Architecting and Building the Future of Healthcare Informatics: Cloud, Containers, Big Data and *CHIPS*

Rudolph Pienaar, Jorge Bernal,
Nicolas Rannou and P. Ellen Grant
Fetal-Neonatal Neuroimaging and Developmental Science Center
Boston Children's Hospital
Boston, MA 02115
Email: rudolph.pienaar@childrens.harvard.edu

Daniel Hähn
Department of Computer Science
Harvard University
Email: haehn@seas.harvard.edu

Ata Turk and Orran Krieger
Department of Computer Science
Boston University
Boston, MA

***Abstract*—New trends in software engineering are reshaping the computing landscape – computation is increasingly portable, storage is increasingly elastic, and data accessibility is increasingly "always on" and "always available" to an exponentially increasing variety of applications and devices. While the effects of these trends in the larger "compute-verse" are profound, this paper will discuss and consider how these trends are affecting specifically healthcare informatics. Indeed, end users will experience this trend in applications that are web-centric and mobile-friendly. Such apps will be increasingly used as gateways to powerful backend services (such as analytics and deep learning), while offering local client-side specialization (rich, immersive visualizations and collaborations). The paper offers some perspectives and presents some unmet needs in medical informatics and seeks to provide a viewpoint into how the "next wave" of computing might present itself. In particular the paper presents a web-based medical image data and information management software platform called *CHIPS* (*C*loud *H*ealthcare *I*mage *P*rocessing *S*ervice). This cloud-based service uniquely provides an end-to-end service that can connect data from deep within a Hospital securely to the cloud and allow for powerful collaboration – both on medical image data but also on image processing pipelines, allow for complex processing and enable computational research, and provide a vision of decentralized, large-scale data analysis that can fuel Big Data on medical bioinformatics.**

Keywords: Web based neuroimaging, big data, applied containerization, telemedicine, cloud-storage

I. Introduction

Historically, the tides in the information processing "ocean" have ebbed and flowed between centralization and de-centralization – with current trends moving to decentralization as offered by the cloud and away from traditional client heavy or client-only processing [1], [2], [3]. A heavily client-centric approach, however, is still prevalent in healtchare informatics – and the balkanization of data into disconnected silos is a defining feature of the healthcare informatics landscape [4], [5], [6], [7], [8]. Data silos are rarely interconnected, often the product of different incompatible solutions provided by different vendors with specific and sometimes limited scope – as a result computation on medical data and informatics has been low in comparison with many other technical and scientific fields. The reasons for this are complex and also partially historical: a combination of technical considerations, legacy deployments, as well as regulatory issues. Arguably, some of the reasons also stem from the basic cultural reality of medicine which is intensely personal and based off a one-to-one relationship between a single provider (i.e. a single physician) and a patient. The scope is differential and not integrative – medical practice is designed to focus primarily on a single patient (i.e. a single data point) and differentiate symptomatically how this patient fits in a larger picture of health. Clinical practice conventionally is not one of integration, but differentiation.

Informatics in healthcare reflects this fundamental relationship. Data repositories are not integrated and data processing is geared to the single case. Hospital medical images are typically stored in a closed Picture Archive and Communication System (PACS) with restricted accessibility and limited search capability, as well as mutiple structural incompatibilities with new trends in data informatics [9]. Medical image processing occurs inside hospitals that lack cutting edge computational resources and deep computational expertise. The computational tools that do exist are difficult to install and maintain in a typical hospital environment. Finally, it is difficult to share medical images between institutions or with interested third parties who have deep computational expertise due to the lack of standardized protocols for anonymization approved by local ethics review boards. Therefore, people with the skills to develop new computational tools do not have access to the imaging data. Technologies that enable access to anonymized medical images, provide platforms for sharing and support complex image computations are desperately needed to fully realize the promise of Big Data science on medical image data.

However, computation and devices are accelerating in a trend to decentralizatoin and interconnectedness, with more than 18 billion (and growing) devices projected to be exchanging information of some sort by 2020 [10] (see Figure 1). The application of rich computational tools and Big Data techniques to medical data, an in particular medical *imaging* data has the potential to transform medicine. Imagine a

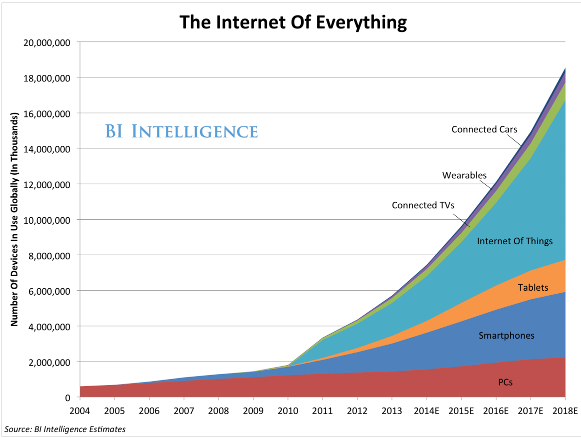

Fig. 1: Projected growth in interconnected devices.

world where millions of anonymized images, along with rich clinical data and massive computational resources are available to researchers exploring different diseases or developing new computational algorithms, available to clinicians from any device, and controlled by patients themselves. Conceivably, computational analytical techniques (such as machine learning, pattern classifiers, etc) mining this data could provide more accurate diagnoses, optimize treatments or identify early indicators of rare diseases. Although many researchers, have focused on developing such computational approaches, there is a pressing need to address fundamental platform issues.

This paper presents a discussion of some of these issues, and then offers a specific compute architecture to help realize the potential of this new trend in informatics, especially as applied to imaging healthcare. We present a solution that is "cloud-ready", based on using the web as interface and complex visualization, coupled with pervasive use of containerization technologies to distribute data and processing to mutiple networks. To our knowledge, no web-based platform currently exists that provides data *and* compute agnostic services (some services, such as CBRAIN [11] and LONI [12] provide conceptually similar approaches, but do not have deep connectivity to typical hospital database repositories), in particular collection, management, and real-time sharing of medical data, as well as access to pipelines that process that data. Our solution, *CHIPS* (**C**loud **H**ealthcare **I**mage **P**rocessing **S**ervice), is a novel web-based medical data storage and data processing workflow service that provides strict data security while also facilitating secure, real-time interactive collaboration over the Internet and internal Intranets.

CHIPS is able to seamlessly collect data from typical sources found in hospitals (such as Picture Archive and Communications Systems, PACS) and easily export to approved cloud storage. *CHIPS* not only manages data collection and organization, but it also provides a large (and expanding) library of pipelines to analyze imported data, and the containerized compute can execute in a large variety of remote resources. *CHIPS* provides for persistent record and management of activity in *feeds* as well as for powerful visualization of data. In particular, it makes use of the popular `XTK` toolkit which was also developed by our team at the Fetal-Neonatal Neuroimaging and Developmental Science Center, Boston Childrens Hospital[1] for the in-browser rendering and visualization of medical image data and can be freely downloaded from the web[2] [13].

II. Changing Trends

To support new paradigms of analytics and Big Data science in medical image research, technologies that enable access to medical images, provide platforms for image sharing, and provide access to infrastructures that support complex image computations are needed.

i. **The medical imaging future is demanding access to large anonymized datasets.** In order to develop novel image processing applications that utilize machine learning and Big Data analytics, access to sufficiently large pools of anonymized medical imaging data is required. However, access to image databases in most medical institutions is extremely complex due to both technical and policy reasons. It is also difficult to share such databases among institutions and with image processing experts due to the lack of standardized procedures for anonymization and sharing. Therefore, people with the skills to develop new computational solutions do not have access to the imaging data. *There is need for frameworks that integrate natively and securely with existing hospital services and provide easy access to imaging data making sharing, collaboration and development on imaging and relevant data intuitive and simple.*
ii. **Medical image processing infrastructures inside medical institutions lack cutting edge computational resources.** Novel image processing techniques that exploit cutting edge computational hardware systems (e.g. GPUs, accelerators) and/or large-scale parallelism are difficult to develop within medical institutions since clinical establishment do not have the expertise and economical incentive to set-up these special infrastructures as they become available. Modern technical solution to address these issues is to exploit cloud computing platforms that offer computational infrastructure on-demand. *There is need for frameworks that abstract-away and enable offloading of the computational requirements of image processing solutions to cloud computing platforms in order to ease development/testing/adoption of new approaches.*
iii. **Utilizing advanced healthcare solutions must be made simpler.** Effective usage of image processing approaches may require a significant amount of learning and effort on the side of the practitioners, which prevents adoption of these solutions in clinical settings. *There is need for intuitive, user-friendly and innovative interface designs to ease interaction with image processing applications by offering them as services and lowering their adoption barrier.*
iv. **Medical institutions shy away from partnering with large public cloud providers due to privacy considerations associated with sharing patient health care**

[1] http://fnndsc.babymri.org
[2] http://goxtk.com

data with these entities and fears of being locked into one commercial solution. *There is need for secure, non-commercial, and opensource frameworks that enables sharing/collaboration on medical imaging data over public cloud computing frameworks.*

v. **Solutions providing for Big Data should be generalizable across medical data modalities**. Infrastructure that consolidates all types of health care data and provides for processing on this data should not be specific to a single modality. *There is a need for solutions that allow for Big Data analytics irrespective of data type. Systems that process image data, for example, should be designed so that the same infrastructure can be used for other health care data such as genetics and other electronic health care records*[14].

The basic premise of this paper is that, to unleash medical imaging innovation, **there is need for an end-to-end framework that integrates securely with existing hospital image storage services, enables sharing and collaboration on image data, and interacts with cloud computing solutions to process this data on rich off-site computing platforms.**

INNOVATION

Medical image analysis research is facing challenges in utilizing the innovation offered by Big Data science. This is due to the difficulties and complexities associated with accessing medical imaging data, and whenever such access is available, obtaining access to cutting-edge computational frameworks to be able to develop Big Data science solutions for medical imaging. **Cloud Healthcare Image Processing Solution (*CHIPS*) is a fully integrated attempt to address these challenges by offering a framework that leverages cloud computing and social networking technologies to enable Big Data analytics research for medical image processing.**

i. ***CHIPS* is designed to drastically simplify the ability to access anonymized image data securely from within hospital data silos (such as PACS, providing more opportunity for Big Data science in imaging research.** *CHIPS* elevates data to first class citizen status and uses modern web-based approaches to collect, interact, and disseminate this data. *CHIPS* integrates natively with existing hospital services, in particular PACS databases, and provides services to query/retrieve image data. In addition, privacy and data-sharing are integrated in the very design of *CHIPS*. User access as well as data and analysis result access is only permitted under approprite IRB provisions. Where data is to be transmitted across compute networks, *CHIPS* will by default strip any patient identifying information from any image transmitted.

ii. ***CHIPS* is a pervasively container-based platform that standardizes the development of applications for medical image processing, allowing researchers to develop new algorithms, deploy/test these on powerful compute platforms such as public clouds and make them available to a broad community.** *CHIPS* encourages the deployment and sharing of new algorithms by providing a plug-and-play interface inspired by container-based solutions developed as part of cloud computing systems. This enables testing over hardware configurations that may not be available in the researcher's host environment. Any new application uploaded to *CHIPS* would in effect be available to all users of the system. By building a system that exposes a complete "end to end" workflow and that intelligently combines a multitude of existing tools, *CHIPS* will hopefully accelerate the usage of advanced research-based techniques within clinical workflows.

iii. ***CHIPS* is designed to provide a user interface for Radiology workflows that mimics social networking solutions. It transforms collaboration and interaction with medical imaging systems and enables real time collaboration/consultation through the interface using social media metaphors.** This aspect of *CHIPS* will significantly reduce the barrier of entry for medical image processing researchers and practitioners while offering/using new solutions. *CHIPS* front-end is web-based, and utilizes familiar social networking metaphors to ease usage and collaboration. Activity in the system is presented and logged persistently in rich "feeds", visual representations of user activity streams, that also contain the data operated on. Parameter choices in applications are automatically tracked. Experimental notes and image acquisition parameters can be added to feeds. Results and even feeds can be shared with other users enabling easy collaboration.

iv. *CHIPS* supports core application plugins implementing commonly used functionality in medical image analysis research. ***CHIPS* offers several neuroimaging plugins by default**, including tools such as FreeSurfer (for volumetric segmentation and brain surface reconstruction), Diffusion Toolkit (for white matter tractography), and 3DSlicer components

A number of technologies try to address certain segments of what *CHIPS* proposes, such as the GIFT-cloud [15], LORIS [16] and others [17], [18], [19] that ease presentation and sharing of imaging data by integrating with clinical systems, but do not provide mechanisms to link these systems with cloud solutions. Other frameworks such as Gadgetron [20] and SQUAREMR [21] try to harness the power of cloud computing systems for improving the performance of specific imaging applications such as MRI reconstruction, but do not tackle the core problems of data access or general computational infrastructure requirements. Pipelining solutions such as LONI [22] exist that are designed for planning, executing, monitoring and sharing scientific workflows, but are not directly compatible with hospital or cloud systems. Finally, large cloud providers cloud providers such as Amazon [23], IBM [24] and Microsoft [25] offer generic HIPAA compliant services that aim to attract medical applications to their platforms but do not offer open source, cross-cloud applicable solutions. *Currently, there are no solutions that service all the needs highlighted above in one single end-to-end solution. CHIPS* is the only system that is designed to offer a free and open, publicly accessible, cloud-based and completely integrated and managed experience for medical image data analysis that offers mechanisms for collecting data, consulting and collaborating with other researchers over the data, as well as processing, analyzing, visualizing and returning new data, all through intuitive and modern web interfaces.

III. Approach

Modern web browsers are becoming powerful platforms for advanced application development [26], [27]. New advances in core web application technologies such as the modern web browsers' universal support of ECMAScript 5 (and 6) [28], CSS3 and HTML5 APIs have made it much more feasible to implement powerful middle-ware platforms for data management and powerful graphical rendering, as well as real-time communication purely in client-side JavaScript [29], [30]. The last decade has seen a slow, but steady, shift to fully distributed solutions using web-standards [31], [32], [11], [33], closely tracked by expressiveness of the JavaScript programming language. Web-based solutions are especially appealing as they do not require the installation of any client-side software other than a standard web browser which enhances accessibility and usability.

Unrelated to rise of web-technologies, a new emerging trend is the rapid adoption of containerization technologies. These have enabled the concept of *compute* portability in a similar sense to *data* portability. Just as data can be moved from place to place, containerization allows for operations on that data to also be moved from place to place.

CHIPS is able to seamlessly collect data from typical sources found in hospitals (such as Picture Archive and Communications Systems, PACS) and easily export to approved cloud storage. *CHIPS* not only manages data collection and organization, but it also provides a large (and expanding) library of pipelines to analyze imported data, and the containerized compute can execute in a large variety of remote resources. *CHIPS* provides for persistent record and management of activity in *feeds* as well as for powerful visualization of data. In particular, it makes use of the popular `XTK` toolkit which was also developed by our team at the Fetal-Neonatal Neuroimaging and Developmental Science Center, Boston Childrens Hospital[3] for the in-browser rendering and visualization of medical image data and can be freely downloaded from the web[4] [13].

IV. Architectural Overview

A. Scope

The creation of *CHIPS* has been motivated by both clinical and research needs. On the clinical side, *CHIPS* was built to provide clinicians with easy access to large amounts of data (especially from PACS), to provide for powerful collaboration, and to allow for easy access to a library of analysis processes or pipelines. On the research side, *CHIPS* was designed to allow computational researchers to test and develop new algorithms for image processing across heterogeneous platforms, while allowing life science researchers to focus on their research protocols and data processing, without needing to spend time on the minutiae of performing data analysis.

[3]http://fnndsc.babymri.org

[4]http://goxtk.com

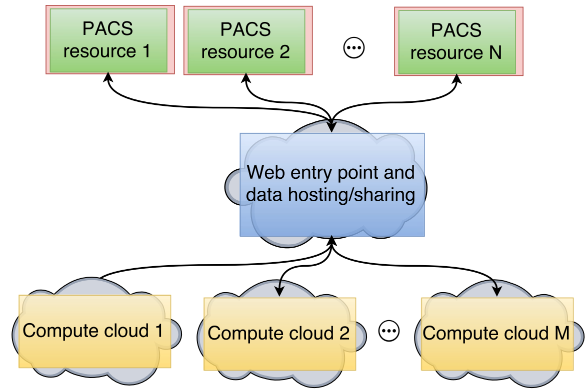

Fig. 2: *CHIPS* connects multiple input PACS sources to multiple "cloud" compute nodes.

The system design is highly distributed, as shown in Figure 2, which shows a *CHIPS* deployment connected to multiple input sources and multiple compute sources. Though the figure suggests a single, discrete central point, components of *CHIPS* do reside on each input (PACS) and compute location.

B. Distributed Component Design

Architecturally *CHIPS* is not a single monolithic system, but a distributed collection of interconnected components, including a front-end webserver and web-based UI; a core RESTful back-end central server that provides access to all data, feeds, users, etc; a DICOM/PACS interface; a set of independent RESTful microservices that handle inter-network data IO and also remote process management, and a core cloud-based computational platform that orchestrates offloading of image processing pipelines to some remote cloud-based compute – see Figure 3.

The top the red box of Figure 3 contains the *PACS node* and represents the Hospital image data repository. The second blue box, labeled *Web-entry point and data hosting node* contains the main *CHIPS* backend and is presented as being in a "cloud" (i.e. some resource that is accessible from the Internet). Finally, the bottom yellow box is shown on a separate "cloud" to emphasize that it is topologically distinct from the *Web-entry point*.

The logical relationships between data (represented as the rectangles with a tree structure) and compute elements denoted by the named hexagons is shown by either data connectors (thick blue arrows) or control connections (single line arrows). In the syntax of the diagram, the stylized cloud icon touching some of the boxes denotes that these compute elements are controlled by a REST API, while the sphere icon denotes web-access.

An remote compute is denoted by `plugin`, which is controlled by a `manage` component. In the most abstract sense, the `plugin` processes an input data structure, and outputs a transformed data structure (the two tree graphs as shown). File transfer between the data cloud and compute cloud is performed by the file `IO` handler component. A `query/retrieve` process in the data cloud connects to an

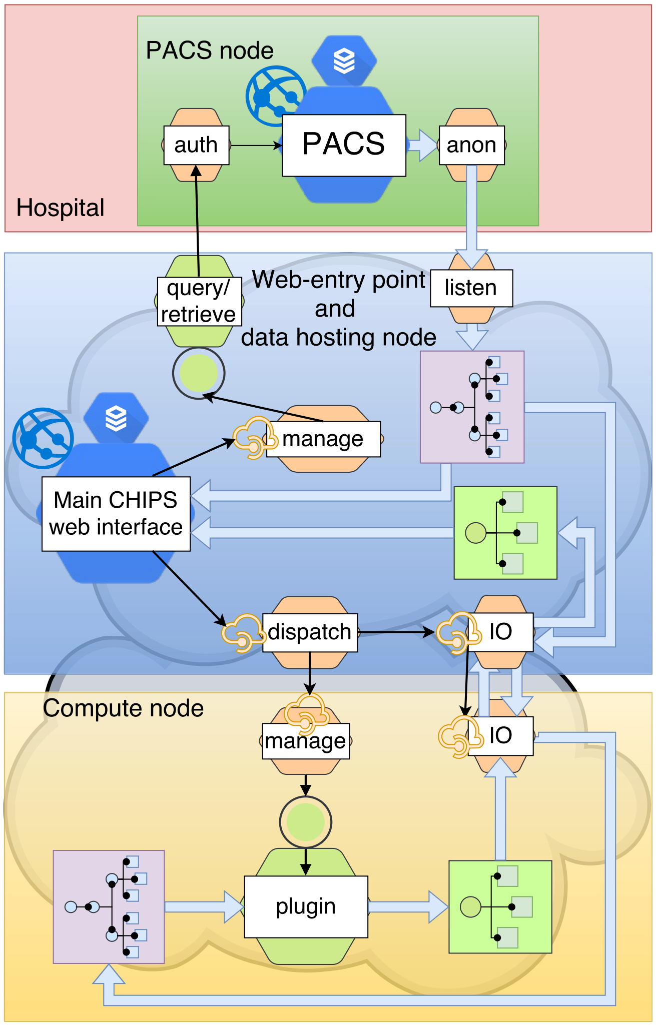

Fig. 3: The internal *CHIPS* logical architecture.

authentication process, `auth` in the Hospital network, while on-the-fly anonymization of DICOM images is handled by process anonymizer `anon`. Finally the `dispatcher` is a component that determines what compute node (or cloud) is best suited for the data analysis at hand. The circle icon attached to the `manage` and `plugin` icons implies the attached process and can provide real-time feedback information to other software agents about the controlled process via its own REST interface.

C. Pervasive containerization

CHIPS is designed as a distributed system, and the underlying components are containerized (currently using docker[5]). Docker is a rapidly maturing technology that is offered as a first-class citizen on many commercial cloud offerings such as Amazon Web Services, Microsoft Azure, Google Compute Engine and IBM Softlayer among others. It is designed to be an infrastructure level technology for managing and deploying software at scale [34]. Docker's main components are a command-line program, a background daemon, and a set of remote services that together simplify installing, running, publishing, upgrading and removing software. This is accomplished by using Linux containers which provide a sandboxed environment for running software applications and all of its dependencies. These containers are at the heart of the strong security features provided by Docker as they isolate a process from all computer resources except where explicitly allowed. Docker also includes several ways to package and distribute software through Docker images. A Docker image is a bundled snapshot of all the files that should be available to a program running inside a container. Many containers can be created from the same image but they do not share changes to their file system [35]. Software within docker containers see the same execution environment and initial state regardless of underlying hardware, operating system or computer state. *CHIPS* will use docker containers for all its plugins as its secure application distribution model.

In Figure 3, the *Main CHIPS web interface* and associated backend database is housed within a single container[6]. Input data and processed results are accessible in the hosting node and volume mapped as appropriate to this back end. Other components of *CHIPS* in the web-entry node are similarly containerized. This includes the `manage`[7] block, which is responsible for spawning processes on the underlying system. Not only does `manage` provide the means to start and stop processes, but it also tracks the execution state, termination state, and standard output/error streams of the process. The `manage` component has a REST interface through which clients can start/stop and query processes.

Also containerized is the `IO`[8] component that can transfer entire directory trees across network boundaries from one system to another as well as the `dispatch`[9] component that can orchestrate multiple processing jobs as handled by `manage`. The `plugin` container houses the particular compute to perform on a given set of data, and is spawned by the `manage` component under direction of the `dispatch`. Since the compute typically occurs on a separate system to the data hosting node, the `IO` containers perform the necessary transmit of data to this compute system, as well as the retrieve of resultant data back to the data node, allowing the web container to present (and visualize) results to the user.

A critical part of *CHIPS*'s architecture is the means by which plugins are shipped to and run on the remote compute resource. We propose two main mechanisms for this: (1) allow external parties to create their own docks which then are added to the backend system by *CHIPS* development team after vetting of the plugin's suitability; and (2) develop a sandboxed/automated docker build system for low-dependency plugins, where an external developer uploads the source code of their plugin using an "upload" feature of the web front end. This upload will capture sufficient information to allow the build/execution of the plugin on the remote resource, and will then create a simple dock for the plugin which in turn will be automatically added to the library of available docks. We propose to sandbox such auto-created plugins in an isolated filesystem and even on an isolated compute environment (such as a virtual machine) for a period of evaluation and testing of such auto-generated

[5]https://www.docker.com

[6]https://github.com/FNNDSC/ChRIS_ultron_backEnd

[7]https://github.com/FNNDSC/pman

[8]https://github.com/FNNDSC/pfioh

[9]https://github.com/FNNDSC/swarm

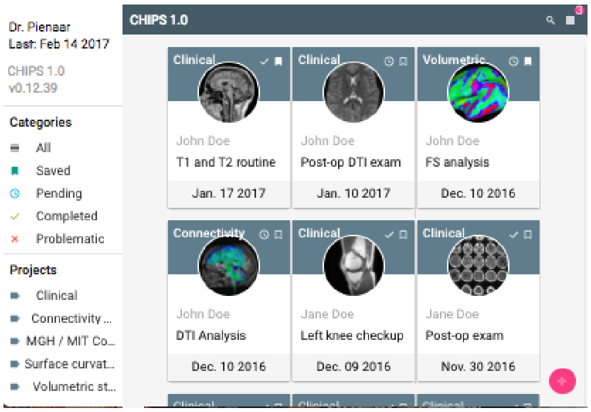

Fig. 4: *CHIPS* home page with a “cards” organization.

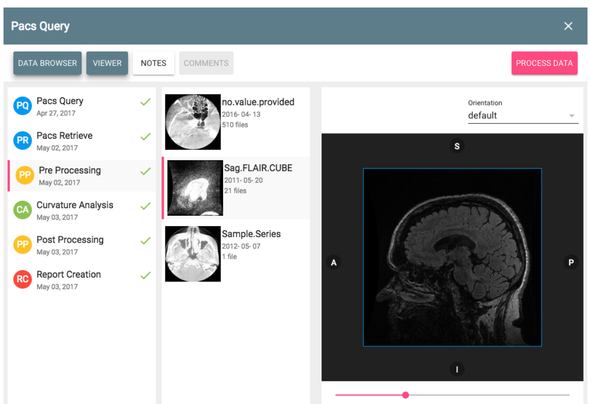

Fig. 5: Visualizing pulled and processed data.

plugins. Also important is the ability for *CHIPS* (via the `manage` container) to support automatically middleware functionalities such as service discovery, queueing, and routing once a cloud platform has been tied to *CHIPS*. By enforcing distribution and management of individual plugin applications in their own Docker images and containers, we open the path to leveraging existing scheduling and management tools aimed at solving those specific problems such as the Docker Swarm or OpenShift[10] and Kubernetes.

V. UI Considerations

Figure 4 shows the home page view on first logging into the system. Studies that have been “sent” to *CHIPS* appear in their own “cards” on the user’s home page with a small visualization of a represented image set of the study. Various control on this home page allow users to organize/tag “cards” in specific projects (or folders), remove cards, bookmark for easy access, etc. New cards can be generated by clicking on the ⊕ icon and choosing an activity (such as PACS Query/Retrieve), and any card can be seamlessly shared with other users of the system.

On selecting a given feed, the core image data in that feed is visualized in a rich, web-based viewer – see Figure 5.

[10] https://github.com/FNNDSC/openshiftmgr

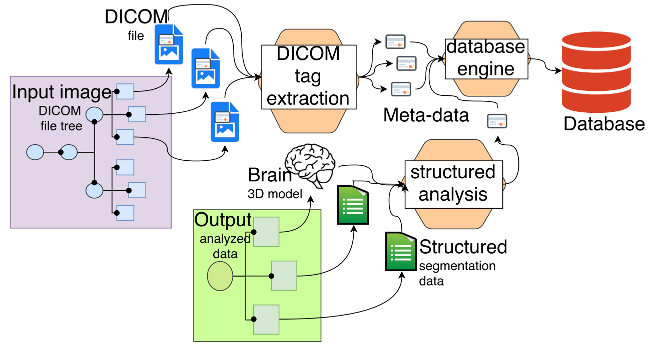

Fig. 6: Big data pre-processing.

Various tabs and elements of the feed view provide different perspectives on the data, and also provide the ability to annotate notes, or add comments. As in the feed view, a ⊕ icon is also present, and if selected, opens a ribbon of “plugins” (or “apps”) to run on the data contained in the feed. For example, certain plugins might perform a surface reconstruction of the brain surface with tissue segmentation (for example, a FreeSurfer plugin).

The interface semantics within a feed are straightforward: a user clicks on the feed and enters the top level data view. Once a plugin from the ⊕ is applied, the feed data is processed accordingly. When the plugin is completed, its output files are also organized in the feed in a logical tree view (accessible via the left “Data” tab) in a manner akin to an email thread. In this manner, the thread of execution from data → plugin → data is defined – in effect building a workflow.

Any image visualized can also be shared in real-time using collaboration features built into the viewer library and leveraging the Google Drive API and Google Realtime API [30].

VI. Big Data Infrastructure

An important component of *CHIPS* lies in creating a foundation suitable for future support of “data mining”. Recently, the term *Big Data* has come into common parlance, especially in the context of informatics [36], [37], [38]. Despite the term and the use of *Big*, the concept often refers to the use of predictive analytics and other advanced data analytics tools that extract meaning from sets of data and does not necessarily to the particular size of the data set.

In healthcare, big data analytics has impacted the field in very specific areas such as clinical risk intervention, waste and care variability reduction, and automated reporting. However, as a field, biomedical imaging has not especially benefited from big data approaches due to the unstructured nature of image data, complexity of results from analysis in terms of data formats (again usually unstructured), simple quality issues such as noise in image acquisitions, etc.

CHIPS constructs a framework to allow big data methods to be used in this image space. Consider that the incoming source data to *CHIPS* are DICOM images that by their nature contain a large amount of meta information, most of which is non PHI and will be left unchanged by the anonymization processes. Information about the scanning/imaging protocol, acquisition parameters, as well as certain non-PHI demographics such as patient sex and age can be meaningfully databased. Moreover, the application of an analysis pipeline to an image data-set can in turn result in large amounts of meaningful data that can be databased and associated with the incoming source data. For example, FreeSurfer, which is dockerized as a plugin in the *CHIPS* system produces volumetric segmentations and surface reconstructions on raw input MRI T1 weighted data [39], [40], [41].

In Figure 6 input raw DICOM (purple block) and output processed data from the DICOMs (green block) are shown. A `DICOM tag extraction` process removes the image meta data and associates this information with the particular image record. DICOM data is regularly formatted and easily extracted. Importantly, for the output data, and assuming the output data is a 3D surface reconstruction and tables of brain parcellation volume values, a `structured analysis` process regularizes all this information into meta data that will be added to the space of data pertaining to this image record. This processing will lay the ground work on which data analytics can explore and mine for relations between (for example) input acquisition parameters and pipeline output results, or simply mine across output results for hidden trends in data trajectories (for example volumetric changes with age or sex).

VII. Privacy Handling and Patient-centric trends

Medical data, like financial data, is subject to regulatory constraints governing its dissemination on computer networks. While for the most part medical data is usually generated, consumed, and stored within a single institution, it is important to consider the patient's position in data ownership [42], [43]. Complex issues such as regulatory forces, historical inertia, incompatible data formats, etc. all contribute to placing "lag" or barriers to the flow of data to distributed locations such as the cloud and the ability for distributed compute to perform analytics on data.

Nevertheless, it is reasonable to assume that the overwhelming trend of computing is towards portable data, and in healthcare this can lead to changes in the concept of ownership. While this might contribute to a delay in healthcare data systems to be more integrative and distributed, the movement of the computing industry and expectations of patients arguably make such an outcome very likely.

While in this paper concerned itself primarily with computing trends in general and technical considerations of anticipating these trends in healthcare informatics, and while privacy issues are complex and should not be underestimated, our position is that data will move fluidly across networks and compute and that the patient himself/herself will become a primary role-player in the portability and access to their own data. To this end, in *CHIPS* data privacy as understood by HIPAA is paramount, and addressed by specific modules that authenticate to, and anonymize and image data received from a PACS.

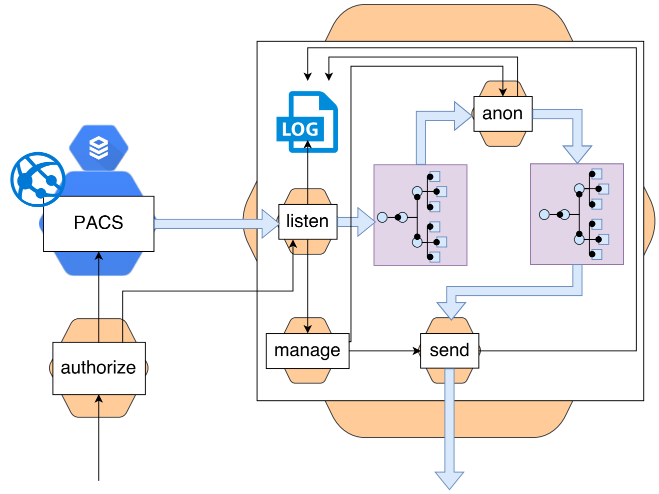

Fig. 7: *CHIPS* authorization to PACS and dynamic anonymization engine.

The anonymization component of Figure 3 is shown in more detail in Figure 7. In Figure 7, the receipt of DICOM data by a listener process triggers a cascade of events. When a user has been authenticated as being allowed to perform a PACS Q/R, a listener process, `listen` is informed by the `authorize` process of the user and the data that has been queried and subsequently tagged for retrieval. As data arrives at `listen`, it logs all incoming files as well as the original *CHIPS* user that is performing this Q/R. Data is parsed and packed out to the filesystem of the host computer which is receiving the PACS data.

Once unpacked, anonymization on the received data tree is performed by the `anon` process – note that multiple off-the shelf and opensource DICOM anonymization applications are available. On successful anonymization, the process manager, `manage`, will invoke a transmission process, `send`, that will read the anonymized files and transmit to the *CHIPS* cloud. Note that as shown in the Figure, all components of this pipeline log all activity.

VIII. Conclusion and Future Directions

With an expected exponential increase in connected devices, an accelerating trend to distributed-but-centralized computing across multiple environments, the informatics landscape in healthcare is at the cusp of fundamental changes. We anticipate that data and services on data in healthcare will migrate every increasingly to cloud environments, both for storage of data and also for group-based analytics (providing data for more accurate group-based models, deep learning and Big Data) as well as single patient specific processing (process a single patient image data for tumors, for example).

In this paper, we discussed some of these larger trends, as well as very specific needs in healthcare, and presented a system called *CHIPS* that is one possible platform that is designed for being future-ready. *CHIPS* is a distributed system that provides a single, cloud-based, access point to a large family of services. These include: (a) accessing medical image data securely from participating institutions with authenticated access and built-in anonymization of collected image data; (b) organizing collected data in a modern UI that allows for easy data management and sharing; (c) performing processing on images by dispatching data to remote clouds and controlling/managing remote execution on these resources; (d) powerful real-time collaboration on images using secure third party services (such as the Google RealTime API); and intuitively constructing medical image processing workflows. *CHIPS* is not only a medical data management system, but strives to improve the quality of healthcare by allowing clinical users the ability to easily perform value added processing and sharing of data and information. Current and future directions for *CHIPS* include facilitating the construction of big-data frameworks and allowing for users to simply construct experiments for data analytics and various machine learning pipelines.

REFERENCES

[1] L. Kappelman, E. McLean, V. Johnson, and R. Torres, "The 2015 sim it issues and trends study." *MIS Quarterly Executive*, vol. 15, no. 1, 2016.

[2] D. Peak and M. Azadmanesh, "Centralization/decentralization cycles in computing: Market evidence," *Information and Management*, vol. 31, no. 6, pp. 303 – 317, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378720697000025

[3] S. Wattal and A. Kumar, "Cloud computing-an emerging trend in information technology," in *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*. IEEE, 2014, pp. 168–173.

[4] A. Seth, "The future of healthcare informatics: it is not what you think," *Glob Adv Health Med*, vol. 1, no. 4, pp. 5–6, Sep 2012.

[5] R. Haux, "Health information systems past, present, future," *International Journal of Medical Informatics*, vol. 75, no. 34, pp. 268 – 281, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505605001590

[6] E. Wicklund. (2014) 'Silo' one of healthcare's biggest flaws. [Online]. Available: http://www.healthcareitnews.com/news/silo-one-healthcares-biggest-flaws

[7] D. Shaywitz. (2015) Data Silos: Healthcare's Silent Shame. [Online]. Available: https://www.forbes.com/sites/davidshaywitz/2015/03/24/data-silos-healthcares-silent-tragedy/#7859c8242907

[8] B. Clough. (2016) Healthcare Big Data Silos Prevent Delivery of Coordinated Care. [Online]. Available: http://healthitanalytics.com/news/healthcare-big-data-silos-prevent-delivery-of-coordinated-care

[9] D. Dennison, "PACS in 2018: an autopsy," *J Digit Imaging*, vol. 27, no. 1, pp. 7–11, Feb 2014.

[10] D. Mercer. (2014) Connected World The Internet of Things and Connected Devices in 2020. [Online]. Available: https://www.strategyanalytics.com/access-services/devices/connected-home/consumer-electronics/reports/report-detail/connected-world-the-internet-of-things-and-connected-devices-in-2020#.WRn7tVTyvMg0.76480659467625770.5684050139533516

[11] T. Sherif, P. Rioux, M.-E. Rousseau, N. Kassis, N. Beck, R. Adalat, S. Das, T. Glatard, and A. C. Evans, "Cbrain: a web-based, distributed computing platform for collaborative neuroimaging research," *Frontiers in neuroinformatics*, vol. 8, 2014.

[12] D. E. Rex, J. Q. Ma, and A. W. Toga, "The LONI Pipeline Processing Environment," *Neuroimage*, vol. 19, no. 3, pp. 1033–1048, Jul 2003. [Online]. Available: http://www.hubmed.org/display.cgi?uids=12880830

[13] D. Haehn, N. Rannou, B. Ahtam, E. Grant, and R. Pienaar, "Neuroimaging in the browser using the x toolkit," in *Frontiers in Neuroinformatics Conference Abstract: 5th INCF Congress of Neuroinformatics (Munich)*, 2014.

[14] W. L. Schulz, T. J. S. Durant, A. J. Siddon, and R. Torres, "Use of application containers and workflows for genomic data analysis." *Journal of pathology informatics*, vol. 7, p. 53, Dec 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/28163975

[15] T. Doel, D. I. Shakir, R. Pratt, M. Aertsen, J. Moggridge, E. Bellon, A. L. David, J. Deprest, T. Vercauteren, and S. Ourselin, "Gift-cloud: A data sharing and collaboration platform for medical imaging research," *computer methods and programs in biomedicine*, vol. 139, pp. 181–190, 2017.

[16] S. Das, A. Zijdenbos, D. Vins, J. Harlap, and A. Evans, "Loris: a web-based data management system for multi-center studies," *Frontiers in Neuroinformatics*, vol. 5, p. 37, 2012.

[17] I. B. Ozyurt, D. B. Keator, D. Wei, C. Fennema-Notestine, K. R. Pease, J. Bockholt, and J. S. Grethe, "Federated web-accessible clinical data management within an extensible neuroimaging database," *Neuroinformatics*, vol. 8, no. 4, pp. 231–249, 2010.

[18] A. Scott, W. Courtney, D. Wood, R. De la Garza, S. Lane, R. Wang, M. King, J. Roberts, J. Turner, and V. Calhoun, "Coins: An innovative informatics and neuroimaging tool suite built for large heterogeneous datasets," *Frontiers in Neuroinformatics*, vol. 5, p. 33, 2011.

[19] R. L. Harrigan, B. C. Yvernault, B. D. Boyd, S. M. Damon, K. D. Gibney, B. N. Conrad, N. S. Phillips, B. P. Rogers, Y. Gao, and B. A. Landman, "Vanderbilt university institute of imaging science center for computational imaging xnat: A multimodal data archive and processing environment," *NeuroImage*, vol. 124, pp. 1097–1101, 2016.

[20] H. Xue, S. Inati, T. S. Sørensen, P. Kellman, and M. S. Hansen, "Distributed mri reconstruction using gadgetron-based cloud computing," *Magnetic resonance in medicine*, vol. 73, no. 3, pp. 1015–1025, 2015.

[21] G. Kantasis, C. G. Xanthis, K. Haris, E. Heiberg, and A. H. Aletras, "Cloud gpu-based simulations for {SQUAREMR}," *Journal of Magnetic Resonance*, vol. 274, pp. 80 – 88, 2017.

[22] I. D. Dinov, J. D. Van Horn, K. M. Lozev, R. Magsipoc, P. Petrosyan, Z. Liu, A. Mackenzie-Graham, P. Eggert, D. S. Parker, and A. W. Toga, "Efficient, Distributed and Interactive Neuroimaging Data Analysis Using the LONI Pipeline," *Front Neuroinformatics*, vol. 3, pp. 22–22, 2009. [Online]. Available: http://www.hubmed.org/display.cgi?uids=19649168

[23] A. W. Services, "Aws: Cloud computing in healthcare," https://aws.amazon.com/health/, 2017.

[24] IBM, "Ibm cloud solutions for healthcare," https://www.ibm.com/cloud-computing/solutions/healthcare/, 2017.

[25] Microsoft, "Microsoft in health," https://enterprise.microsoft.com/en-us/industries/health/, 2017.

[26] D. Haehn, N. Rannou, P. E. Grant, and R. Pienaar, "Slice:drop: Collaborative medical imaging in the browser," in *ACM SIGGRAPH 2013 Computer Animation Festival*, ser. SIGGRAPH '13. New York, NY, USA: ACM, 2013, pp. 1–1. [Online]. Available: http://doi.acm.org/10.1145/2503541.2503645

[27] D. Ginsburg, S. Gerhard, J. E. C. Calle, and R. Pienaar, "Realtime visualization of the connectome in the browser using webgl," *Frontiers in Neuroinformatics*, 2011.

[28] F. Khan, V. Foley-Bourgon, S. Kathrotia, E. Lavoie, and L. Hendren, "Using javascript and webcl for numerical computations: A comparative study of native and web technologies," in *ACM SIGPLAN Notices*, vol. 50, no. 2. ACM, 2014, pp. 91–102.

[29] F. Mwalongo, M. Krone, G. Reina, and T. Ertl, "State-of-the-art report in web-based visualization," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 553–575.

[30] J. L. Bernal-Rusiel, N. Rannou, R. Gollub, S. Pieper, S. Murphy, R. Robertson, P. E. Grant, and R. Pienaar, "Reusable client-side javascript modules for immersive web-based real-time collaborative neuroimage visualization," *Frontiers in Neuroinformatics*, vol. 11, p. 32, 2017.

[31] P. Eckersley, G. F. Egan, E. De Schutter, T. Yiyuan, M. Novak, V. Sebesta, L. Matthiessen, I. P. Jaaskelainen, U. Ruotsalainen, A. V. Herz *et al.*, "Neuroscience data and tool sharing," *Neuroinformatics*, vol. 1, no. 2, pp. 149–165, 2003.

[32] J. Millan and L. Yunda, "An open-access web-based medical image atlas for collaborative medical image sharing, processing, web semantic searching and analysis with uses in medical training, research

and second opinion of cases," *Nova*, vol. 12, no. 22, pp. 143–150, 2014.

[33] D. Wood, M. King, D. Landis, W. Courtney, R. Wang, R. Kelly, J. A. Turner, and V. D. Calhoun, "Harnessing modern web application technology to create intuitive and efficient data visualization and sharing tools," *Frontiers in neuroinformatics*, vol. 8, p. 71, 2014.

[34] J. Nickoloff, *Docker in Action*. Manning Publications Co., 2016.

[35] K. Matthias and S. P. Kane, *Docker: Up & Running*. " O'Reilly Media, Inc.", 2015.

[36] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51–59, Feb 2013. [Online]. Available: http://dx.doi.org/10.1089/big.2013.1508

[37] M. Swan, "The quantified self: Fundamental disruption in big data science and biological discovery," *Big Data*, vol. 1, no. 2, pp. 85–99, Jun 2013. [Online]. Available: http://dx.doi.org/10.1089/big.2012.0002

[38] C. S. Greene, J. Tan, M. Ung, J. H. Moore, and C. Cheng, "Big data bioinformatics," *Journal of Cellular Physiology*, vol. 229, no. 12, pp. 1896–1900, 2014. [Online]. Available: http://dx.doi.org/10.1002/jcp.24662

[39] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical Surface-Based Analysis – I. Segmentation and Surface Reconstruction," *NEUROIMAGE*, vol. 9, pp. 179–194, 1999.

[40] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system," *NeuroImage*, vol. 9, pp. 195–207, 1999.

[41] "FreeSurfer," http://surfer.nmr.mgh.harvard.edu/.

[42] M. A. Rodwin, "The case for public ownership of patient data," *Jama*, vol. 302, no. 1, pp. 86–88, 2009.

[43] ——, "Patient data: Property, privacy & the public interest," *American journal of law & medicine*, vol. 36, no. 4, pp. 586–618, 2010.