KP Andriole
Wednesday, February 29, 2012

RSNA QIBA–RIC IMAGING DATA WAREHOUSE

Abstract

Discoverability of and accessibility to lasting imaging warehouses with tools and datasets of sufficient quality, size, diversity, and inclusive of relevant metadata are critically essential to the development, validation and deployment of quantitative imaging biomarkers and applications for the detection, diagnosis and management of disease. Relatively modest progress has been made to date.

An *ad hoc* Open Image Archives (OIA) Committee of the Quantitative imaging Biomarkers Alliance (QIBA) was formed to assess what could be done to improve the creation and sustained growth of imaging archives. These efforts have transitioned to a committee of combined QIBA, OIA and Radiology Informatics Committee (RIC) members which was formed to provide a platform of collaboration, to optimize synergy between overlapping areas of expertise and interests; and in particular to support and augment the OIA activities, drafting a plan for potential RSNA involvement for imaging data warehouses going forward.

Four classes of QIBA use cases were defined: A. Comparative Evaluation of Imaging Biomarker Performance versus Gold Standard; B. Public Resource Shared Data (e.g., Image Processing Algorithm Development); C. FDA Approval of Clearance of Imaging Tests; and D. Pharma Clinical Trials with Imaging Biomarkers as Endpoints; for two quantitative imaging biomarker projects: 1) CT volumetric image analysis for management of patients with lung cancer, and 2) quantification of tumor metabolism using FDG-PET standardized uptake value (SUV) image analysis.

Imaging data warehouse needs for each of the QIBA Technical Committee Working Groups (DCE-MRI, FDG-PET, Volumetric-CT, fMRI, and COPD-Asthma) were summarized and common features noted. These included the requirement to accommodate different image and non-image data formats (including and in addition to DICOM, XML, TIFF, NiFTI, etc.) and ; a wide variety of relevant clinical metadata; data input and search and query-retrieve capabilities; image de-identification, data security and user authentication with group sharing; and data output statistics and analytics functions, though not necessarily image display applications.

Existing tools and databases including The Cancer Imaging Archive (TCIA), the National Biomedical Image Archive (NBIA), Laboratory of NeuroImaging (LONI), eXtensible Neuroimaging Archive Toolkit (XNAT), and MIDAS were examined, and current limitations detailed. Of concern is the lack of and/or need for a "trusted third party", the need to promote a culture of sharing perhaps with a reward system or participation, and a business model for long-term sustainability. Additional limitations of existing image data archives include ease-of-use around tool downloads, data uploading, tool configurability and functional enhancements. The need for front-end image and metadata collection tools, security control, advanced search, and back-end data analytics components was also noted.

A way forward is to perform a proof-of-concept implementation using projects from each of the QIBA work groups demonstrable at the RSNA 2012 Annual Meeting, by beginning with one existing image data archive with the most flexible, modularizeable architecture (deemed to be MIDAS). Next steps would be to enhance and/or create an easy-to-use input portal as well as a back-end analytics portal, with the aim of augmenting the architecture to be generalizable to other data archives. The benefits of having the RSNA as convener of the imaging data warehouse are manifold. Direction is needed from the RIC regarding development, implementation and service models (e.g., open-source by committee, industry development or RSNA in-house development similar to MIRC activities). In addition, funding will be sought to support ongoing and future QIBA-RIC imaging data warehouse efforts.

QIBA-RIC Committee Members

Katherine P. Andriole, Ph.D., Chair	Brigham and Women's Hospital Harvard Medical School	RIC, QIBA
Rick Avila, MS	Kitware	QIBA, OIA
Enrique Avilés	Critical Path Institute	QIBA
Andrew J. Buckler, MS	BBMSC	QIBA
David A. Clunie, MBBS	RadPharm	QIBA, IHE
Bradley J. Erickson, MD, PhD	Mayo Clinic Rochester	RIC, QIBA, CaBIG, CTSA
Ross Filice, MD	FDA	QIBA
Dmitry B. Goldgof, PhD	University of South Florida	QIBA
Paul E. Kinahan, PhD	University of Washington	QIBA
Michael McNitt-Gray, PhD	UCLA	QIBA
Daniel L. Rubin, MD, MS	Stanford University	RIC, QIBA, CaBIG, RadLex
Eliot L. Siegel, MD	Baltimore VAH University of Maryland	RIC, QIBA, CaBIG, NCI
Daniel C. Sullivan, MD	Duke University	QIBA Chair
James T. Voyvodic, PhD	Duke University	QIBA
E. Maureen White, MD, MS, MBA	General Electric Healthcare Devices	QIBA
Gudrun Zahlmann, PhD	F. Hoffmann, La Roche	QIBA, OIA

Linda B. Bresolin, PhD, MBA	RSNA Staff
Chris Carr	RSNA Staff
Steve Drew	RSNA Staff
Joe Koudelik	RSNA Staff
Madeleine McCoy	RSNA Staff
Joan McMillen	RSNA Staff
Fiona Miller	RSNA Staff

List of Abbreviations

AIM	Annotation Image Markup, NCI
BIRN	Biomedical Informatics Research Network, NIGMS
CaBIG	Cancer BioInformatics Grid, NCI
CDA	Clinical Document Architecture
CIP	Cancer Imaging Program
CSV	Comma Separated Values, for converting spreadsheet data
CTP	Clinical Trials Protocol of MIRC
CTSA	Clinical and Translational Science Awards, Imaging Work Group
DICOM	Digital Imaging and Communications in Medicine Standard
FDA	Food and Drug Administration
HIT	Health Information Technology
HL7	Health Level Seven interoperability standard for HIT
HTTPS	Hypertext Transfer Protocol Secure, HTTP with SSL
IHE	Integrating the Healthcare Enterprise, RSNA
JPEG	Joint Photographic Experts Group Image File format .jpg
LIDC	Lung Image Database Consortium
LONI	Laboratory of NeuroImaging, UCLA
MIDAS	Server, client and stand-alone tools for data archiving, access, analysis
MIRC	Medical Imaging Resource Center, now called RSNA TFS
NBIA	National Biomedical Image Archive, formerly NCIA
NCIA	National Cancer Image Archive, NCI
NCI	National Cancer Institute

NIfTI	Neuroimaging Informatics Technology Initiative file format for fMRI
NIAMS	National Institute of Arthritis and Musculoskeletal and Skin Diseases
NIGMS	National Institute of General Medicine Sciences
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
OIA	Open Image Archives
QIBA	Quantitative Imaging Biomarkers Alliance, RSNA
PDF	Portable Document Format
QIN	Quantitative Imaging Network
RIC	Radiology Informatics Committee, RSNA
RSNA	Radiological Society of North America
SSL	Secure Sockets Layer, encryption protocol for Internet
SQL	Structured Query Language for relational database management systems
TCIA	The Cancer Imaging Archive, Washington University, formerly NBIA
TFS	RSNA Teaching File System, formerly MIRC
TIFF	Tagged Image File Format, e.g., used in Pathology
TQI	Towards Quantitative Imaging, Reading Room of the Future, RSNA
VTK	Visualization Toolkit
XForms	XML format for specification of data processing model and user interface
XHTML	eXtensible HyperText Markup Language
XML	eXtensible Markup Language document format
XNAT	eXtensible Neuroimaging Archive Toolkit

1. Introduction

A. Need for Imaging and Relevant Data Warehouse Tools

The field of Radiology is increasingly looking toward quantitative imaging methods to provide evidence-based objective measures with which to detect, diagnose and treat disease. The development, validation and deployment of quantitative imaging biomarkers are dependent upon the quality, size, diversity as well as discoverability of and accessibility to image databases with corresponding relevant information including "ground truth" and/or clinical outcomes.

Creating and supporting large open-access image archives with clinical application focus has the potential to accelerate the development and scientific acceptance of quantitative imaging methods. Once an institution or company has developed a new quantitative imaging method, there exists an even greater need to use large collections of radiological images and associated metadata to reach scientific consensus and regulatory approval of the approach.

Despite the growing importance of this objective and the numerous attempts to create open-access image archives, their remains limited availability to high quality, large and diverse image archives with sufficient metadata to develop quantitative imaging applications.

B. History of Committees, Roles and Accomplishments

i. QIBA and QIBA Activities – Open Image Archives (OIA)

Given the significance of open image archives and the relatively modest progress made to date, the Imaging Biomarkers Roundtable, coordinated by the Radiological Society of North America (RSNA) established an *ad hoc* Open Image Archives (OIA) committee of the Quantitative Imaging Biomarkers Alliance (QIBA) to determine what could be done to improve the creation and sustained growth of image archives. The goal of the OIA was to make recommendations that have the potential to significantly improve the number, size and quality of open image archives. This committee reviewed past and current open image archive initiatives; identified the main challenges, incentives and hurdles associated with building and sustaining open image archives; identified key characteristics and requirements of successful image archives based on input from each of the QIBA Technical Committees Working Groups; and prioritized a set of use cases for testing potential solutions recommended by this committee.

ii. Formation of the QIBA-RIC Committee

The formation of a committee consisting of members from both QIBA and the RSNA Radiology Informatics Committee (RIC) was proposed by QIBA Chairman Dr. Daniel Sullivan at the May 11-12, 2011 RIC Retreat, and the committee formed with Dr. Kathy Andriole as Chair June 30, 2011. Committee members include representatives from all workgroups of QIBA, the RIC, the Cancer BioInformatics Grid (CaBIG) of the National Cancer Institute, Clinical and Translational Science Awards (CTSA) Imaging Work Group, the Food and Drug Administration (FDA), the RSNA's Integrating the Healthcare Enterprise (IHE), OIA, and the Quantitative Imaging Network (QIN). The first group call was held July 12, 2011. Subsequent calls and activities have been on-going at least monthly and more frequently as needed, with a face-to-face held September 27, 2011.

The Committee Mission and Statement of Purpose:

- Provide a platform of collaboration between QIBA and RIC
- Optimize synergy between overlapping areas of expertise, interest and activity
- Address informatics needs of the QIBA-RIC community
- Accelerate advancement of industry tools and standards for generating and managing quantitative imaging information

The initial focus of the QIBA-RIC group has been to support and build upon the QIBA OIA activities, with the aim of identifying gaps from an informatics perspective, deciding what is implementable currently, discerning whether it would be beneficial for the RSNA to be involved in next steps, and if so identifying the technical specifications and drafting a plan for potentially developing a prototype for proof-of-concept implementation, as well as exploring future acquisition of funding for subsequent activities.

The first thing the QIBA-RIC did was to replace the term archive with "warehouse" to communicate the full functionality and interactivity of the tool(s) envisioned. Noted among the major challenges are policy issues and creating a community of sharing amongst collaborators.

It is anticipated that the *ad hoc* OIA Committee will transition into the QIBA-RIC Committee going forward.

{Relevant Documents: Appendix 1. Open Image Archive Documents – Statement of Purpose}

2. Use Case Summary

Whether viewed as ends unto themselves or as providing a test-bed capability that may be extended to a series of quantitative imaging biomarkers, two projects were selected as illustrative and for first consideration to utilize open image archives or data warehouses: 1) CT volumetric image analysis for management of patients with lung cancer, and 2) quantification of tumor metabolism using FDG-PET standardized uptake value (SUV) image analysis. The specific aims of such projects are to develop the capability to meet targeted levels of accuracy and reproducibility for the putative biomarker(s), to identify and create mitigation strategies for all meaningful sources of variability in these measurements, and to compare performance to alternative measures such as the diameter-based RECIST criteria.

Four general classes of use cases were defined as follows: A. Comparative Evaluation of Imaging Biomarker Performance versus Gold Standard; B. Public Resource Shared Data (e.g., Image Processing Algorithm Development); C. FDA Approval of Clearance of Imaging Tests; and D. Pharma Clinical Trials with Imaging Biomarkers as Endpoints. From an informatics standpoint, use cases C and D are generalizable to projects requiring restrictive security, authentication and authorization capabilities of the imaging warehouse, whereas use cases A and B, may allow for more open or public use.

{Relevant Documents: Appendix 2. Imaging Biomarker Roundtable Ad Hoc Committee on Open Image Archives (OIA): Example Use Cases for OIA.}

3. QIBA Technical Committee Working Groups Needs Summary

QIBA currently has five active technical committees:

- DCE-MRI: dynamic contrast enhanced – magnetic resonance imaging
- FDG-PET: fluorodeoxyglucose – positron emission tomography
- Volumetric-CT: 3D computed tomography
- fMRI: functional magnetic resonance imaging
- COPD-Asthma: chronic obstructive pulmonary disease – Asthma

Imaging warehouse data needs common to these groups include coverage across the range of imaging acquisition devices and different manufacturers as well as imaging examination acquisition protocols, and a diversity of patient populations with the full spectrum of disease severity. Common features include the need to accommodate different image and non-image data formats (including and in addition to DICOM, XML, TIFF, NiFTI, etc.) and a wide variety of relevant clinical metadata; data input and search and query-retrieve capabilities; image de-identification, data security and user authentication with group sharing; and data output statistics and analytics functions, though not necessarily image display applications.

{Relevant Documents: Appendix 3-10. Open Image Archive Key Attributes; OIA User Requirements Specification; OIA Requirements Project for each Work Group}

4. Overview of Existing Tools

A. Available Archives and Tools

Several medical imaging archives exist to provide the biomedical research community, industry and academia with access to images. Examples include The Cancer Imaging Archive (TCIA) of the Cancer Imaging Program (CIP); the National Biomedical Image Archive (NBIA) of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and of the Cancer BioInformatics Grid (caBIG) of the National Cancer Institute (NCI); the Laboratory of

NeuroImaging (LONI) of the University of California at Los Angeles (UCLA); the eXtensible Neuroimaging Archive Toolkit (XNAT) used by the Biomedical Informatics Research Network (BIRN) and others; several organizational efforts using MIDAS (Kitware Inc). The RSNA Medical Imaging Resource Center (MIRC) in conjunction with the Clinical Trials Protocol (CTP) has also been used as an imaging archive infrastructure for research as well as education.

The National Institute of Standards and Technology (NIST) has instantiated separate image databases based on TCIA, MIDAS and XNAT image archive software applications. Each of these open-source software tools is under active development, is web-based, supports DICOM tags and some metadata formats for input and query, DICOM and HTTPS transfer protocols, role-based security capabilities with public access, and supports Linux and Windows operating systems (OS), though XNAT does not support Mac OS.

An excellent summary is provided on the website of the Cancer Imaging Program. Image Archive Software Solutions. 2011; Available from: https://wiki.nci.nih.gov/display/CIP/CIP+Survey+of+Biomedical+Imaging+Archives.

B. Summary of Limitations

Though a number of imaging archives exist, it is the opinion of the QIBA community and the QIBA-RIC Committee that none are sufficient. Of significant concern is the lack of and/or need for a "trusted third party", the need to promote a culture of sharing perhaps with a reward system for participation, and a business model for long-term sustainability.

Across all QIBA constituents and other interested parties queried including the FDA and NIST, it was expressed that data upload to existing systems was challenging, that downloading and instantiating an archive in individual laboratories was problematic, that overall ease-of-use was lacking, and configuration of tools and enhancement to functionality was extremely difficult. Further, existing structures currently cannot support outcomes data that would be necessary for the purposes of FDA qualification of imaging biomarkers or for algorithm or metric comparison.

Noted ideal attributes, currently lacking in existing imaging archives include standardization of and improved tools for collecting metadata at the front-end and better data input; data validation and data curation tools; more flexible user-authentication, role assignment and control of access authorization, and the ability to provide public and private use models; more advanced searching, data mining and discoverability tools attached to the data stores; and a well-developed back-end analytics component with different data views that can be extracted into models.
{Relevant Documents: Erickson, Marcus and Pan Whitepapers on Imaging Infrastructure for Research, CTSA Imaging Informatics Working Group.}

5. Possible Approaches Forward

A. Start from Scratch – *No*

Starting from scratch and building anew does not seem to be the best option. This would be a duplicative effort, time consuming and costly, and achievement of a functional imaging data warehouse would potentially be delayed.

B. Integration and Enhancement of Existing Tools – *Approach Chosen*

i. Start with One and Generalize to Others

Rather, a way forward may be to perform a proof-of-concept implementation using projects from each of the QIBA work groups that would be demonstrable at the RSNA 2012 Annual Meeting, by beginning with one existing image data archive with the most flexible, modularizeable architecture; keeping in mind that in time this would be generalized and expanded to other data archives (e.g., TCIA, XNAT).

The recommendation is to begin with the open-source MIDAS software as a base for implementing an imaging and relevant data warehouse because it appears to be the most flexible and is highly customizable. The package has an installation process and is the easiest to implement. The web-based tool accommodates multiple image and other file formats and metadata, and has hooks for validation and editing of submitted metadata. MIDAS is written in a modular fashion, as plug-in architecture with an integration framework.

ii. Build Input Portal and ii. Analytics Back-End

Next steps would be to enhance and/or create an easy-to-use input portal as well as a back-end analytics portal, with the aim of augmenting the architecture for use with multiple existing archives and tools. The Input Portal would ideally be able to accept diverse image and metadata file formats, have an easy-to-use installation process, data input validation tools, and an intuitive graphical user interface. A Back-End Analytics Portal enabling more advanced searching, data mining and discoverability tools attached to the data stores, along with a rich analytics component for data output statistics with different visualization methods (though not necessarily image display applications as it is anticipated that researchers will want to have their own display tools) and data views that can be extracted into models would need to be integrated into the

imaging data warehouse.

C. Software Development Team

i. RSNA In-House or ii. Others (including industry) with RSNA as Convener

The benefits of having the RSNA as host of the imaging data warehouse are manifold. The RSNA can act as an impartial trusted third party, advising on policy issues and encouraging data sharing among collaborators, facilitating data solicitation, collection, public archival and use. The RSNA may be in a better position than other archive entities have been to provide a commitment towards continuous development and enhancement. Because of the RSNA's standing with industry partners, Pharma is more likely to support QIBA efforts and submission of data sets. The RSNA could serve as a convener of developers of existing archives to facilitate interoperability and best practices and standards. And the RSNA can also more broadly inform and educate the quantitative imaging community through their large membership base and standing with industry partners.

Direction is needed from the RIC regarding development, implementation and service models (e.g., open-source by committee, industry development or RSNA in-house development similar to MIRC activities).

iii. Grant Funding to Support Activities

Future funding will be sought to support ongoing and future QIBA-RIC imaging data warehouse efforts.

Bibliography

QIBA-RIC Committee Wiki
See http://qibawiki.rsna.org/index.php?title=QIBA/RIC_ctte#Working_Documents and http://qibawiki.rsna.org/index.php?title=QIBA/RIC_ctte#Reference_Materials.

Cancer Imaging Program. Image Archive Software Solutions. 2011; Available from: https://wiki.nci.nih.gov/display/CIP/CIP+Survey+of+Biomedical+Imaging+Archives.

Committee on Open Image Archives (OIA). 2012; Available from: https://qibawiki.rsna.org/index.php?title=Committee_on_Open_Image_Archives.

MIDAS – The Digital Archiving System. 2012; Available from: http://www.kitware.com/products/midas.html/.

National Biomedical Image Archive (NBIA). 2012; Available from; https://gforge.nci.nih.gov/frs/?group_id=312.

National Cancer Institute. The Cancer Biomedical Informatics Grid (CaBIG). 2011; Available from: http://cabig.cancer.gov/.

Quantitative Imaging Biomarkers Alliance (QIBA) of RSNA, 2012; Available from: http://www.rsna.org/Research/QIBA/.

QI-Bench, 2012; Available from: http://www.qi-bench.org/.

Quantitative Imaging Network of NCI. 2012; Available from: https://wiki.nci.nih.gov/display/CIP/QIN.

The Cancer Imaging Archive (TCIA) of NCI. 2012; Available from: http://cancerimagingarchive.net/.

XNAT. 2012; Available from: http://www.xnat.org/.

MacKenzie-Graham, Van Horn JD, Woods RP, Crawford KL, and Toga AW. Provenance in Neuroimaging. NeuroImage, 42 (2008) 178–195.

Sagotsky JA, Zhang L, Wang Z, Martin S, and Deisboeck TS. Live Sciences and the Web: A New Era for Collaboration. Molecular Systems Biology 4:201; doi:10.1038/msb.2008.39, July 1, 2008.

MacKenzie-Graham, Van Horn JD, Woods RP, Crawford KL, and Toga AW. Provenance in Neuroimaging. NeuroImage, 42 (2008) 178–195.

Zhao J, Miles A, Klyne G, and Shotton D. Linked Data and Provenance in Biological Data Webs. Briefings in Bioinformatics.10(2):139-152, doi:10.1093/bib/bbn044, December 6, 2008.

Van Horn JD and Toga AW. Is It Time To Re-prioritize Neuroimaging Databases and Digital Repositories? NeuroImage, 47 (2009) 1720-1734.

Goecks J, Nekrutenko A, Taylor J. Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible and Transparent Computational Research in the Life Sciences. Genome Biology 2010, 11:R86 http://genomebiology.com/2010/11/8/R86.

Erickson BJ, Pan T, Marcus DJ. Whitepaper on Imaging Infrastructure for Research Part 1: General Workflow Considerations, CTSA Imaging Informatics Working Group November 2011.

Marcus DJ, Erickson BJ, Pan T. Whitepaper on Imaging Infrastructure for Research Part 2: Data Management Practices, CTSA Imaging Informatics Working Group November 2011.

Pan T, Erickson BJ, Marcus DJ. Whitepaper on Imaging Infrastructure for Research Part 3: Security and Privacy, CTSA Imaging Informatics Working Group November 2011.

Appendix

1. Imaging Biomarker Roundtable Ad Hoc Committee on Open Image Archives (OIA): Statement of Purpose. April 21, 2010.

2. Imaging Biomarker Roundtable Ad Hoc Committee on Open Image Archives (OIA): Example Use Cases for OIA. V1.0, November, 2010.

3. Open Image Archives Key Attributes. May 18, 2011.

4. QIBA Open Image Archives: User Requirements Specification. May 18, 2011.

5. RSNA QIBA Open Image Archives Requirements Project. FDG-PET Work Group, October 21, 2011.

6. RSNA QIBA Open Image Archives Requirements Project. DCE-MRI Work Group, October 26, 2011.

7. RSNA QIBA Open Image Archives Requirements Project. Volumetric CT Work Group, October 31, 2011.

8. RSNA QIBA Open Image Archives Requirements Project. fMRI Work Group, November 2, 2011.

9. RSNA QIBA Open Image Archives Requirements Project. COPD-Asthma Work Group, November 2, 2011.

10. RSNA QIBA DCE-MRI Technical Committee: Data Warehouse Use Case and Business Applications, December 2011.

6. Imaging Data Warehouse Proposal

Executive Summary

What is proposed is a pilot program to implement an open image data warehouse hosted by the RSNA based on an existing code-base (MIDAS) and to augment it with support for forms-based and bulk loading, storage, retrieval and mining of structured and unstructured related non-image data (covariates, clinical, pathology results, protocol descriptions, etc.), including a pattern-based conversion ability to extract from a multitude of alternative tagging, markup, annotation and quantitative encoding formats, and to make available the information to standard SQL database mining, reporting and statistical analysis tools.

Background

The joint taskforce of the Quantitative Imaging Biomarker Alliance (QIBA) and the RSNA Radiology Informatics Committee (RIC) have consulted broadly with the active QIBA Sub-committees, the Imaging Biomarker Roundtable Committee on Open Image Archives (OIA), and reviewed white papers from the Clinical and Translational Science Awards (CTSA) Imaging Informatics Sub-committee. Currently deployed open-source open image archives such as NBIA, MIDAS, XNAT and LONI have been reviewed, as have informal practices such as physical media exchange (e.g., DVD, hard drives) and the use of ftp and similar web-based drop box solutions.

Limitations of Existing Solutions

Consistent patterns and clear gaps have been identified with respect to both operational use cases and requirements for long-term secondary re-use.

1. The trustworthiness, sustainability and longevity of individual research group and/or federally funded activities, whether localized within institutions or institutes, out-sourced or centralized, is open to question in the face of significant economic and deferral government financial uncertainty. RSNA emerges as a potentially reliable long-term organization with a continuing revenue stream and mandate to serve its community, and a mission to promote the development and adoption of quantitative imaging.
2. Existing OIA solutions support the submission and retrieval of images in standard formats (like DICOM), but are in many cases burdened by unwieldy submission mechanisms or administrative procedures, with both technical and procedural barriers that particularly hamper small scale, informal or unfunded experiments.
3. The ability to query OIA content by all available meta-data, or to “mine” the content of different types of data or collections within an OIA, or federate such queries across different OIA implementations is absent or limited.
4. The ability to reliably associate structured and unstructured information about the images, the subject of the images, or the conduct of the experiment or protocol is very limited, if

present at all. Formal and informal standards or conventions exist for encoding such information (e.g., XHTML forms, XForms, CSV files, HL7 V.2 messages, HL7 Clinical Document Architecture (CDA) documents, proprietary XML schema documents, DICOM Structured Reports, PDF files), and for automating capture workflow (IHE Retrieve Form for Data Capture), but these are not widely implemented in existing OIAs, particularly with respect to associating these with the images, or making the contained metadata available for query and data mining.

5. The ability to "tag" ("annotate" in the most crude sense) a particular image or set of images, as possessing a particular feature is limited, if present at all, as is the ability to import or export a collection with such tags in either a standard or proprietary form.
6. The ability to import, export and index for searching (with other metadata) categorical or quantitative information related to or extracted from image content ("annotations", "markup", "regions of interest"), whether generated by a human or an automated process, is limited, and is particularly confounded by the need to handle a variety of existing commonly implemented industrial standards (e.g., DICOM Structured Reporting (SR), DICOM Radiotherapy (RT) Structure Sets, DICOM Presentation States), as well as research formats specific to certain groups (e.g., NCI Annotation Image Markup (AIM)), domain-specific activities (e.g., Neuroimaging Informatics Technology Initiative (NIfTI)), specific projects (e.g., Lung Image Database Consortium (LIDC) XML) and specific platforms (e.g., Visualization Toolkit (VTK)) or tools (e.g., 3D Slicer, FreeSurfer), not to mention commercial proprietary formats (or extensions to standard formats), whether documented or "reverse-engineerable".

Goal

The goal of the proposed pilot project is to prove the concept of a sustainable OIA or imaging data warehouse that has sufficient submission and querying capabilities in order to support both operational needs for basic research into quantitative imaging, and secondary re-use of acquired images and metadata.

Proposal

Realistically an OIA solution requires flexibility and adaptability to accommodate the use of "best of breed" tools and formats chosen to solve a particular problem expeditiously. Accordingly, a practical OIA cannot dictate the choice of standard used for submission and retrieval of images or associated information.

Storage and distribution is essentially a solved problem, technologically. The preference of the committee based on review of existing open source solutions is to base work on the MIDAS product from KitWare. This tool in its current form is agnostic to the format of the content stored, and allows indexing of the content by category.

What is proposed is a pilot project to first deploy an instance of MIDAS from its source code and dependencies either on RSNA hardware, or by virtual machines at a co-location or cloud-sourced facility that is under RSNA control.

Then, it is proposed to extend this instance, either with additional interfaced components or by modification of the source code, using a combination of encoding-format-specific toolkit for parsing (e.g., into XML), generic pattern matching (e.g., using XSL-T) and database insertion (using standard SQL without proprietary extensions) to add the ability to:

1. Query for database-indexed metadata extracted from information contained in image data headers (e.g., to extract and index the slice thickness or reconstruction kernel or similar), using a combination of image-format-specific toolkits for parsing (e.g., into XML) and generic pattern matching (e.g., using XSL-T), and to implement instances of this (from DICOM images, using both standard and private data elements, and from NIfTI-1 images)
2. Directly associate other submitted data (such as spreadsheets of covariates like lab test values or histo-pathological diagnosis) with the image datasets, whether by direct linkage or shared identifiers of subjects and visit or date (especially if de-identified), and to implement several instances of this (from a CSV spreadsheet with a header row containing common data elements, and shared subject and visit identification columns, and an XForms form data instance consistent with the Integrating the Health Care Enterprise (IHE) Retrieve Form for Data Capture (RFD) profile as a Form Archiver actor).
3. Populate a template-driven form via a web-browser accessible user interface to submit such form information as described above (e.g., as an IHE RFD Form Filler).
4. Receive and store image-content related data (such as annotations, markup and regions of interest or other quantitative derived information), associate it with the relevant image datasets, extract database-indexed meta-data for querying as above, and to implement three instances of this (from DICOM SR, DICOM RT Structure Sets, and NCI AIM version 3.0).

In such cases where images, documents and other files are submitted, the original submitted form shall be retained and retrievable; the objective is to provide an index of extracted information that may be queried, not to “convert” file formats per se (in this phase).

This pilot phase does NOT propose the development of a specific query user interface, but rather that off-the-shelf existing database and data mining query tools (that can take advantage of access to a standard SQL database) can be used to perform the query. The pilot phase does require that such accessibility be demonstrated, however, from an open source database-agnostic data mining tool (e.g., RapidMiner), and from an open source database-agnostic statistical

package (e.g., "R" DBI package).

Further requirements are that the database of indexed meta-data:

1. Shall have an extensible schema to allow the addition of new patterns of extraction by the user, without requiring a database rebuild (i.e., it will be dynamically extensible); and it is anticipated that users will contribute such tools to the community activities.
2. A means shall be provided to re-index existing files when updated patterns that pertain to the same content type are provided by the user.

Out of Scope

It is not the intent to attempt to overload the concept of an OIA with a complete imaging clinical trials infrastructure implementation within the scope of this pilot project. Such features as de-identification, management of pseudonymous identifiers, quality control of incoming data and matching against protocol requirements, tracking of submitted data against expected data, querying of sites for missing or bad data, etc., are out of scope at this time.

It is also not a goal to develop commercial or regulatory compliant grade tools, for example, to provide for double-data entry of paper Case Report Form (CRF), to meet 21 CFR Part 11 requirements with respect to electronic records (including electronic signatures and audit trails), or to document the design, development testing, validation, and deployment process with the degree of rigor necessary for commercial clinical trials of drugs, biologics or devices.

It is not the goal to include in the OIA an image "viewer", since the expectation is that images (and associated information) will be retrieved and viewed locally; further, a viewer capable of rendering images of every modality with any possible annotation format is a non-trivial problem. Future extensions might include a generic plugin viewer capability (such as via the DICOM WG 23 API, for instance). Further, the extensibility to bulk data other than images (e.g., MR raw data), militates against a built-in generic rendering capability.

Evaluation and Criteria for Success

Each of the QIBA specialty Technical Committee Work Groups (CT Volumetrics, DCE-MRI, PET, fMRI and COPD) will adapt (one or a sufficient subset) of their experimental datasets and experimental designs to test that the pilot project deliverables by:

1. Loading and indexing their phantom or in vivo bulk image data, in one of the supported formats (expected to be DICOM for most except perhaps fMRI)
2. Loading and indexing their covariate, subject description and experimental design information
3. Loading and indexing their human or machine generated derived information (e.g.,

ROI values such as size or density with image or 3D-relative coordinates) in whatever format they were created or acquired, and if necessary design a "pattern" to add to the pilot project implementation to extract specific meta-data for indexing

4. Demonstrating the use of a query, mining or statistical tool to successfully access the indexed meta-data.

Said sub-committees will then make recommendations to RSNA as to the utility and sufficiency of the pilot solution, in order to guide RSNA in assessing the value of sustaining the project, and any future development phases.

QIBA Workgroup Project Datasets for use in the Proof-of-Concept Implementation

- Volumetric computed tomography phantom data for lung nodule quantitation
- FDG-PET digital reference object data for validation of SUV calculations and ROIs
- fMRI data collections from Duke and Medical College of Wisconsin in DICOM and NIfTI file formats (from different scanner manufacturers) and associated brain-function metadata stored in binary and text file formats to test quantitative reproducibility.

Future Design Considerations

To the extent that the MIDAS proof-of-concept implementation is successful, a similar interface is expected to be required for the NBIA, given its installed base and extensive existing content. The same might also be said for XNAT, and even teaching file solutions like MIRC using the Clinical Trials Protocol, which have some utility in their own right for research. To this end, the design and development in the pilot phase should take this factor into consideration, even to the extent that the project might be built as a separate tool and database that indexes MIDAS, rather than is directly incorporated into it.

Security Considerations

MIDAS contains access-control capabilities, and to the extent possible, these should be reflected in any additional content submission interface and the SQL-database access interface, such that the same credentials, authentication and access controls are applicable to both.

References

See http://qibawiki.rsna.org/index.php?title=QIBA/RIC_ctte#Working_Documents and http://qibawiki.rsna.org/index.php?title=QIBA/RIC_ctte#Reference_Materials.

7. Action Items

1. Presentation of activities and proposed plan to the RSNA RIC on February 14, 2012.

2. Get direction regarding development, implementation and service model:
   - open-source by committee
   - industry development or
   - RSNA in-house development similar to MIRC activities.

3. Write one-page proposal/report for presentation to the RSNA Board of Directors (in the suggested format below).

4. Seek funding to support ongoing and future QIBA-RIC imaging data warehouse efforts.

DRAFT Report to the Board of Directors

Ronald L. Arenson, MD
Liaison for Annual Meeting and Technology
March 2012

QIBA-RIC Imaging Data Warehouse

Action Requested:
Approval for the QIBA-RIC Committee to move forward with a proof-of-concept implementation of an RSNA-supported imaging data warehouse, used by the QIBA Technical Work Groups, with anticipated demonstration at the RSNA 2012 Annual Meeting.

Goal: ***{Whatever Board Goals fit here}***
#1 RSNA will advance the radiological sciences and foster the development of new technologies.
1.5 Promote translation of radiologic science and quantification to clinical care.
#4: RSNA will facilitate informatics strategies to improve the efficiency and effectiveness of healthcare.
4.2 Develop and promote informatics performance solutions.

Issue: ***{Need RIC advice here}***

Background:
The field of Radiology is increasingly looking toward quantitative imaging methods to provide evidence-based objective measures with which to detect, diagnose and treat disease. The development, validation and deployment of quantitative imaging biomarkers are dependent upon the quality, size, diversity as well as discoverability of and accessibility to image databases with corresponding relevant information including "ground truth" and/or clinical outcomes. Despite the growing importance of this objective and the numerous attempts to create open-access image archives, their remains limited availability to high quality, large and diverse image archives with sufficient metadata to develop quantitative imaging applications.

Fiscal Note: ***{Need RIC advice here}***