorithms as simply being additional "plug-ins" to the enterprise service bus of a new generation VNA called VNAi Services. This later approach would indeed be the most efficient method of automatically presenting new data to the AI algorithms, and delivering results to the various users.

> The healthcare organizations that will achieve the Learning Health System being promoted by the ONC will be those that have successfully integrated their various data repositories, and the key to that integration will be the deployment of a CDM that will simplify the exchange of all clinical data objects across the new generation VNA's service bus.

Problems and Solutions

Here is an overview of the major questions that focus on technical issues that need to be addressed by the “front-end”/“back-end” pre-processing platform that is required to efficiently incorporate AI in the enterprise data management infrastructure, infrastructure necessary to achieving a Learning Health System.

1. What is the appropriate platform for accessing, processing, presenting data to the AI and analytics applications, collecting and presenting the results? How will this new platform fit into the existing enterprise data management systems the healthcare organization already has (VNA, ECM, EMR)? Is this platform simply a new application added to an existing platform like the VNA, or is it a separate platform that interacts with the existing systems?

   Ideally this new platform, this adjunct to the current data management environment, would be a standalone suite of applications, data interfaces and supporting hardware that could work seamlessly with the existing systems. Bear in mind that this is also about expanding the functionality of analytics applications to include collection and analysis of data produced and/or managed by all those other systems besides the Enterprise Imaging systems (PACS and the VNA), effectively expanding data mining and analytics to the HIS, RIS, EMR and ECM.

2. How does the Healthcare Organization, specifically the larger Independent Delivery Network (IDN), that has already invested in VNA and EMR technology go about beginning the deployment of numerous AI algorithms, designed to process all forms of clinical information? What would be the first steps?

   This is basically figuring out where these new analytics tools and AI algorithms will be hosted. Where (in what system or subsystem) do these new tools and algorithms physically reside, in a PACS, in the EMR or in the VNA, or in a completely standalone platform that can be interfaced to a pre-existing system?

   This also includes figuring out a methodology for “listening” for new data, interfaces for accessing that data, a process for contextualizing and normalizing the data, and determining how these new processes and methodologies can be automated. Then how will those analytics and AI applications access the necessary clinical data? What system-to-system data interfaces will be required? If the data is to be manually moved to the algorithms, when will the data be moved and by whom?

   If we are going to automatically move the data to the algorithms, how do we do that? Assuming the current environment of multiple data sources and repositories (HIS/RIS, PACS, VNA, EMR, ECM, etc.), and a healthcare organization’s desire to leverage all of these existing systems, what data interface connections will be required between all the data sources and repositories and whatever platform is hosting the new analytics tools and the AI applications?

3. If the Learning Health System is not almost entirely automated, how practical is it to think that users will manually execute the process of running data through an AI application? Requiring already-busy physicians to perform additional computer work would probably be a tough sell. How can any measure of efficiency be accomplished if the required processes are largely manual, even if pro-active, assuming that they will involve potentially thousands of clinical users?

   Most system developers believe that it would be too time consuming and labor intensive to run the required front-end processes on-demand, manually forwarding individual data sets to standalone AI platforms.

   It is also recognized that you cannot run an automated version of the required processes on an EMR or even on an existing VNA, because those platforms cannot sustain the run rates required to accomplish those processes required by AI.

   What is needed is a dedicated, expanded, high performance service bus either built in the VNA or hosted in a separate “platform” that could be interfaced to the existing VNA. This high-performance service bus is referred to in this paper as VNAi Services.

To help determine the best answers to the questions above please consider the associated data flow associated with the AI and advanced analytics environment, described below.

Up to this point in the paper, we have talked about certain front-end processes that are required to acquire and process incoming data that will have to be routed to the AI applications, and rear-end processes that are required to support the reporting of the results of the AI applications to the downstream users via an existing or new dashboard or a new or existing display application. For the sake of simplicity, let’s assume a model where all types of new and related prior data are going to be routed through a new-generation VNA, or a suite of standalone applications that are an adjunct to the existing VNA.

The following is a proposed data pathway (data flow) into and out of a new generation VNA, or the standalone software applications acting as an adjunct to an existing VNA. The proposed data flow also covers the final destination of the AI results, an existing or advanced display dashboard integrated into the range of display applications (Advanced Visualization, Diagnostic or Enterprise Clinical Viewer) used by the various consumers of the AI results.

Step 1.

The first step in the data pathway is the **acquisition** of new data. The incoming data object types to include DICOM, HL7, XML, XDS, JSON, REST, FHIR (it looks like V4 of FHIR will allow DICOM access) as access. Using JSON stored in Elasticsearch is what will allow the access to the above data object types. As a reminder, JSON is the open standard file format or data interchange format that uses human-readable text to transmit data objects consisting of attribute–value pairs and array data types.

The physical data interfaces would include DICOM, HL7, XDS, RESTful, XML drop to Epic, and FHIR Patient service. The interfaces would be supported by a new generation VNA or by a standalone platform interfaced to the existing VNA.

Step 2.

The next step in the data pathway is **normalizing** the data, as in modifying the headers to a standard syntactical format and correcting the metadata. This process would be applicable to all of the various incoming data types. This process would take place in a software application running in the new generation VNA software suite or in a standalone application running in a separate server interfaced to the existing VNA's service bus.

Step 3.

The next step in the data pathway is **converting** the data into a format that is consumable by the various downstream applications, including the AI algorithms and the business analytics that represent the machine "user", as well as the worklists, dashboards and display applications that face the human "users". Clinical/human workflow is the norm today, but machine/data is the new "user" that will and is overwhelming these environments. So, it will be necessary to make the data conversion process as robust as possible.

For the image data, this is primarily a conversion of the metadata and/or headers associated with the image pixel data. However, some of the AI algorithms already in use or on the drawing board will actually modify the image pixel data (i.e. US, MRI, CT). For the non-image data (HL7, XML, XDS, etc.) being created by or being managed by the HIS/RIS, EMR, ECM, the conversion of the actual clinical data would only be from an accuracy standpoint. Any human input needs to be examined as humans are capable of making many mistakes.

Once again, JSON is the contextual format that the non-image data is converted to, making it consumable by the downstream "users". JSON is the native data format to Elasticsearch, and JSON is the native format of FHIR, which many expect to dominate by the end of 2020. Note that JSON is now accepted in DICOM as a report format.

This data conversion process like the "normalizing" process would take place in a software application running in the new generation VNA software suite or in a standalone application running in a separate server interfaced to the existing VNA's service bus. The conversion would occur when the DICOM, HL7 or JSON message arrives.

The data normalizing and converting processes would effectively create a CDM for all types of clinical data that will live within the new enterprise data management infrastructure.

Step 4.

The next step in the data pathway is **storing** whatever data will be managed by the VNA. Assuming image pixel data is not converted to any other format, storing the image data would be the typical DICOM store by the VNA. The exception to this is the Part 10 components that presently would probably be tar-balled and moved in a tier manager that would create a metadata file on disk. This is very similar to the way several current VNA's do it now. A new format is in the works by the DICOM standards committees. If there is any header/metadata conversion to the DICOM image data, the original incoming metadata is stored in the VNA along with each of the converted versions and both the original and the file revisions can be searched and accessed. All changes can be seen and most importantly can be searched.

This header metadata would be stored in a file alongside the study, similar to what is being proposed in the DICOM standards committee for a change to Part 10.

As for all of the non-image data, assuming that there is a conversion of both the header/metadata and the actual clinical data to the downstream consumable format, the VNA would store both the original and all converted versions of this data. Both logical and physical deletions would also be supported.

Step 5.

The next step in the data pathway is "auto-routing" the converted data to the downstream applications ("users"). This auto-routing would be completely automated based on the functional requirements of the downstream applications. For example, a stroke AI protocol would require an input of relevant current and prior image data and pieces of non-image clinical data (ie, series description, CT head). This auto-routing would be initiated by a rules engine, yet another software application running in the new generation VNA software suite or in a standalone application running in a separate server interfaced to the existing VNA's service bus. The rules engine's functionality would be cross checked by the AI algorithm. For example, if the downstream "user" is an AI algorithm for stroke, the rules engine would determine if the incoming study is a brain or a breast and execute the auto-routing accordingly. The rules engine would determine what internal interface methodologies would be used to accomplish this auto-routing.

Step 6.

The next step in the data pathway is the **execution** of the downstream application(s) and a subsequent **reporting** of the results (all-inclusive) by the AI or analytics applications to a **report manager**. The report manager is another new software application that could be considered a 'governor" or "AI workflow director." The report manager would run in the new generation VNA software suite or in a standalone application running in a separate server interfaced to the existing VNA's service bus.

Step 7.

The last step in the data pathway is the actual **delivery** of the AI and analytics results/report to the human users. There are a number of possibilities here, ranging from interfacing to existing reading worklists to interfacing to a new generation of advanced worklist residing in either new or existing display applications. In all cases, the objective is to notify users that AI and/or analytics results are available for a given patient/study.

For the radiologist, this report manager application could integrate with an existing radiology worklist, so the AI and/or analytics results would present themselves to the radiologist through buttons on the reading worklist. In this manner, the radiologist would be informed that an AI and/or analytics result was available for a specific patient whose new imaging study was about to be interpreted, thus removing the need for the radiologist to go searching for the result. Another possibility is the report manager application updating an advanced radiology worklist application (button activations, etc.) that notifies the radiologist that results are available for a study.

This updating process could simultaneously transfer clinical content to an existing or an advanced display dashboard integrated into an existing or new display application. The variety of display applications referred to here include advanced visualization, diagnostic display application used by the interpreting physicians, and the clinical display application(s) used by the referring physicians, the so-called clinical viewer accessible through the EMR. One possible report manager-to-display application messaging methodology could be Apache Kafka, an open-source stream-processing software platform developed by LinkedIn.

Figures 3 and 4 below summarize the two options described in the above data pathway: software plug-ins to the service bus of a new generation VNA or standalone software applications residing in a separate platform interfaced to the existing VNA service bus.

Applications supporting the AI / Advanced Analytics Environment as Components of a New Generation VNA

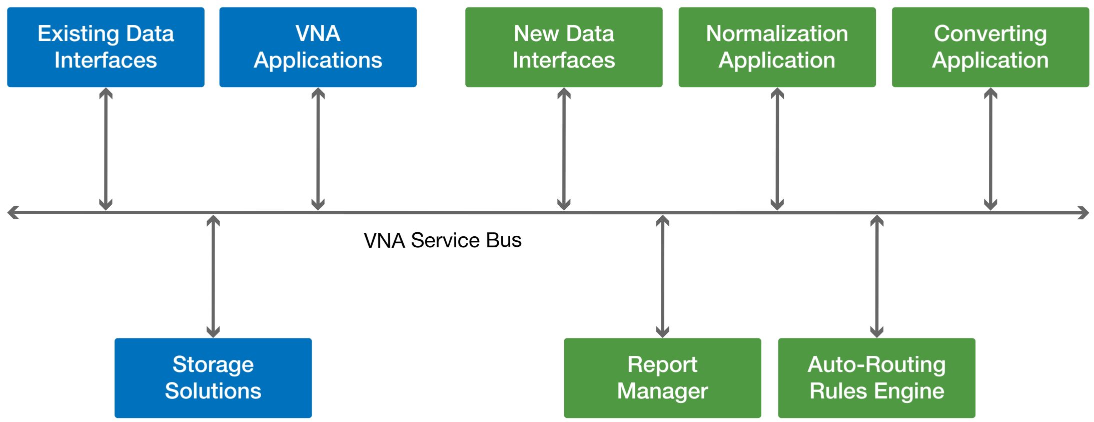


Figure 3. Applications supporting the AI / Advanced Analytics Environment as Components of a New Generation VNA

Applications supporting the AI / Advanced Analytics Environment as Components of a Standalone Platform

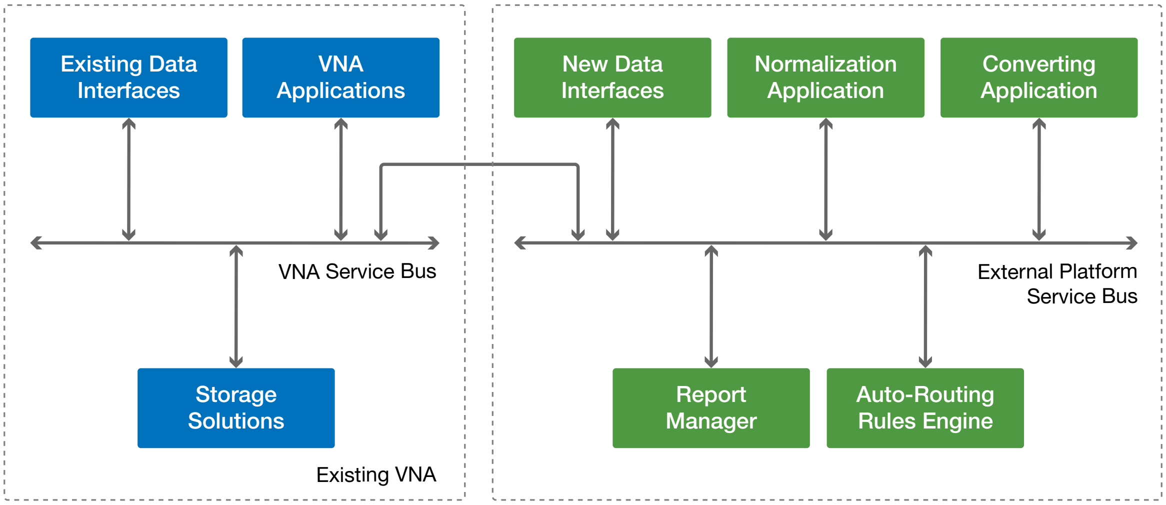


Figure 4. Applications supporting the AI / Advanced Analytics Environment as Components of a Standalone Platform

Conclusion and Recommendations to the Reader

First of all, it's time to start the education process. Today's healthcare organization's leadership, especially the thought leaders in the IT department, need to fully understand the changes that are coming in healthcare, including the initiatives of the ONC, and the concept of a Learning Health System. Today, we not only have to think about accessing all of the patient's clinical data that is relevant to an episode of care, no matter where it is created or where it is being managed, but also make the mining and conceptualization process almost entirely automatic and proactive. This new workflow concept is an example of machine learning, and it is the cornerstone of a Learning Health System.

Tomorrow's healthcare organization needs a process that brings context and clarity to clinical data, **in real time**, to help all of the physicians associated with the patient's episode of care. The data management systems now in place will need to shift from the relatively simple data-to-disk persistence model to a data interoperability perception model. IT leadership must now begin to consider the technical implications required to achieve this shift and to understand the required technology shift, therefore, it is necessary to start the learning process. Some of the key terms and concepts used in this paper include syntactical persistence, semantical perception, ESB, and the CDMs.

While the broader goal of the healthcare organization is moving towards achieving the Learning Health System, a more immediate goal might be determining how best to deploy Artificial Intelligence algorithms in the existing enterprise imaging and clinical data management environment, the environment comprised of the VNA, the EMR, the ECM and the myriad collection of diagnostic and clinical display applications. AI is real and it's coming, so it's time for the healthcare organization, specifically the planners in IT, to start thinking about how the organization is going to deploy it.

Secondly, as part of this education process, it is necessary for IT leadership to fully understand the infrastructure challenges and the potential solutions to those challenges in order to successfully deploy Artificial Intelligence applications. IT leadership must consider:

1. How will the AI servers access the relevant clinical data? How much of this process will be automated?
2. Will those AI servers have to physically manage all of that clinical data?
3. How will the person or the application that is forwarding the data, determine what clinical data is relevant?
4. What are the best interfacing methodologies?
5. How and in what system will the clinical data be translated into the format required by the AI algorithm?
6. How will the results of the AI algorithm (findings, predictions, recommendations, etc.) be presented to the various users? Will this presentation be automated, or will the user have to go looking for the results?

The above challenges require practical solutions to the three major technology issues discussed in this paper. The **first major technology issue** is addressing the interfacing and data translation requirements of the AI environment. The **second major technology** issue is deciding what will be the most efficient way to access, move, and store all of this clinical data that will be used by the AI algorithms. The **third major technology issue** is figuring out how to make the results of the AI and analytics applications **automatically** available to the many different users of the patient's clinical information.

The healthcare organization that has already deployed a VNA will most likely be unwilling to replace it with a new generation VNA that is equipped with those VNAi Services discussed in this paper. This organization's IT department will surely want to look for a separate technical solution that could plug into the existing VNA and therefore enable that existing VNA to interact with the AI environment. Such a solution was described in this paper as a "front-end"/"back-end" pre-processing platform, sort of a standalone adjunct to the existing VNA.

So, where to start? My recommended first step is to begin by researching the market for available technology solutions that address one or more (if not all) of the issues addressed in this paper that are associated with the successful deployment of an AI environment. If the organization is to have any chance of achieving the Learning Health System by the 2024 date recommended by the ONC, the organization is going to have to begin installing (as soon as possible) enough of this "front end" to begin accessing data and then processing it for either an advanced analytics application, or for the artificial intelligence application(s) that will plug into this front end.

My recommended second step is to install as soon as possible the "back-end" platform, the dashboard application that organizes the results, with the appropriate interfaces to existing diagnostic and clinical display applications that will allow display of the new dashboard information.

Achieving nationwide interoperability to enable a Learning Health System by 2024 is an ambitious program, and if IT leadership hasn't already started to think about the new AI environment, and the many complications presented by that environment, the organization has very little chance of successfully deploying an AI environment, much less achieving the ONC goal.

VITAL®

Canon
CANON GROUP

Vital Images, Inc. | 850 Opus Parkway, Suite 300 | Minnetonka, MN 55343 | USA | +1 866.433.4624
Vital Images Europe | Wiltonstraat 41 | 3905 KW Veenendaal | The Netherlands | +31 31.872.8900

M-06570B