Specific Aims

Imaging is an invaluable tool in cancer research. Indeed, over 6,000 registered cancer trials depend on one or more forms of imaging, including MRI, CT, PET, X-Ray, mammography, and SPECT. Yet, despite its ubiquity in clinical studies, the full power of imaging remains largely untapped. Advanced image acquisition methods, quantitative image biomarkers, and multivariate analyses across imaging, genomics, pathology, and other domains, all have the potential to dramatically impact cancer research and clinical practice. The NCI has recognized this potential and recently established the Quantitative Imaging Network (QIN) to develop and validate advanced imaging biomarkers. Nonetheless, imaging remains underutilized in cancer research due in large part to a lack of informatics technology. New informatics tools are needed to capture, organize, secure, and verify data collected in clinical studies; to automate image processing routines; to integrate imaging data with correlative phenotypic and genetic data; and to facilitate replication and re-analysis of published results. In human translational research settings, informatics tools are also needed to connect clinical information systems and research computing environments, where data can be processed using cutting edge algorithms. We propose to establish the **Integrative Imaging Informatics for Cancer Research (I3CR)**, a new U24 center to develop, distribute, and support open source software to address these needs.

This application is in response to the "Advanced Development of Informatics Technology (U24)" funding opportunity announcement, which calls for "advanced development and enhancement of emerging informatics technologies to improve the acquisition, management, analysis, and dissemination of data and knowledge in cancer research." The emerging informatics technology upon which the I3CR's platforms will be built is XNAT, a proven open source imaging informatics system. XNAT provides general imaging informatics capabilities, including data management workflows, data access control structures, and ergonomic user interfaces. XNAT will be enhanced to create a data management platform to manage imaging and related data in a range of cancer research contexts. In addition, a knowledge management platform will be implemented to document and publish the data, procedures, and resources that underlie scientific knowledge. The specific aims of the I3CR are:

Aim 1: Data Management Platform (DMP). I3CR will enhance XNAT to support common cancer research paradigms. I3CR enhancements will include additions to XNAT's core infrastructure as well as cancer-specific extensions to XNAT's existing data model, application programming interface (API), and user interface. Many of these extensions will focus on enabling external software applications to interoperate with the DMP. We have formed partnered with developers of many of the most widely used open source and commercial visualization and analysis platforms, database systems, and clinical devices to ensure that the DMP if fully interoperable with all aspects of research and clinical workflows.

Aim 2: Knowledge Management Platform (KMP). I3CR will build a knowledge management platform that integrates seamlessly with the DMP. The KMP will introduce a novel data structure – "knowledge portfolios" – to document and preserve the source data, derived data, analytic methods, and computational resources used to generate a scientific finding. Portfolios will be stored in the DMP's searchable content repository and references to source and derived data will link to the DMP database. A second novel element – the KMP "machine library" – will build on the Docker virtualization environment to store, version, and distribute the actual computing systems used to run processing and analysis routines. Together, these components comprise a platform that will encourage reproducible research and executable publications.

Aim 3: I3CR Pilot Network. A federated network of prominent cancer imaging programs will be established to evaluate and guide I3CR technology development. The network will use I3CR tools to undertake two pilot projects: 1) dose optimization in radiation therapy through image-guided dose sculpting and 2) glioblastoma tumor segmentation and patient outcome prediction using multi-parametric imaging. The successful execution of these projects will confirm our central hypothesis that the I3CR's integrative informatics technology enables cancer researchers to overcome technical hurdles that currently impede scientific progress.

The I3CR platforms will fill the gaps in informatics technology that currently hinder the use of advanced imaging in cancer research. The platforms will impact cancer research in clear and measurable ways: error rates and missing data will be reduced in clinical trials, the scale and scope of data repositories will expand, cutting edge imaging methods will be available in the clinic, and imaging studies will be more easily replicated and validated. As an early indicator of the importance of the I3CR and its likeliness of reaching a broad user base, a large number of high impact cancer research programs have already committed to using the I3CR's technology and participating in the pilot network.

Research Strategy

Significance

Why is imaging important in cancer research and patient care? Over 60% of randomized phase III cancer trials yield negative results [1]. To avoid these frustrating and expensive outcomes, researchers are turning to quantitative imaging endpoints in earlier phase trials [2]. While semi-quantitative measurements like the RECIST [3] or MacDonald criteria [4] have been used for many years, quantitative imaging markers are emerging as extremely sensitive surrogate markers for clinical outcomes [5], [6], and the importance of integrating imaging into response assessment has been formally recognized (in neuro-oncology for instance, with RANO [7]). The Radiological Society of North America, in partnership with the pharmaceutical industry, formed the Quantitative Imaging Biomarkers Alliance (QIBA) in 2007 to assist in validating specific biomarkers and establishing best practices for their use [8]. QIBA members have now demonstrated the utility of a number of quantitative imaging biomarkers including volumetric CT [9] and FDG-PET [10]. The National Cancer Institute (NCI) has provided major funding through the Quantitative Imaging Network (QIN) to explore an array of additional quantitative imaging markers ranging from quantitative MRI in prostate cancer to quantitative FDG- and FLT-PET in breast cancer. These programs illustrate the extraordinary potential of quantitative imaging endpoints to impact cancer research. In addition, specific research on imaging methods that have the potential to be used as biomarkers is an active area of investigation that is yielding advances that directly impact patient care. Novel MR methods [11]–[13] and PET imaging tracers [14], [15], for example, have improved the accuracy of diagnoses, enabled patients to be stratified by prognosis, and provided quantitative markers to guide therapy and assess patient response. Such advances in diagnostic imaging require careful research and validation before they can be used effectively in the clinic.

What are the imaging informatics needs? As stated in the Advanced Development of Informatics Technology Funding Opportunity Announcement to which we have applied: “Biomedical informatics is no longer an option but an integral component of all biomedical research”. Despite this truism, significant gaps exist in the imaging informatics resources currently available to cancer researchers:

- *Big data management*. Imaging studies often rely on large and complex protocols that may include multiple imaging modalities sampled at numerous time points over the course of a study. The increasing prevalence of quantitative acquisition and analysis approaches depend on sophisticated computational methods that generate additional derived data. Given these “big data” challenges, informatics tools are needed that have the capacity to organize data structures, enforce quality assurance practices, generate audit trails and provenance records, provide detailed reports and data tracking tools, and ultimately facilitate data analysis.
- *Data integration*. Imaging data are seldom acquired in isolation. Typically, they are but one aspect of a multi-domain research protocol that may include clinical, histopathological, genomic, and any number of other data types. Typically, these data are managed in specialized databases optimized for the particular domain (e.g. clinical data is often captured using a REDCap database). Informatics tools are needed to bridge these data management systems and enable users to efficiently mine data across domains.
- *Application interoperability.* Quantitative imaging measurements are typically made using specialized imaging workstations or software platforms. These measurements may be relatively straightforward manual measures of tumor size or volume, more complex automated segmentations of multiple tumor tissue types, or advanced image parameter estimates. Informatics tools are required to enable imaging workstations and software platforms to seamlessly exchange raw data following acquisition with archival tools (such as XNAT) and store derived measures alongside the source imaging data.
- *Human translational imaging*. Image acquisition and processing methods developed in the laboratory often fail to be efficiently validated and translated to clinical practice. This is often due to challenges in connecting clinical information systems to processing and analysis systems in research labs. Informatics tools are needed to facilitate the integration of experimental imaging methods into clinical practice.
- *Data sharing*. Despite increasing awareness of the value of data sharing and pressure from the NIH on investigators, data sharing is still a relatively rare practice. While there are many reasons for this [16], it often comes down to the technical challenges of anonymizing, organizing, and transferring large amounts of complex data to a public sphere. Informatics tools are needed to organize, prepare, and securely transfer data to managed data sharing environments like The Cancer Image Archive (TCIA)[17].
- *Knowledge management*. Data repositories like TCIA are designed to store and publish standard imaging data and a limited set of metadata. They are not equipped to manage the full spectrum of information and

resources that underlie scientific knowledge. Informatics tools are needed to generate, manage, and distribute these assets with the primary aim of enabling reproducible research.

These informatics gaps limit large-scale investigation, validation, and translation of experimental imaging biomarkers. ***We therefore propose to develop the Integrative Imaging Informatics for Cancer Research (I3CR) platform, a suite of informatics tools that integrates image acquisition, analysis workflows, phenotypic and genomic data sources, clinical information and treatment systems, and scientific knowledge (Figure 1).*** I3CR will focus on the provision of core data management services to organize imaging and complementary data into a formal hierarchy and enforce structured quality assurance protocols. It will use modern web service architectures paired with DICOM, the imaging industry standard data format and network protocol, to enable and streamline data exchange between workstation software, complementary database systems, and clinical information systems. Finally, it will include technology to streamline research operations, simplify data mining and analysis practices, and facilitate sharing of data and knowledge.

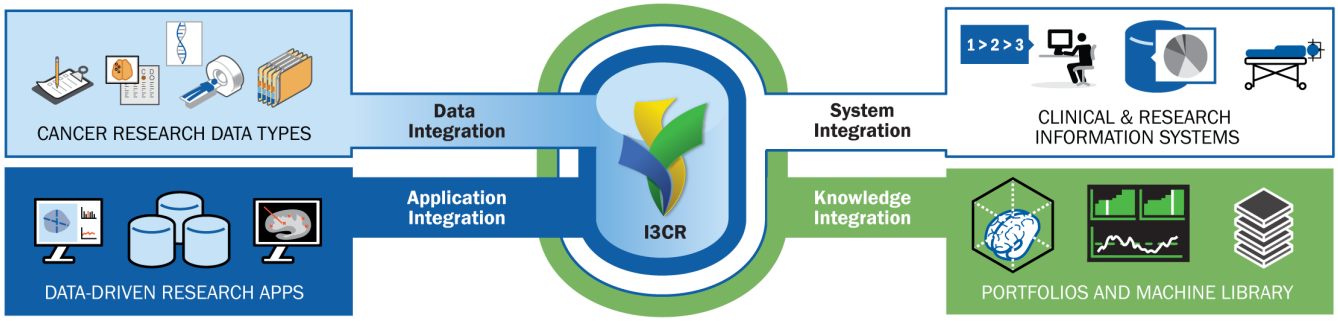

Figure 1. I3CR technology focuses on four levels of integration: data, applications, systems, and knowledge.

How will I3CR meet these needs? The I3CR team will tackle the ambitious goal of developing a comprehensive cancer imaging informatics technology by leveraging our experience with the XNAT imaging informatics platform [18]. Over the last decade, XNAT has emerged as the most widely used imaging informatics platform across a number of research fields. XNAT supports a large number of high impact programs, including many in the neurosciences [19]–[22], cardiovascular research [23], and a range of other fields [24]–[27], and has been commercialized to support FDA-regulated clinical trials requiring 21 CFR Part 11 compliance. XNAT has been open source since its inception and code contributions have been made by dozens of developers [28]. Its ongoing development is based in Dr. Marcus' laboratory at Washington University. XNAT will serve as the underlying framework upon which the *I3CR Data Management Platform (DMP)* proposed in Aim 1 will be implemented. By building on a stable existing framework, we ensure a low-risk path to producing a reliable production-grade platform for the cancer research community. In addition, components developed for XNAT by other projects can be incorporated into I3CR to augment its capabilities. For example, the Comprehensive Neuro-oncology Data Repository (CONDR) project, under the direction of Dr. Marcus, has developed XNAT modules to facilitate clinical research of glioblastoma multiforme (GBM) and metastatic brain tumors [25]. CONDR extensions to XNAT include data types to capture, enter, and display clinical histories, surgical encounters, biopsy records, histopathology assessments, volumetric region of interest representations, and quantitative representations of imaging phenotypes. These XNAT extensions will be adopted and expanded by I3CR to support a broad range of cancer research. Similarly, the Human Connectome Project (HCP) informatics team, under Dr. Marcus' direction, has developed an XNAT data mining module that

What exactly is XNAT?

XNAT is a web-based software platform designed to facilitate common management and productivity tasks for imaging and associated data. It consists of an image repository to store raw and post-processed images, a database to store metadata and non-imaging measures, and user interface tools for accessing, querying, visualizing, and exploring data. XNAT supports all common imaging methods, and its data model can be extended to capture virtually any related metadata. XNAT includes a DICOM workflow to enable exams to be sent directly from scanners, PACS, and other DICOM devices. XNAT's web application provides a number of productivity features, including data entry forms, searching, reports of experimental data, upload/download tools, access to standard laboratory processing pipelines, and an online image viewer. A fine- grained access control system ensures that users are restricted to accessing only authorized data. XNAT also includes a web services API for programmatic access and an open plugin architecture for extending XNAT's core capabilities.

enables users to run analyses on high performance computing and cloud computing resources. This engine will be a key component of the knowledge management platform proposed in Aim 2.

What is knowledge management? Knowledge management includes the generation, provenance, and distribution of derived scientific "knowledge" obtained through processing, modeling, and analysis of source data. At the most abstract level, knowledge includes the assertions and conclusions that are often made in publications, such as: "Increased rCBV [relative cerebral blood volume] measures are associated with poor overall survival in GBM" [29]. This sort of high-level statement is highly derivative and, without access to the underlying data and methodology, is difficult to verify. The methods sections of publications provide readers with a synopsis of this information and lab notebooks provide a more detailed though frequently inscrutable private record. However, recent technological advances, including computer system virtualization, data provenance reporting formats, and workflow management tools, provide an opportunity to formally document and publish a fully realized rendering of this information. In addition, thinkers at the forefront of the open science movement have begun to establish best practices for conducting and publishing reproducible research [30]–[34]. In Aim 2, we will leverage these technologies and guidelines to create the *I3CR Knowledge Management Platform*, a framework for generating scientific knowledge through a reproducible, self-documenting process. While knowledge management is a broad topic of interest across all scientific research, and the proposed I3CR platform will be useful beyond cancer research, cancer research in particular has been highly susceptible to irreproducible results [35], [36]. Better knowledge management is an imperative step towards rectifying the reproducibility problem in cancer research. Furthermore, imaging science procedures, with frequent manual interventions (e.g. generating regions of interest) and complex computational dependencies, are particularly difficult to execute consistently over the course of a study and to describe comprehensively in a publication. Knowledge management technology addresses these issues by formalizing workflows, recording data provenance, and linking to source and derived data.

How is I3CR different? There have been past attempts to develop integrative imaging informatics systems for cancer research [37]–[40]. Unfortunately, these projects have not resulted in widely adopted applications. Ongoing projects similarly underserve the cancer research community. The NCI Cancer Imaging Program's survey of available imaging informatics systems and archives [41] reveals striking gaps in capabilities. Most systems were designed for neuroimaging (e.g. XNAT, LONI IDA) or generic documents (e.g. MIDAS) and thus lack support for common cancer imaging measurement protocols (e.g. RECIST), data formats (e.g. DICOM RT), and annotations. I3CR will implement these much-needed features. The most widely used informatics system in the field is The Cancer Image Archive (TCIA), but it is a public resource that is limited to distribution of "final" data sets. I3CR on the other hand will cover the full research cycle, including data collection, integration, and processing, and will link directly with clinical information systems at study sites. Our proposed software development approach will also avoid the top-down "build it and they will come" development process that has plagued past high profile cancer informatics efforts [42]. I3CR development will be use-case driven and tied closely to a network of real world pilot projects and research programs, including the QIN.

Why is this the right team? The interdisciplinary I3CR team includes expertise in informatics, cancer imaging and treatment, and cancer biology. Dr. Marcus, co-PI, is the leader of the XNAT program and a Co-PI on the CONDR project, which together provide relevant expertise in both informatics and cancer imaging. Dr. Wahl has been developing and validating cancer imaging biomarkers and response criteria for three decades. Drs. Michalski and Robinson bring expertise as radiation oncologists with expansive research programs. The remainder of the team provides engineering, coordination, and quality assurance expertise.

A note on this submission. This proposal is a new submission based on a previous proposal that was reviewed very positively but unfortunately was not funded. With this new proposal, we have addressed the all major concerns expressed by past reviewers. The investigator team has been expanded to include cancer domain experts in imaging biomarkers (Wahl), response assessment (Wahl, Robinson), radiation therapy (Robinson, Michalski, Doran), and clinical trials (Michalski). We propose deeper support for radiation therapy, including collaborative relationships with the leading vendors of RT systems, additional processes for capturing and analyzing dosing plans, and longitudinal data models for tracking information related to specific lesions.

Innovation

In Aim 1, we will implement the I3CR Data Management Platform, a relatively low-risk but high-yield effort. This work will result in the stable, production-grade informatics technology that is desperately needed by cancer researchers. The platform will enable highly innovative hypothesis-driven research that would not be possible without the technology infrastructure. **This approach to improving on existing technology is**

explicitly stated as the goal of the FOA to which we have responded: "The purpose of this… FOA is to invite applications for advanced development and enhancement of emerging informatics technologies to improve the acquisition, management, analysis, and dissemination of data and knowledge in cancer research."

In Aim 2, we propose a higher risk undertaking that will produce a novel mechanism for managing the knowledge generated in cancer imaging studies. The I3CR Knowledge Management Platform will enable researchers to publish the data and methods underlying their scientific findings at an unprecedented level of detail, enabling their colleagues to validate, replicate, and extend their work. Cancer research has been notoriously difficult to reproduce [35], [36], leading to many failures in the drug development pipeline. The proposed platform is an imperative step toward improving reproducibility in cancer clinical research.

In Aim 3, we propose to develop a network of early adopters and specific pilot projects that will be used to evaluate I3CR technology. This approach provides a unique opportunity for receiving rapid feedback that can be used to adjust I3CR development to meet real world needs. The pilot projects in particular are designed to yield scientific results that in the absence of I3CR technology would not be possible.

Approach

Specific Aim 1. Data Management Platform (DMP)

Overview. The I3CR Data Management Platform (DMP) will enable cancer researchers to manage their imaging, clinical, histopathological, and genomic data in a single integrated database accessible from a secure user-friendly web application. The web interface will include data entry forms and file importers for many common cancer imaging data types and file formats. Its search interface will allow users to query across complex combinations of imaging and non-imaging criteria to build, analyze, and share custom data sets. In addition to its own web interface, the DMP will include an application programming interface (API) for interacting with the system from external software applications. The DMP will integrate with complementary data management environments (e.g. REDCap) and cancer imaging repositories (e.g. TCIA) and will interoperate seamlessly with the I3CR Knowledge Management Platform. Specific advances to be included with the DMP are listed in Table 1.

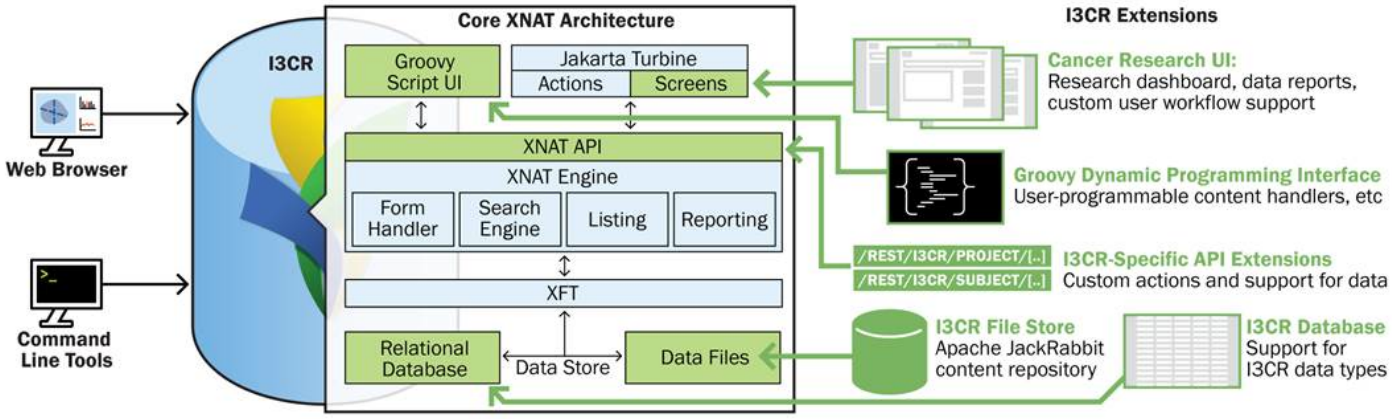

Figure 2. The Data Management Platform architecture builds new functionality and data structures into XNAT, including a searchable content repository, dynamic programming interfaces, and workflow engine.

Aim 1.1 Platform Implementation

Architecture. The DMP architecture (**Figure 2**) will build on the XNAT imaging informatics platform. XNAT uses the Jakarta Turbine framework to implement a Java-based model-view-controller architecture, the PostgreSQL relational database system to store metadata, and Spring Framework to implement security, messaging, and caching. I3CR's use of this infrastructure will allow the DMP to leverage XNAT's extensible data model, web services API, and plugin system. A representative set of these extensions is listed in **Table 1** and described in more detail in the sections below. In addition, the DMP architecture will include several significant modifications and additions to XNAT. First, XNAT's standard data file management component will be replaced with the Apache Jackrabbit content repository platform. Jackrabbit's features, which include advanced metadata management, versioning, and linked files, provide more robust and flexible file management than XNAT's directory-tree based file storage. Files on remote object stores like Amazon S3, for example, could be linked into a DMP database through Jackrabbit. In addition, Jackrabbit's native hierarchical structure is an ideal match for the portfolio structure used by the Knowledge Management Platform (see Aim

2). Second, a dynamic scripting module will be implemented to allow custom code to be inserted into the DMP application. The module will support code written in the Groovy programming language, which can be dynamically compiled and executed within the Java virtual machine. The module will include a web-interface for creating and editing code and a rule engine to configure when the code is executed. The module will be used throughout the DMP to provide a streamlined mechanism for extending XNAT functionality. Custom file importers, for example, will be generated using this feature (see *File Formats* below). Finally, a workflow engine will be implemented using the Activi Business Process Management Platform. Activi allows users to create, execute, and monitor multi-step workflows. Activi differs from and complements "pipeline" systems [43]–[46] in that it goes beyond sequential machine processes to include higher level workflows that include manual procedures ("review image quality") and general automation ("generate quarterly report"). Activi's internal API will be embedded within the DMP API. Custom UI components will be created in the DMP web application for interacting with Activi workflows.

File Formats	
Data Format	**Description**
Annotation & Imaging Markup (AIM)	NCI-supported format for recording qualitative and quantitative imaging measurements.
DICOM RT	Industry standard format for recording radiation therapy plans as well as research ROI measures.
DICOM Segmentation Objects	Industry standard format for recording 2D and 3D structures.
Elekta DVH	Lesion-specific dose profiles generated in Elekta radiation treatment plans.
Analyze Objects	Widely used format for recording 2D ROIs.
REDCap	XML-formatted export of REDCap case report form data.
Excel/CSV	Generic and Microsoft formats for storing tabular data.
Data Model	
Data Type	**Description**
WHO Criteria	Response criteria for solid tumors using bi-linear summation measurements of radiographic images.
Response Evaluation Criteria in Solid Tumors (RECIST)	Response criteria for solid tumors using single linear summation measurements of radiographic images.
Response Assessment in Neuro-Oncology (RANO)	Response criteria for solid tumors using radiographic and clinical factors.
PET Response Criteria in Solid Tumors (PERCIST)	Response criteria for metabolic tumor response assessment with (18)F-FDG PET
Deauville Criteria	Response criteria for reviewing interim PET scans in advanced Hodgkin lymphoma
Hopkins Criteria	Response criteria for reviewing PET/CT in head and neck squamous cell cancer patients (HNSCC)
VASARI	Assessment of human gliomas based on anatomic MRI scans, includes 26 separate observations.
Histopathology Assessment	Assessment of histopathological characteristics of a tissue biopsy, based on CONDR prototype.
Surgical Encounter	Captures surgical outcomes and biopsy details, based on CONDR prototype.
DTI	Standard DTI-based measures, including mean diffusivity and fractional anisotropy.
DCE	Standard DCE-based measures, including blood flow and blood volume.
Networking, Communications, and Programming Interfaces	
Service	**Description**
DICOM web services	Adapter for DICOM's REST web services protocols (WADO-RS, QIDO, STOW)
REDCap import/export	Adapter for REDCap data feed utilizing REDCap's XML format.
TCIA export	Metadata and image transfer service to securely export full data sets to TCIA.
CDISC export	Adapter for exporting data in CDISC format, the standard for clinical trial data sets.
File Parser	Programmable service to enable extraction of file-format specific metadata and data.
User Interface	
Module	**Description**
Workflow UI	Modules to manage, utilize, and review workflows in the I3CR workflow engine.
File Parser editor	Administrative module to create and edit specific parsers for extracting information from files.
Dashboards	Screen for displaying reports, charts, and operations specific to use case configurations.
Integration tools	Modules for data integration tasks, including REDCap and TCIA data exchange.

Table 1. Representative DMP extensions to the XNAT platform.

Data Model. Cancer studies depend on a variety of qualitative and quantitative assessments of imaging data, often as surrogate endpoints in clinical trials where a primary outcome of overall survival may take many years to obtain. The DMP will add a number of new data types to the XNAT data model to represent assessments commonly used in the field including standards such as RECIST [3], PERCIST [47], Deauville criteria [48], MacDonald criteria [28], Hopkins criteria [49], and RANO [29] (**Figure 3**). The data types will be created using XNAT's standard XML Schema-based extension mechanism. For each data type, XNAT will automatically generate the model and controller components of the web application, including database tables and UI views of the data, and enforce data access restrictions (see *Security* below).

File Formats. Cancer researchers depend on a number of file formats in their work, including DICOM and Analyze for binary image data, AIM and DICOM Segmentation Objects for image measurements, and XML and CSV for clinical data. XNAT already includes the ability to upload files of any type and associate them with a

particular data element. However, XNAT simply stores the files and does not parse them to extract the metadata and data encoded in the files. Using the dynamic programming module describe above, format-specific parsers will be developed to support common file formats used in cancer research. Each parser will include the necessary logic to read the file, identify extractable content, convert that content into an XNAT-compliant format, and store the extracted to the DMP database. Once a parser has been created, it will be applied whenever a file of the matching type is uploaded to the system. **Figure 4** details this process, using as an example AIM-formatted annotations of tumor measurements and the ePAD visualization tool [50].

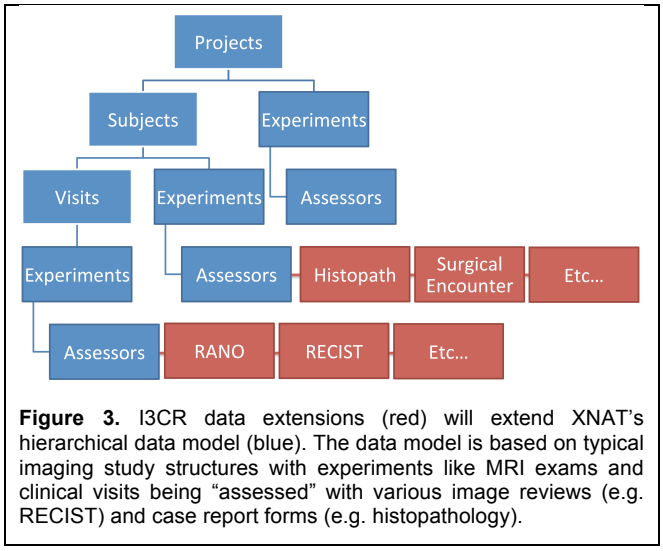

Figure 3. I3CR data extensions (red) will extend XNAT's hierarchical data model (blue). The data model is based on typical imaging study structures with experiments like MRI exams and clinical visits being "assessed" with various image reviews (e.g. RECIST) and case report forms (e.g. histopathology).

Networking, Communications, and Programming Interfaces. The primary communication protocols used by the DMP will include DICOM and HTTPS. DICOM is the industry-standard network protocol for data exchange between medical imaging systems, and XNAT already includes a number of DICOM services, with a particular focus on storage of DICOM data. XNAT's storage service directs incoming data to appropriate projects based on values extracted from the DICOM metadata. As the DICOM files are being stored, site-wide and project-specific anonymization profiles are applied, a configurable set of DICOM tags are extracted into the XNAT database, and a validation service is executed to evaluate whether incoming data comply with the project's protocol. I3CR extensions to XNAT's DICOM services will include the addition of DICOM query/retrieve to locate and import DICOM studies from clinical PACS as well as DICOM export to push DICOM data to the PACS.

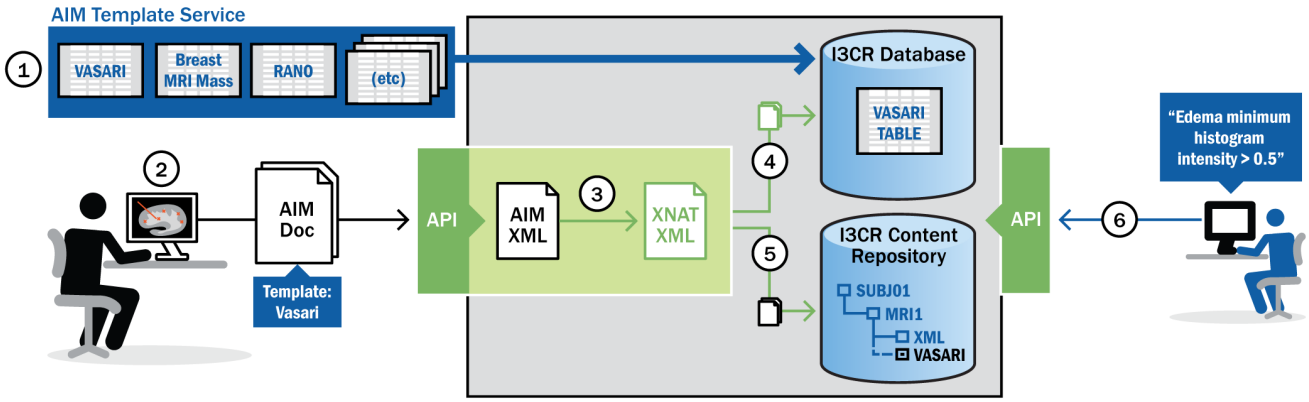

Figure 4. Import of files into the I3CR database is illustrated in this example of importing an AIM document from the ePAD program. This process entails several steps: 1) Prior to import, tables and other XNAT structures are created in the I3CR database through an automated script that transforms AIM templates stored in the AIM Template Service into XNAT schemas, including the VASARI brain tumor feature set as illustrated here, 2) AIM documents are uploaded from ePAD via the DMP API, 3) Custom Groovy import code identifies the AIM template that matches the AIM XML and converts the document to the matching XNAT XML, 4) The XNAT XML is stored in the database via standard XNAT processes, 5) the original AIM document is stored in the content repository, and 6) Users are able to query the database for data matching arbitrary criteria from the AIM markup – here a user searches for cases where mean "Edema minimum histogram intensity" is greater than 0.5. As described in the letter provided by Dr. Rubin, ePAD will be modified to support this process.

While DICOM serves as the standard within medical imaging systems, HTTPS is the standard for the vast majority of web communication and will be used for most of the DMP's services. XNAT's HTTPS-based REST [51] web services API provides broad access to XNAT functionality, including services for storing and retrieving data, querying the database, creating and managing projects, performing administrative tasks, and running processing pipelines. The API is used throughout XNAT itself and also by an array of external applications that integrate with XNAT, including the PyXNAT Python library [19]. I3CR will leverage XNAT's API extension mechanism to expose a number of new capabilities aimed primarily at integrating the DMP with external applications. Primary amongst these, we will implement DICOM's own emerging HTTPS web services. These include WADO-RS for retrieving DICOM objects [52], QIDO for querying DICOM repositories [53], and STOW for storing DICOM objects [54]. This standard enables access to DICOM objects through a REST services

interface that is highly compatible with XNAT's own API and user authentication mechanisms. Additional API services will be implemented in support of workflow extensions, file parsing, and integration with the KMP.

User Interface (UI). The DMP will include a user-friendly web UI based on the XNAT 2.0 interface (**Figure 5**). XNAT's uses jQuery and the Bootstrap JavaScript framework to generate highly dynamic and responsive pages. The interface can be extended and customized at various levels. Additional or alternative page elements can be incorporated on any of the standard pages and entirely new pages can be embedded to create custom views and workflows. This customizability will be leveraged heavily by the I3CR to create the context-specific configurations required by each of the driving use cases detailed in Aim 3. Key features common to all configurations will include: detailed project, subject, and experiment reports; a dynamic search engine that enables complex searches across multiple data types; a web-based image viewer; and administrative tools to manage, monitor, and customize the site.

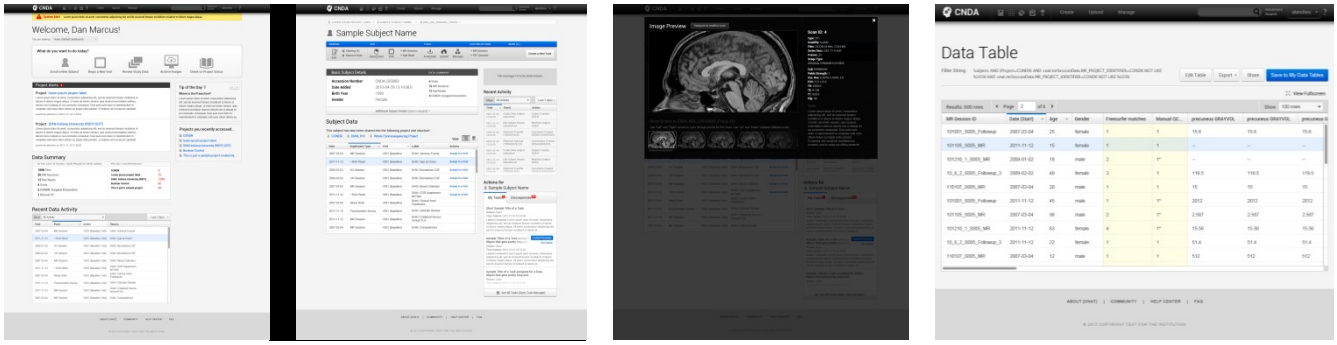

Figure 5. The DMP user interface. These examples of (from left) the default homepage dashboard, subject report, image viewer, and data table illustrate the look and feel of the UI. The UI's plugin model allows custom views to be added to specific pages or to create entirely new "dashboard" screens to create custom workflows.

Security & Access Control. The DMP will utilize the security, privacy, and encryption mechanisms included in XNAT to control access to data and prevent exposure to protected health information. All access will be controlled by password-protected user accounts and all traffic will be SSL-encrypted. In XNAT, a user's access to data is controlled at the project level, where the user's role controls access privileges for each type of data in the project. The XNAT web application has passed numerous security audits, including comprehensive assessments conducted by the University of Texas Chief Security Office and the Columbia University Central IT Division. The I3CR development process will include a vulnerability assessment prior to every release.

Aim 1.2 Application Integration

The DMP will be designed to integrate seamlessly with a wide variety of external applications and information systems. Initial integration targets will focus on several widely used applications, which will provide immediate impact to a large user community (**Table 2**). These initial targets were selected to cover a number of different categories – image visualization and analysis, databases and data repositories, surgical navigation and radiation therapy systems – to ensure that I3CR's integration approaches are broadly applicable.

Visualization and Analysis Applications	
Application	**Description**
3D Slicer	Open source visualization platform supported by the QIICR U24 program.
ePAD	Web-based image visualization and annotation application supported by the Stanford QIN.
ClearCanvas	Commercial open source visualization platform with full AIM support.
XIP	NCI supported workstation and reference implementation of DICOM Application Hosting interface.
Databases	
Database	**Description**
REDCap	Case report form-based database and data entry application.
TCIA	NCI-supported cancer research imaging data repository.
Clinical Systems	
Vendor	**Description**
Varian	RT treatment systems
Elekta	Radiosurgery treatment systems

Table 2. Representative applications that will be integrated with DMP.

Visualization and Analysis Applications. Integration with image visualization and analysis applications will include the ability to retrieve images from an XNAT instance to the application for visualization, processing and analysis, and subsequent export of derived data and image back to the DMP instance. Either the standard XNAT API or the emerging DICOM web services will be used as deemed appropriate for each application. A UI

component will be added to application to enable navigation of the DMP-hosted data. Functional prototypes using our proposed integration approach have already been implemented in 3D Slicer, ClearCanvas MITK, ePAD, and several other applications (see letters from Drs. Kikinis and Rubin).

Database Applications and Data Repositories. Our set of initial target databases is motivated by several factors: good fit with our core use cases, high overall impact, and reusability of the generated code and design patterns. REDCap is widely used within the oncology research community to capture case report-based clinical data including assessments, questionnaires and surveys, and has production and in-development implementations of many common cancer assessments (see letter from Dr. Paul Harris, director of REDCap). Integration with REDCap will include single sign-on across DMP and REDCap instances, automated real-time exchange of data, dynamic querying, and unified data entry workflows. This integration will give investigators a pairing of cooperative databases that together provide seamless support for studies that include clinical case report forms and imaging data. Our second database target, TCIA, is a rapidly growing NCI resource for hosting and sharing high impact imaging-based cancer research data [17]. Integration with TCIA will include automated upload and download of data between TCIA and local DMP instances, giving users freedom to explore and analyze the data using their own methods and to share their data through a robust and highly visible resource. An XNAT-based data exchange prototype has already been implemented on TCIA that allows TCIA-hosted data sets to be accessed via XNAT web services.

Clinical Systems. A key component of human translational research is the ability to integrate research workflows with clinical information systems and treatment platforms. Two representative vendors have been selected as initial partners for clinical integration. Varian Medical Systems is the world's largest provider of external beam radiation, brachytherapy, and proton therapy systems. Likewise, Elekta is a world leader in radiation therapy systems, including the market leading GammaKnife image-guided radiosurgery systems. I3CR will enable export of advanced image processing output generated through DMP workflows, including functional localization maps, fiber tracts, tissue segmentation, molecular imaging parameters, and optimized dose plans, to these platforms. Precise dosimetry data generated during a patient's treatment is stored as a DICOM RT file and provides complex dosimetric data that can be considered as variables in outcomes studies as well as potential source data for subsequent ROI and VOI generation. I3CR will automate the import and processing of these data. From these two initial target clinical systems, we will develop a generalized library for interfacing with clinical systems, using standards like DICOM and IHE wherever possible. Both Varian and Elekta are members of the I3CR pilot network (see Aim 3 below) and are committed to providing the necessary technical support to achieve I3CR's aims (see letters of support from Dr. Khundia and Ms. Gilmore-Lawless).

DMP Federation. Data federation technology will be implemented to support sharing of data between DMP instances. Globally unique identifiers will be produced within each instance to avoid collision of data elements between systems [55]. A SAML-based single sign-on system will be implemented to enable users to be authorized and authenticated across instances [56]. A central data portal developed to support the pilot network (see Aim 3) will be expanded as I3CR technology matures to provide a global registry of available data and analytic tools across all I3CR users who choose to participate in the registry.

Specific Aim 2: Knowledge Management Platform

Overview. The I3CR Knowledge Management Platform (KMP) will allow researchers to generate, document, and share scientific finding and the source data, software, and other resources associated with them. The KMP will integrate seamlessly with the I3CR DMP, and it will be designed to interoperate with and complement the capabilities of widely used workflow management systems.

Architecture. As illustrated in **Figure 6**, the KMP will be comprised of novel data structures called *knowledge portfolios* and several infrastructure components: 1) a *content repository* for storing and searching portfolios, 2) an *application programing interface (API)* for creating, modifying, and accessing portfolios from external applications, 3) a *machine library* for storing and accessing the computing resources used to process and analyze the data included in portfolios, and 4) a *user interface* for interacting with portfolios. The KMP components will be implemented as XNAT modules, allowing them to be easily integrated into the I3CR Data Management Platform. Like all data within XNAT, access to KMP portfolios will be limited to authorized users.

Knowledge Portfolios. The knowledge portfolio data structure that will be implemented to record the layers of source and derived data and processing and analysis conducted to obtain a scientific finding. Portfolios will be distributed in the JSON format, an open standard that is human readable, compact, easily generated and parsed, and supported by content repositories. Portfolios will include both scientific and supporting elements

(**Table 3**). Portfolio elements will be designed to embed content directly in the document or, for large content, to link to it as an external resource. The portfolio format will be structured hierarchically in order to capture the sequential nature of image processing and analysis, allowing for multiple stages of processing and analysis.

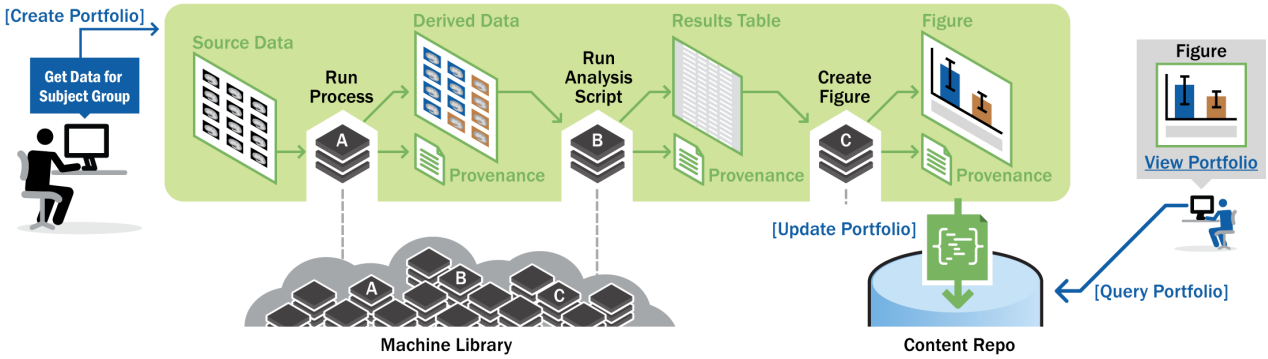

Figure 6. The I3CR Knowledge Management Platform builds on the knowledge portfolio data structure (green box), which includes links to the source data, derived data, data tables, and figures that underlie a scientific finding. The scripts and applications used to generate the finding are preserved in Docker containers in the Machine Library. Portfolios are stored in a searchable Jackrabbit content repository. The KMP includes an API (blue text) to interact with portfolios from external applications and a UI for user-focused browsing portfolio catalogs and deep exploration of individual portfolios.

Scientific Content: core assets used to gain knowledge, interpret results, conduct analyses, and otherwise get science done	
Source Data	Source data are the raw data obtained through a data collection event (e.g. a brain scan).
Derived Data	Derived data are the result of an operation conducted on source data or other derived data.
Tables	Tables contain two-dimensional data, often with rows representing individual study subjects and columns representing the various measurements generated for each subject. Tables are a subtype of derived data and result from completing an analysis
Process	Processes are computational or statistical procedures that operate on data (source, derived, or table) generate derived data and figures. The Process element's attributes include a human readable description and the algorithm used to conduct the analysis.
Figures	Figures are graphical renderings of tables and other data, typically generated as part of a statistical analysis.
Assertions	Assertions are conclusions or "knowledge" obtained from analyses. Assertions are typically stated as facts though they may require additional context, theory, and discussion outside the scope of the portfolio to interpret. Assertions may eventually even be proven false. Nonetheless, documenting an assertion directly in the portfolio provides a link between emerging scientific knowledge and the data and procedures underlying that knowledge.
Supporting Content: ancillary structures used to tracking and documenting the scientific assets	
Data Dictionary	Data dictionaries formally define the measures included within a portfolio, linking when possible to published ontologies. The columns within a table, for example, would each be associated with a dictionary entry.
Provenance	Provenance entries detail the sequence of operations that were undertaken to generate a particular asset, including the names and versions of all scripts and applications along with input arguments provided to the executables. Provenance records will typically be generated by workflow systems like Taverna, Galaxy, and XNAT Pipeline Engine and will follow the PROV-DM format
Script	Scripts are machine executable code to complete an analysis or other data processing. Script attributes will include the name, content of the script, and execution environment for running the script.
Machines	Machine elements document the computational resource that was used to a processing or analysis procedure (as documented in the Process element). Machine element attributes link to entries in the Machine Library.
Metadata	Metadata elements document information about the version, commentary, status, security and other content for managing a portfolio

Table 3. Portfolio Content

Content Repository. Portfolios will be stored in the DMP's Apache Jackrabbit content repository, which uses a hierarchical structure that is well-suited to the proposed portfolio format. As portfolio documents are stored in the repository, typically through the platform's API, the contents will be extracted from the document and stored as nodes in the repository. Jackrabbit nodes are fully searchable through Jackrabbit's native query mechanisms. Thus a user could easily locate all portfolios that include a particular subject or that include analyses using a particular script. Jackrabbit also enables versioning of portfolios. Because portfolios will be very small in size (large content like images will be captured as links), the repository will be able to manage a virtually limitless number of portfolios, including many derivate "forks" of portfolios. In addition to storing the portfolios, the repository will include mappings between portfolios and publications and to specific contents within a publication (e.g. a graph or a figure).

Application Programming Interface. The API will provide services for external applications (including the KMP's own user interface) to interact with portfolios (**Table 4**). The API will be implemented using Apache Sling, which provides a REST interface to content repositories like Jackrabbit. Sling includes a JSON parser that will be configured to automatically translate between portfolio documents and the Jackrabbit repository.

Machine Library. A unique feature of the KMP is that it will allow users to preserve and reuse the computing systems on which processing and analyses are done, including the operating system, system configuration, and installed applications. This capability will be built using the Docker virtualization platform, which packages all of the components needed for a machine as Docker "containers" that can be saved, distributed, and run

anywhere. Containers have many virtues that will benefit the KMP, including that they can be versioned and saved with full state preservation at any point time. Thus a KMP user will be able to interact with a container as if it were a typical machine, installing and running applications and scripts to conduct an analysis, and upon completing the analysis, save the container in the KMP machine library. Containers saved in the library can then be documented in portfolios and reinstated at any time from anywhere. This ability has the potential to be truly transformative: users will be able to run any analysis contained in any of their own portfolios or portfolios shared with them. An analysis associated with a figure in a paper, for example, could be validated by retrieving the associated portfolio and launching the relevant container. Users will be able to copy a portfolio, make modifications to specific elements, and rerun the analysis. By revising the portfolio source data, a finding could be quickly replicated in a new subject group. By modifying a computational model, alternative methods could be quickly compared. Unlike other virtualization systems, the Docker infrastructure manages containers highly efficiently, so the KMP machine library will be able to store a virtually limitless number of machines.

Service	Function
Create Portfolio	Creates a new portfolio in the system with default metadata, including date of creation and private access privileges, and returns portfolio's URL. The portfolio will be permanently accessible to authorized users from the provided URL.
Update Portfolio	Adds to or modifies elements in the portfolio. With each update, a new version of the portfolio is created and previous versions are accessible by setting a version number attribute on the portfolio's URL.
Fork Portfolio	Creates an exact copy of the portfolio with a new URL. The new portfolio can be modified without affecting the original.
Update Portfolio Access Privileges	Sets the privileges for a portfolio, including who can access, modify, and copy the portfolio. By default portfolios will be private. Access to particular data linked to within the portfolio will be separately controlled by the system that manages the data.
Update Metadata	Modifies the portfolio metadata to the indicated values. Metadata fields include status, name, description, and notes.
Anonymize Portfolio	Forks a portfolio and replaces PHI with anonymized values suitable for open access sharing.
Search Repository	Searches enable users to find portfolios that match specified criteria. Criteria can be included for any element in the portfolio.

Table 4. The KMP's core API services.

User Interface. The UI will allow users to manage, share, discover, and fork portfolios. Much of UI's style and functionality will be modeled after github.com, a popular environment for publishing software code. The interface will include a high level page for searching and browsing the catalog of portfolios. For each portfolio, a detail page will present the hierarchical content of the portfolio in a concise, consumable format. The UI will be embedded within the DMP using XNAT's module packaging capability. By linking the UI into the DMP, we will be able to leverage the DMP's security and access control system and to link portfolios directly to the data contained in them. For example, a list of all analyses that a particular subject has been part of can be displayed directly on the subject detail page in the DMP.

Specific Aim 3: Pilot I3CR technology in a network of real world cancer research programs.

The central hypothesis of the I3CR program is that our integrative informatics technology will enable cancer researchers to overcome technical hurdles that currently impede scientific progress. To test this hypothesis, a network of early adopters will use the I3CR platforms to share their data, tools, and computational resources with others in the network. Network members will use I3CR technology in their own studies and participate in two specific pilot projects. Successful execution of these projects will confirm I3CR as a solution to technical challenges that currently block scientific progress.

The I3CR Network. The I3CR pilot network includes 13 organizations (**Table 5**), including 6 QIN projects, 5 additional cancer imaging technology programs, and 2 commercial partners. Each organization in the network will use the I3CR Data Management and Knowledge Management Platforms to manage, document, and share data. As described in the letters of support provided by each of the participating sites, the shared data will include a range of imaging modalities, quantitative derived metrics, quantitative histopathology, treatment plans, and clinical and outcomes data. Dr. Pamela LaMontagne, the I3CR outreach coordinator, will assist each site in using I3CR technology and will elicit feedback from participants to guide I3CR development.

Supporting Technology. The DMP instances deployed at each site will be linked using the data federation tools described in Aim 1.2 to enable secure sharing of data on the network. This federation architecture allows owners of data to maintain complete control over which users on the network have access to their data. A central web-based hub will be implemented to provide users with a global, searchable index of data available on the network. The hub will be based on a system developed under Dr. Marcus' direction for the Dementias Platform UK (DPUK) data federation (see letter from Dr. Ourselin). This system includes a data staging module that enforces DICOM anonymization profiles [57], masks sensitive data fields, and recodes patient identifiers to ensure patient anonymity.

Pilot Projects. These projects were selected because they cover a range of imaging modalities, therapeutic approaches, and response metrics, and they present significant technical barriers that the I3CR platforms are

intended to eliminate: complex multi-modal data, distributed heterogeneous data sources, computationally intensive analytic workflows, and dependence on clinical information systems and treatment platforms. For each project, I3CR technology will be applied to address these challenges in ways that we expect will lead to measureable improvements in clinical applications.

I3CR Pilot Program Members				
Program	**PI**	**Disease Focus**	**Therapeutic Focus**	**Pilot Contributions**
QIN – MGH	Rosen	GBM	DSC-MRI	GBM mapping imaging biomarkers; shared patient cases; I3CR platform evaluation
QIN – BWH	Fennessy	Prostate	Multi-parametric MRI	Shared patient cases; I3CR platform evaluation
QIN – Iowa	Buatti	Head and Neck	Multi-tracer PET	Shared patient cases; I3CR platform evaluation
QIN – Stanford	Rubin	Lymphoma, Colon	Multiple	Integration of image annotation markup with I3CR technology; shared patient cases; I3CR platform evaluation
QIN – Michigan	Cao	Head and Neck	DCE-, DW-MRI	RT optimization imaging biomarkers; shared patient cases; I3CR platform evaluation
QIN – OHSU	Huang	Breast, soft tissue sarcoma	DCE-MRI	RT optimization imaging biomarkers; shared patient cases; I3CR platform evaluation
CONDR -- WashU	Marcus, Fouke	GBM	Image-guided surgery	GBM mapping pipeline components; shared patient cases; I3CR platform evaluation
Inst. for Cancer Research	Doran	Breast, Others	RT	Integration of analytic tools into I3CR pipelines; shared patient cases; I3CR platform evaluation
Univ. College London	Ourselin	Multiple	Multiple	Integration of analytic tools into I3CR pipelines; I3CR platform evaluation
DR THERAPAT	Multiple	Prostate, Cervical	RT	RT optimization pipeline components; shared patient cases
QICCR	Kikinis	Multiple	Multiple	Integration of 3D Slicer-based visualization and analysis platform with I3CR technology
Varian	Khuntia	Multiple	RT	Integration of RT planning/treatment system with I3CR technology
Elekta	Gilmore-Lawless	Multiple	RT	Integration of RT planning/treatment system with I3CR technology

Table 5. I3CR pilot network members

Project 1: Optimizing dose in radiation therapy. *Background:* While radiation therapy is widely used to treat solid tumors throughout the body, the method is far from optimal. In particular, with current methods, radiation dose is typically delivered uniformly to a defined tumor target even though tumors are radiobiologically non-uniform. A number of I3CR pilot investigators are developing quantitative imaging methods that can be used to generate spatially optimized dose plans and measure response to resulting treatment. The Univ. of Michigan QIN project, for example, is using dynamic contrast enhanced (DCE) and diffusion weighted MRI to predict which regions of the tumor are likely to fail to respond to standard dose levels [58], [59]. Similarly, the OHSU QIN project is developing a DCE biomarker to predict response based on regional tumor vascularization [60], [61]. The DR THERAPAT project, a consortium of European cancer research centers, is developing analytic workflows to produce high spatial resolution maps of predicted tissue response to RT. These maps can then be used to boost the dose at RT resistant points in the tumor (**Figure 7**) [62]. *Approach*: While these methods individually show great promise, they are being developed largely in isolation from one another using limited patient data and treatment platforms. I3CR technology will be applied in 4 specific ways to expand on the work of these individual projects: 1) The DR THERAPAT workflow will be implemented as an I3CR pipeline that allows the QIN imaging biomarkers to be inserted into its parameter set; 2) additional patient cases shared within the I3CR pilot network will be used to improve the modelling and validation of the pipeline; 3) the pipeline will be integrated with a broader range of planning and treatment platforms (Philips, Varian, Elekta, and 3D Slicer) to expand its clinical utility, and 4) the pipeline will be distributed to the network via KMP portfolios and virtual machines to enable it to be validated in a range of clinical settings. The I3CR dose optimization pipeline will be modularized so that additional imaging modalities and alternative analytic methods can be evaluated.

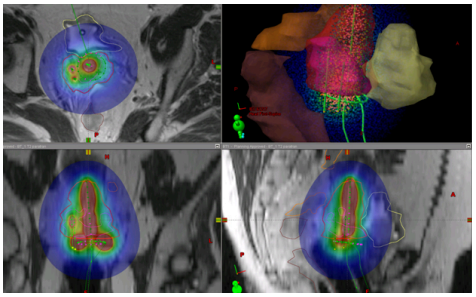

Figure 7. DR THERAPAT models enable homogenous dose plans (top left) to be "sculpted" to target radiation-resistant tumor regions (bottom).

Project 2. Multi-modal glioblastoma mapping. *Background:* Glioblastoma is the most common and lethal type of malignant brain tumor, largely due to their infiltrative growth habit [63]. Because prognosis in GBM patients is highly dependent on the extent of surgical resection [64], [65], a number of QIN and other quantitative imaging research groups are developing imaging biomarkers to accurately identify tumor from healthy tissue as well as to identify proximal eloquent cortex [66]. The MGH QIN project, for example, is developing advanced MRI of vascularization in GBMs [67]. The CONDR project under Dr. Marcus' leadership

has developed methods to register quantitative histopathology to pre-operative imaging, which allows precise spatial correlation of imaging characteristics with actual tumor pathology (**Figure 8**). *Approach*: The Medical Image Computing and Computer Assisted Intervention (MICCAI) Society has developed a "grand challenge" mechanism where scientists and engineers compete to build the best algorithm to solve a particular medical image analysis problem. We will develop a MICCAI grand challenge to build improved GBM analysis tools. The challenge will be similar to past challenges [68] but will introduce two new evaluation criteria: 1) image-registered digitized pathology slides will be used as an improved "gold standard" over manual image segmentation and 2) entries will be required to predict clinical outcome and genetic factors. Given these additional criteria, I3CR technology will play a critical role in the challenge. Data sets, including multi-modal imaging, digitized pathology slides, and clinical outcomes, contributed by CONDR and the MGH QIN will be shared with challenge contestants over the I3CR federated data network. KMP portfolios will be developed to document training and test data sets and entrants will be expected to document and distribute their tools through their own KMP portfolios and contributions to the KMP machine library. This approach ensures that challenge entries are well-documented, and makes high potential entries immediately available for deploying on the I3CR network for validating on additional patient cases.

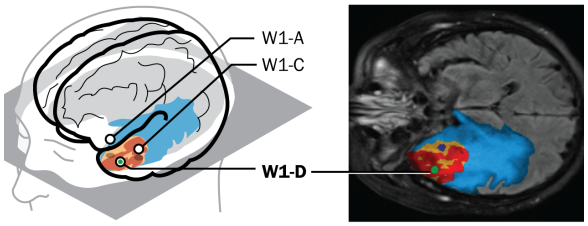

Figure 8. CONDR enables tissue samples (W1-A,C,D) to be precisely correlated with image parameters.

PROGRAM-WIDE ACTIVITIES

Evaluating I3CR Technology. We have hypothesized that I3CR technology will enable better science by providing a common framework for managing, documenting, and sharing data and tools. Based on this hypothesis, we expect several measurable outcomes: 1) The DMP will enable investigators to collect and share more data, so we expect that imaging biomarkers produced on I3CR technology will be more stable across a broad range of patient cases; 2) the KMP will enable source data and software components to be transparently documented and distributed, so we expect that imaging biomarkers produced on I3CR technology will be evaluated and improved on by a larger number of independent researchers; and 3) the I3CR platforms will be closely integrated with clinical systems, so we expect that imaging biomarkers produced on I3CR technology will be quickly translated to patient care. To evaluate these predictions, we will compare the stability, degree of independent verification, and clinical adoption of 10 imaging biomarkers developed with I3CR technology with 10 biomarkers developed without the benefit of I3CR. To ensure a level of fairness, candidate biomarkers for each type will be selected based on the most widely downloaded publications in prominent radiology and oncology journals. If these predicted outcomes are obtained, we will consider the I3CR program a success.

Software Development and Outreach. Input from the pilot network will drive an agile software development process based on short development cycles that return functional iterations of the platforms to end-users quickly and frequently. By exposing early adopters to software as it is developed, I3CR developers will get regular feedback to drive continuous improvement. A project timeline is provided in **Table 6.** As I3CR the platforms mature, we will host workshops, participate in MICCAI grand challenges, and publish in journals and conference proceedings. Further details on development and outreach plans to ensure high quality, well-documented software are described in *Resource Sharing*.

Aims	Y1	Y2	Y3	Y4	Y5
Aim 1.1 File Formats	XXX	XX	X		
Aim 1.1 Data Model	X	XX	X		
Aim 1.1 API Extensions	X	XX	XXX	XX	X
Aim 1.1 User Interface	XX	XX	XX	XX	XX
Aim 1.2 Visualization	X	XX	XX	X	
Aim 1.2 Databases		X	XX	XXX	XX
Aim 1.2 Clinical Systems		X	XX	X	
Aim 2 Portfolio	XXX	XX	X		
Aim 2 API		XXX	XX	X	
Aim 2 Machine Library		XXX	XX	X	
Aim 2 User Interface		X	XX	XXX	
Aim 3 Pilot Project 1	X	XX	XXX	XX	X
Aim 3 Pilot Project 2		X	XX	XXX	XX
Aim 3 Evaluation			X	XX	XXX

Table 6. I3CR Timeline. The level of effort on each task is indicated by the number of X's.

Potential Pitfalls. The project's biggest technical risk is the Docker platform, which is still in active development and may prove unstable. However, Docker is open source and has a large community of developers who quickly resolve most issues. A bigger challenge is that some commercial applications may be difficult to distribute within Docker containers. Often this issue can be resolved by linking to license files at runtime. As further mitigation, portfolios will be designed to allow researchers to document external dependencies not fully supported in a linked machine. A potential pitfall is that alternative platforms will emerge that compete with I3CR for attention and resources. We will stay abreast of this and similar work and be prepared to interoperate with systems that reach sufficient maturity. I3CR team members are participants in the relevant informatics working groups, which will us to keep informed of and influence developments in these groups.

References

[1] H. K. Gan, B. You, G. R. Pond, and E. X. Chen, "Assumptions of Expected Benefits in Randomized Phase III Trials Evaluating Systemic Treatments for Cancer," *JNCI J Natl Cancer Inst*, vol. 104, no. 8, pp. 590–598, Apr. 2012.

[2] L. D. Kaiser, "Tumor Burden Modeling Versus Progression-Free Survival for Phase II Decision Making," *Clin Cancer Res*, vol. 19, no. 2, pp. 314–319, Jan. 2013.

[3] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *European Journal of Cancer*, vol. 45, no. 2, pp. 228–247, Jan. 2009.

[4] D. R. Macdonald, T. L. Cascino, S. C. Schold, and J. G. Cairncross, "Response criteria for phase II studies of supratentorial malignant glioma.," *JCO*, vol. 8, no. 7, pp. 1277–1280, Jul. 1990.

[5] N. Butowski and S. M. Chang, "Endpoints for clinical trials and revised assessment in neuro-oncology," *Curr. Opin. Neurol.*, vol. 25, no. 6, pp. 780–785, Dec. 2012.

[6] L. P. Clarke, B. S. Croft, R. Nordstrom, H. Zhang, G. Kelloff, and J. Tatum, "Quantitative imaging for evaluation of response to cancer therapy," *Transl Oncol*, vol. 2, no. 4, pp. 195–197, Dec. 2009.

[7] J. G. Pérez-Larraya, M. Lahutte, G. Petrirena, G. Reyes-Botero, A. González-Aguilar, C. Houillier, R. Guillevin, M. Sanson, K. Hoang-Xuan, and J.-Y. Delattre, "Response assessment in recurrent glioblastoma treated with irinotecan-bevacizumab: comparative analysis of the Macdonald, RECIST, RANO, and RECIST + F criteria," *Neuro Oncol*, vol. 14, no. 5, pp. 667–673, May 2012.

[8] A. J. Buckler, L. Bresolin, N. R. Dunnick, D. C. Sullivan, and Group, "A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging," *Radiology*, vol. 258, no. 3, pp. 906–914, Mar. 2011.

[9] A. J. Buckler, P. D. Mozley, L. Schwartz, N. Petrick, M. McNitt-Gray, C. Fenimore, K. O'Donnell, W. Hayes, H. J. Kim, L. Clarke, and D. Sullivan, "Volumetric CT in lung cancer: an example for the qualification of imaging as a biomarker," *Acad Radiol*, vol. 17, no. 1, pp. 107–115, Jan. 2010.

[10] A. J. Buckler and R. Boellaard, "Standardization of quantitative imaging: the time is right, and 18F-FDG PET/CT is a good place to start," *J. Nucl. Med.*, vol. 52, no. 2, pp. 171–172, Feb. 2011.

[11] G. Murphy, M. Haider, S. Ghai, and B. Sreeharsha, "The expanding role of MRI in prostate cancer," *AJR Am J Roentgenol*, vol. 201, no. 6, pp. 1229–1238, Dec. 2013.

[12] B. Lemasson, C. J. Galbán, J. L. Boes, Y. Li, Y. Zhu, K. A. Heist, T. D. Johnson, T. L. Chenevert, S. Galbán, A. Rehemtulla, and B. D. Ross, "Diffusion-Weighted MRI as a Biomarker of Tumor Radiation Treatment Response Heterogeneity: A Comparative Study of Whole-Volume Histogram Analysis versus Voxel-Based Functional Diffusion Map Analysis," *Transl Oncol*, vol. 6, no. 5, pp. 554–561, 2013.

[13] S. P. Li and A. R. Padhani, "Tumor response assessments with diffusion and perfusion MRI," *J Magn Reson Imaging*, vol. 35, no. 4, pp. 745–763, Apr. 2012.

[14] M.-I. Bittner, N. Wiedenmann, S. Bucher, M. Hentschel, M. Mix, W. A. Weber, and A.-L. Grosu, "Exploratory geographical analysis of hypoxic subvolumes using (18)F-MISO-PET imaging in patients with head and neck cancer in the course of primary chemoradiotherapy," *Radiother Oncol*, vol. 108, no. 3, pp. 511–516, Sep. 2013.

[15] L. Evangelista, C. Mezzato, G. Felloni, and G. Saladini, "Current and future perspectives in diagnostic imaging as a guide to targeted/local therapies in breast cancer recurrence," *Q J Nucl Med Mol Imaging*, vol. 57, no. 4, pp. 367–380, Dec. 2013.

[16] J.-B. Poline, J. L. Breeze, S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, C. Haselgrove, K. G. Helmer, D. B. Keator, D. S. Marcus, R. A. Poldrack, Y. Schwartz, J. Ashburner, and D. N. Kennedy, "Data sharing in neuroimaging research," *Front Neuroinform*, vol. 6, p. 9, 2012.

[17] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *J Digit Imaging*, Jul. 2013.

[18] D. S. Marcus, T. R. Olsen, M. Ramaratnam, and R. L. Buckner, "The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data," *Neuroinformatics*, vol. 5, no. 1, pp. 11–34, 2007.

[19] Y. Schwartz, A. Barbot, B. Thyreau, V. Frouin, G. Varoquaux, A. Siram, D. S. Marcus, and J.-B. Poline, "PyXNAT: XNAT in Python," *Front Neuroinform*, vol. 6, p. 12, 2012.

[20] D. S. Marcus, J. Harwell, T. Olsen, M. Hodge, M. F. Glasser, F. Prior, M. Jenkinson, T. Laumann, S. W. Curtiss, and D. C. Van Essen, "Informatics and data mining tools and strategies for the human connectome project," *Front Neuroinform*, vol. 5, p. 4, 2011.
[21] L. Wang, A. Kogan, D. Cobia, K. Alpert, A. Kolasny, M. I. Miller, and D. Marcus, "Northwestern University Schizophrenia Data and Software Tool (NUSDAST)," *Front Neuroinform*, vol. 7, p. 25, 2013.
[22] C. Haselgrove, J.-B. Poline, and D. N. Kennedy, "A simple tool for neuroimaging data sharing," *Front Neuroinform*, vol. 8, p. 52, 2014.
[23] R. Winslow, J. H. Saltz, I. Foster, J. Carr, Y. Ge, M. Miller, L. Younes, D. Geman, S. Granite, T. Kurc, R. Madduri, T. Rananather, J. Larkin, S. Ardkani, T. Brown, A. Kolasny, K. Reynolds, M. Shipway, and M. Toerper, "The CardioVascular Research Grid (CVRG) Project," *Proceedings of the AMIA Summit on Translational Bioinformatics*, pp. 77–81, 2011.
[24] S. J. Doran, J. d' Arcy, D. J. Collins, R. Andriantsimiavona, M. Orton, D.-M. Koh, and M. O. Leach, "Informatics in radiology: development of a research PACS for analysis of functional imaging data in clinical research and clinical trials," *Radiographics*, vol. 32, no. 7, pp. 2135–2150, Dec. 2012.
[25] S. J. Fouke, T. L. Benzinger, M. Milchenko, P. Lamontagne, J. S. Shimony, M. R. Chicoine, K. M. Rich, A. H. Kim, E. C. Leuthardt, B. Keogh, and D. S. Marcus, "The Comprehensive Neuro-oncology Data Repository (CONDR): A Research Infrastructure to Develop and Validate Imaging Biomarkers," *Neurosurgery*, Oct. 2013.
[26] D. A. Gutman, W. D. Dunn Jr, J. Cobb, R. M. Stoner, J. Kalpathy-Cramer, and B. Erickson, "Web based tools for visualizing imaging data and development of XNATView, a zero footprint image viewer," *Front Neuroinform*, vol. 8, p. 53, 2014.
[27] Y. Gao, S. S. Burns, C. B. Lauzon, A. E. Fong, T. A. James, J. F. Lubar, R. W. Thatcher, D. A. Twillie, M. D. Wirt, M. A. Zola, B. W. Logan, A. W. Anderson, and B. A. Landman, "Integration of XNAT/PACS, DICOM, and Research Software for Automated Multi-modal Image Analysis," *Proc SPIE*, vol. 8674, Mar. 2013.
[28] "Ohloh XNAT Project Page," *Ohloh*. [Online]. Available: www.ohloh.net/p/XNAT. [Accessed: 14-Jun-2014].
[29] R. Jain, L. Poisson, J. Narang, D. Gutman, L. Scarpace, S. N. Hwang, C. Holder, M. Wintermark, R. R. Colen, J. Kirby, J. Freymann, D. J. Brat, C. Jaffe, and T. Mikkelsen, "Genomic mapping and survival prediction in glioblastoma: molecular subclassification strengthened by hemodynamic imaging biomarkers," *Radiology*, vol. 267, no. 1, pp. 212–220, Apr. 2013.
[30] G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, "Ten simple rules for reproducible computational research," *PLoS Comput. Biol.*, vol. 9, no. 10, p. e1003285, Oct. 2013.
[31] C. Stahl, "Experimental psychology: toward reproducible research," *Exp Psychol*, vol. 61, no. 1, pp. 1–2, Jan. 2014.
[32] M. McCormick, X. Liu, J. Jomier, C. Marion, and L. Ibanez, "ITK: enabling reproducible research and open science," *Front Neuroinform*, vol. 8, p. 13, 2014.
[33] J. P. A. Ioannidis and M. J. Khoury, "Improving Validation Practices in 'Omics' Research," *Science*, vol. 334, no. 6060, pp. 1230–1232, Dec. 2011.
[34] R. D. Peng, "Reproducible Research in Computational," *Science*, vol. 334, no. 6060, pp. 1226–1227, Dec. 2011.
[35] F. Prinz, T. Schlange, and K. Asadullah, "Believe it or not: how much can we rely on published data on potential drug targets?," *Nat Rev Drug Discov*, vol. 10, no. 9, p. 712, Sep. 2011.
[36] C. G. Begley and L. M. Ellis, "Drug development: Raise standards for preclinical cancer research," *Nature*, vol. 483, no. 7391, pp. 531–533, Mar. 2012.
[37] S. Langella, S. Hastings, S. Oster, T. Pan, A. Sharma, J. Permar, D. Ervin, B. B. Cambazoglu, T. Kurc, and J. Saltz, "Sharing data and analytical resources securely in a biomedical research Grid environment," *J Am Med Inform Assoc*, vol. 15, no. 3, pp. 363–373, Jun. 2008.
[38] A. Sharma, T. Pan, B. B. Cambazoglu, M. Gurcan, T. Kurc, and J. Saltz, "VirtualPACS--a federating gateway to access remote image data resources over the grid," *J Digit Imaging*, vol. 22, no. 1, pp. 1–10, Mar. 2009.
[39] A. L. Chervenak, T. G. M. van Erp, C. Kesselman, M. D'Arcy, J. Sobell, D. Keator, L. Dahm, J. Murry, M. Law, A. Hasso, J. Ames, F. Macciardi, and S. G. Potkin, "A system architecture for sharing de-identified, research-ready brain scans and health information across clinical imaging centers," *Stud Health Technol Inform*, vol. 175, pp. 19–28, 2012.
[40] K. G. Helmer, J. L. Ambite, J. Ames, R. Ananthakrishnan, G. Burns, A. L. Chervenak, I. Foster, L. Liming, D. Keator, F. Macciardi, R. Madduri, J.-P. Navarro, S. Potkin, B. Rosen, S. Ruffins, R. Schuler, J. A.

Turner, A. Toga, C. Williams, and C. Kesselman, “Enabling collaborative research using the Biomedical Informatics Research Network (BIRN),” *J Am Med Inform Assoc*, vol. 18, no. 4, pp. 416–422, Aug. 2011.
[41] Cancer Imaging Program, “CIP Survey of Biomedical Imaging Archives.” [Online]. Available: https://wiki.nci.nih.gov/display/CIP/CIP+Survey+of+Biomedical+Imaging+Archives. [Accessed: 12-Jun-2013].
[42] “An Assessment of the Impact of the NCI Cancer Biomedical informatics Grid (caBIG).”
[43] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh, “Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python,” *Front Neuroinform*, vol. 5, p. 13, 2011.
[44] I. Dinov, K. Lozev, P. Petrosyan, Z. Liu, P. Eggert, J. Pierce, A. Zamanyan, S. Chakrapani, J. Van Horn, D. S. Parker, R. Magsipoc, K. Leung, B. Gutman, R. Woods, and A. Toga, “Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline,” *PLoS ONE*, vol. 5, no. 9, 2010.
[45] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover, M. R. Pocock, A. Wipat, and P. Li, “Taverna: a tool for the composition and enactment of bioinformatics workflows,” *Bioinformatics*, vol. 20, no. 17, pp. 3045–3054, Nov. 2004.
[46] J. Goecks, A. Nekrutenko, J. Taylor, and $author firstName $author.lastName, “Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences,” *Genome Biology*, vol. 11, no. 8, p. R86, Aug. 2010.
[47] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, “From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors,” *J. Nucl. Med.*, vol. 50 Suppl 1, p. 122S–50S, May 2009.
[48] A. Biggi, A. Gallamini, S. Chauvie, M. Hutchings, L. Kostakoglu, M. Gregianin, M. Meignan, B. Malkowski, M. S. Hofman, and S. F. Barrington, “International validation study for interim PET in ABVD-treated, advanced-stage hodgkin lymphoma: interpretation criteria and concordance rate among reviewers,” *J. Nucl. Med.*, vol. 54, no. 5, pp. 683–690, May 2013.
[49] C. Marcus, A. Ciarallo, A. K. Tahari, E. Mena, W. Koch, R. L. Wahl, A. P. Kiess, H. Kang, and R. M. Subramaniam, “Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes,” *J. Nucl. Med.*, vol. 55, no. 9, pp. 1411–1416, Sep. 2014.
[50] D. L. Rubin, D. Willrett, M. J. O’Connor, C. Hage, C. Kurtz, and D. A. Moreira, “Automated tracking of quantitative assessments of tumor burden in clinical trials,” *Transl Oncol*, vol. 7, no. 1, pp. 23–35, Feb. 2014.
[51] R. T. Fielding and R. N. Taylor, “Principled design of the modern Web architecture,” *ACM Trans. Internet Technol.*, vol. 2, no. 2, pp. 115–150, May 2002.
[52] DICOM WG 27, “Digital Imaging and Communications in Medicine (DICOM) Supplement 161: Web Access to DICOM Persistent Objects by RESTful Services 10 (WADO-RS).” .
[53] DICOM WG 27, “Digital Imaging and Communications in Medicine (DICOM) Supplement 166: Query based on ID for DICOM Objects by REpresentational State Transfer (REST) Services 10 (QIDO-RS).” .
[54] DICOM WG 27, “Digital Imaging and Communications in Medicine (DICOM) Supplement 163: STore Over the Web by REpresentations State Transfer (REST) Services (STOW-RS).” .
[55] S. B. Johnson, G. Whitney, M. McAuliffe, H. Wang, E. McCreedy, L. Rozenblit, and C. C. Evans, “Using global unique identifiers to link autism collections,” *J Am Med Inform Assoc*, vol. 17, no. 6, pp. 689–695, Dec. 2010.
[56] “SAML Specifications.” Mar-2005.
[57] K. A. Archie and D. S. Marcus, “DicomBrowser: software for viewing and modifying DICOM metadata,” *J Digit Imaging*, vol. 25, no. 5, pp. 635–645, Oct. 2012.
[58] R. Farjam, C. I. Tsien, T. S. Lawrence, and Y. Cao, “DCE-MRI defined subvolumes of a brain metastatic lesion by principle component analysis and fuzzy-c-means clustering for response assessment of radiation therapy,” *Med Phys*, vol. 41, no. 1, p. 011708, Jan. 2014.
[59] P. Wang, A. Popovtzer, A. Eisbruch, and Y. Cao, “An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer,” *Med Phys*, vol. 39, no. 8, pp. 5277–5285, Aug. 2012.
[60] S. H. Han, E. Ackerstaff, R. Stoyanova, S. Carlin, W. Huang, J. A. Koutcher, J. K. Kim, G. Cho, G. Jang, and H. Cho, “Gaussian mixture model-based classification of dynamic contrast enhanced MRI data for identifying diverse tumor microenvironments: preliminary results,” *NMR Biomed*, vol. 26, no. 5, pp. 519–532, May 2013.

[61] W. Huang, X. Li, Y. Chen, X. Li, M.-C. Chang, M. J. Oborski, D. I. Malyarenko, M. Muzi, G. H. Jajamovich, A. Fedorov, A. Tudorica, S. N. Gupta, C. M. Laymon, K. I. Marro, H. A. Dyvorne, J. V. Miller, D. P. Barbodiak, T. L. Chenevert, T. E. Yankeelov, J. M. Mountz, P. E. Kinahan, R. Kikinis, B. Taouli, F. Fennessy, and J. Kalpathy-Cramer, "Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge," *Transl Oncol*, vol. 7, no. 1, pp. 153–166, Feb. 2014.

[62] K. Tanderup, S. K. Nielsen, G.-B. Nyvang, E. M. Pedersen, L. Røhl, T. Aagaard, L. Fokdal, and J. C. Lindegaard, "From point A to the sculpted pear: MR image guidance significantly improves tumour dose and sparing of organs at risk in brachytherapy of cervical cancer," *Radiother Oncol*, vol. 94, no. 2, pp. 173–180, Feb. 2010.

[63] L. A. Gloeckler Ries, M. E. Reichman, D. R. Lewis, B. F. Hankey, and B. K. Edwards, "Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program," *Oncologist*, vol. 8, no. 6, pp. 541–552, 2003.

[64] M. M. Grabowski, P. F. Recinos, A. S. Nowacki, J. L. Schroeder, L. Angelov, G. H. Barnett, and M. A. Vogelbaum, "Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma," *J. Neurosurg.*, vol. 121, no. 5, pp. 1115–1123, Nov. 2014.

[65] R. L. Yong, T. Wu, N. Mihatov, M. J. Shen, M. A. Brown, K. A. Zaghloul, G. E. Park, and J. K. Park, "Residual tumor volume and patient survival following reoperation for recurrent glioblastoma," *J. Neurosurg.*, vol. 121, no. 4, pp. 802–809, Oct. 2014.

[66] C. D. Hacker, T. O. Laumann, N. P. Szrama, A. Baldassarre, A. Z. Snyder, E. C. Leuthardt, and M. Corbetta, "Resting state network estimation in individual subjects," *NeuroImage*, vol. 82, pp. 616–633, Nov. 2013.

[67] J. Kalpathy-Cramer, E. R. Gerstner, K. E. Emblem, O. C. Andronesi, and B. Rosen, "Advanced magnetic resonance imaging of the physical processes in human glioblastoma," *Cancer Res.*, vol. 74, no. 17, pp. 4622–4637, Sep. 2014.

[68] B. Menze, M. Reyes, and K. Van Leemput, "The Multimodal Brain TumorImage Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1–1, 2014.