

Open Source Software Projects of the caBIG™ In Vivo Imaging Workspace Software Special Interest Group

Fred W. Prior,¹ Bradley J. Erickson,² and Lawrence Tarbox¹

The Cancer Bioinformatics Grid (caBIG™) program was created by the National Cancer Institute to facilitate sharing of IT infrastructure, data, and applications among the National Cancer Institute-sponsored cancer research centers. The program was launched in February 2004 and now links more than 50 cancer centers. In April 2005, the In Vivo Imaging Workspace was added to promote the use of imaging in cancer clinical trials. At the inaugural meeting, four special interest groups (SIGs) were established. The Software SIG was charged with identifying projects that focus on open-source software for image visualization and analysis. To date, two projects have been defined by the Software SIG. The eXtensible Imaging Platform project has produced a rapid application development environment that researchers may use to create targeted workflows customized for specific research projects. The Algorithm Validation Tools project will provide a set of tools and data structures that will be used to capture measurement information and associated needed to allow a gold standard to be defined for the given database against which change analysis algorithms can be tested. Through these and future efforts, the caBIG™ In Vivo Imaging Workspace Software SIG endeavors to advance imaging informatics and provide new open-source software tools to advance cancer research.

KEY WORDS: Open source, digital imaging and communications in medicine (DICOM), grid computing, image analysis, imaging informatics, caBIG, XIP, AVT

BACKGROUND

The Cancer Bioinformatics Grid (caBIG™) is a program of the National Cancer Institute (NCI) to create a state-of-the-art informatics infrastructure linking NIH-supported cancer centers. caBIG™ provides a common infrastructure, vocabularies, and tools, so that each individual institution connects to data and resources in a way

that was never before possible—catalyzing discovery and advancing the practice of oncology.¹ All caBIG™ projects are based on four fundamental principles:²

- Open source: Standards and software developed within the caBIG™ initiative are licensed as open source and are thus available for implementation, review, or modification without a licensing fee. The caBIG™ open-source license allows vendors to incorporate caBIG™ software into commercial products.
- Open access: The standards and software developed within caBIG™ are freely available for use by health care organizations, biomedical researchers, and vendors who support this community, and anyone is welcome to contribute.
- Open development: caBIG™ is committed to open communication and broad collaboration. Planning for caBIG™ standards and software

¹From the Electronic Radiology Laboratory, Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO, USA.

²From the Department of Radiology, Mayo Clinic, Rochester, MN, USA.

Correspondence to: Fred W. Prior, Electronic Radiology Laboratory, Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO, USA; tel: +1-314-7470331; fax: +1-314-3626971; e-mail: priorf@mir.wustl.edu

Copyright © 2007 by Society for Imaging Informatics in Medicine

Online publication 11 September 2007

doi: 10.1007/s10278-007-9061-4

development is carried out in open meetings, and comments are solicited from all interested participants. Development projects are assigned to particular participants but are carried out iteratively, with multiple opportunities for review and comment by the caBIG™ community at large.

- Federation: caBIG™ software and standards enable local sites, such as cancer centers, to use resources contributed by others and to share computing or data resources with the cancer community at large. Federation implies that these individual resources remain under the control of the local sites but are aggregated for use by all participants as an integrated research tool.

CaBIG™ In Vivo Imaging Workspace

The In Vivo Imaging Workspace is one of four domain workspaces that have thus far been formed within the caBIG™ program to respond to specific needs identified by the NCI-sponsored cancer research centers. The goal of the In Vivo Imaging Workspace is to advance the field of imaging informatics by creating, optimizing, and validating software tools and helping to extract meaning from in vivo imaging data to improve outcomes for patients with cancer.³

To achieve this goal, the workspace created four special interest groups, or SIGs, each focused on a specific area of imaging informatics. The Software SIG was given the charter to create and adapt open source software tools to promote and enhance the use of imaging in cancer research. The SIG chose to focus on tools for image acquisition, management, and analysis for use in clinical trials and tools for enhancing lesion detection, characterization, and change determination.

Like the Imaging Workspace, membership in the Software SIG is open to all interested parties and includes representatives from industry, academia, and government. The Software SIG does not conduct research or development itself but is tasked with identifying and prioritizing projects and defining requirements for these projects. Once a research and development team has been selected via a competitive process conducted by the NCI, the Software SIG helps to guide the development process by defining use cases and clarifying

requirements. It also serves to track the process by reviewing deliverables and progress and regularly reporting to the NCI.

METHODS

The caBIG™ community has created a well defined development process and associated tools. All projects are based on a model-driven development architecture with consistency maintained by an architecture workspace that provides active mentoring to development teams. The caCORE offers tools for building and managing common data elements. NCI maintains GForge and caMP web sites for project management support. All grid-facing interfaces are carefully modeled and the semantics of all data elements are clearly defined and maintained in a common data element repository.

The In Vivo Imaging Workspace Software SIG participates in regular face-to-face workspace meetings and weekly SIG teleconferences. During these open sessions, the needs of the NCI-funded cancer research centers are assessed and specific projects are defined to address these needs. To date, two sets of project proposals have been identified. These are listed in Table 1. The SIG prioritizes the project list and selects the top two or three projects for further definition. A project justification document is created for each of the highest-priority projects in which the project objectives and scope are outlined, and a justification is given for the relative priority of the specified project. A similar process is undertaken by the other SIGs within the workspace. The complete set of project justifications is evaluated and prioritized by the NCI and a subset is selected.

Once a project proposal is approved, the SIG actively solicits and documents requirements. This is commonly accomplished using a Wiki,⁴ which allows the caBIG™ community as a whole to freely propose and modify requirements. Through a series of teleconferences, the Software SIG members reach consensus on a final set of requirements, which are then edited into a formal document and submitted to the entire In Vivo Imaging Workspace, who then recommend projects to the NCI to be considered for funding through the creation of a request for proposals. If

Table 1. Potential Projects Identified by the Software SIG

Projects Identified December 2005: Phase 1	Projects Identified July 2006: Phase 2
Software support for multicenter distributed reader studies and trials	Tools for forming "gold standard" (statistical model of truth)
Database resources and software for change analysis	Database and registry for image analysis software of specific utility in cancer research
Correlation between radiology and pathology imaging	Quality control, quality assurance and curation tools
Algorithm performance characterization tools	Provenance tracking tools for 21 CFR part 11 compliance
Identify and solicit validation and curation tools	Tools for validating and characterizing algorithms
Image acquisition and deidentification software	Image stack and volume synchronization (4D) visualization tools
Software database and repository for algorithm validation	Correlation between radiology and pathology imaging
Consumer reviews of medical imaging software	Ontology of experimental designs and experiment description tools

the NCI approves the project, the caBIG™ primary contractor (Booz, Allen, Hamilton) manages the solicitation of proposals, the selection of the subcontracting team who will carry out the project, and the monitoring of the project's progress, including managing the financial aspects on behalf of the NCI. Figure 1 summarizes the process used by the caBIG™ Software SIG.

To date, two projects proposed by the Software SIG have been selected for implementation. In phase 1, NCI selected the first project on the phase 1 list as the highest-priority task for the SIG and In Vivo Imaging Workspace. For phase 2, a project relating to change analysis was created by combining the first and fifth items on the potential phase 2 projects. Requirement definition for both projects has been completed. The phase 1 project is nearing its first year of development. The phase 2 project is just getting underway.

RESULTS

Software Support for Multicenter Distributed Reader Studies and Trials

The most pressing problem identified by the Software SIG in phase 1 was the need for an extensible open-source platform to support image analysis and visualization. Increasingly, clinical trials rely on imaging-based biomarkers, which in turn rely on precise and repeatable measurements of image features. The requirements for quantitative image analysis to support both research and clinical efforts to detect and diagnose cancer and to track a patient's response to therapy exceed the capabilities of existing commercial imaging products. The Software SIG realized that what was needed was not a new type of imaging workstation but rather a rapid application development

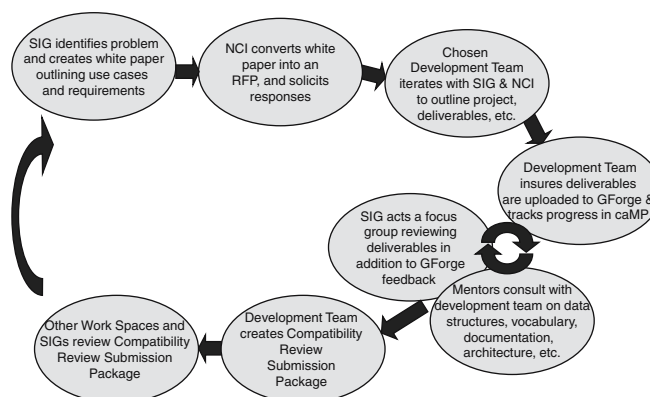


Fig 1. The caBIG™ project management process used to create XIP ensures compliance with caBIG™ architecture and design principles.

environment for creating and optimizing new analysis and visualization tools that could be customized for specific tasks and workflows. The developed tools could then be deployed in both the research lab and the clinical reading room, ideally on any work station.

The eXtensible Imaging Platform (XIP) is an open-source environment for rapidly developing medical imaging applications from an extensible set of modular elements. This platform makes it easier and less expensive to access specific post-processing applications at multiple sites; simplifying clinical trials; and, most importantly, increasing the uniformity of imaging and analysis. Imaging applications developed by research groups are more easily accessible within the clinical operating environment, simplifying workflows and speeding data processing and analysis. Once validated, the software should be readily transitioned into products through streamlined Federal Drug Administration approval processes due to the reuse of already approved libraries and open-source development processes.

XIP supports the rapid development of “plug-in” applications for image analysis and visualization. Applications built by these tools utilize a host-system-independent interface being standardized by the Digital Imaging and Communication in Medicine (DICOM) Working Group 23 (WG-23).⁵ The DICOM WG-23 interface provides a mechanism by which any host supporting a particular profile or version of the interface may control (eg, start, stop, pause, obtain status from) and exchange data with any application that supports the same profile or version of the interface. Through this means, an application programmer need only create one version of an application, which then can be run without significant change on a variety of systems, such as commercial or open-source medical imaging work stations. Such host independence facilitates translational research across multiple centers by allowing the same application to be deployed into a wide variety of settings, both research-oriented and clinically oriented. The interface as defined by DICOM WG-23 also includes abstract models for the data being exchanged, making it possible for an application to interact with existing data and produce new data without regard to how or where the data is actually stored or in what format the data is stored in. To avoid losing the full richness of the underlying

data formats, the DICOM WG-23 interfaces used by XIP also include means to access the native data directly, either through a native parsed model or by accessing files directly.

The XIP package includes reference work station implementations that can utilize DICOM services as well as caGRID data and analytic services⁶ to support hosted XIP applications. The XIP application-building libraries are based on Open InventorTM⁷ classes, with extensions to support medical imaging applications (eg, lesion detection, multidimensional visualization, registration, and fusion). These extensions include both custom-built objects and automatically generated wrapper objects for commonly used toolkits, such as the Insight Took Kit (ITK)⁸ (for segmentation, registration, and image analysis) and the Visualization Tool Kit (VTK)⁹ (for display of multi-dimensional data sets).

Figure 2 illustrates the components of the XIP package, which include:

1. XIP Application Builder—an integrated development environment that allows XIP applications to be constructed by graphically linking modules.
2. XIP libraries—sets of host-independent Open InventorTM objects that may be used to build XIP applications. XIP libraries may be auto-generated from existing class libraries (eg, ITK and VTK) or custom-built from new or existing code. The reference XIP implementation includes the base Open Inventor classes, classes autogenerated from ITK and VTK, and a set of custom XIP classes to support image display, measurements, graphical overlays, and the importing and exporting of DICOM and other data sets through the DICOM WG-23 interface.
3. XIP Reference Implementation, which consists of:
 - (a) XIP Host, which provides the infrastructure in which XIP or DICOM WG-23 applications run. The Host provides data and services to XIP applications (including caGRID interactions and security) and supports the DICOM WG-23 Application Hosting Interface Standard.
 - (b) XIP applications, which operate in the virtual environment provided by the Application Hosting Interface and implement the processing logic to analyze and visualize

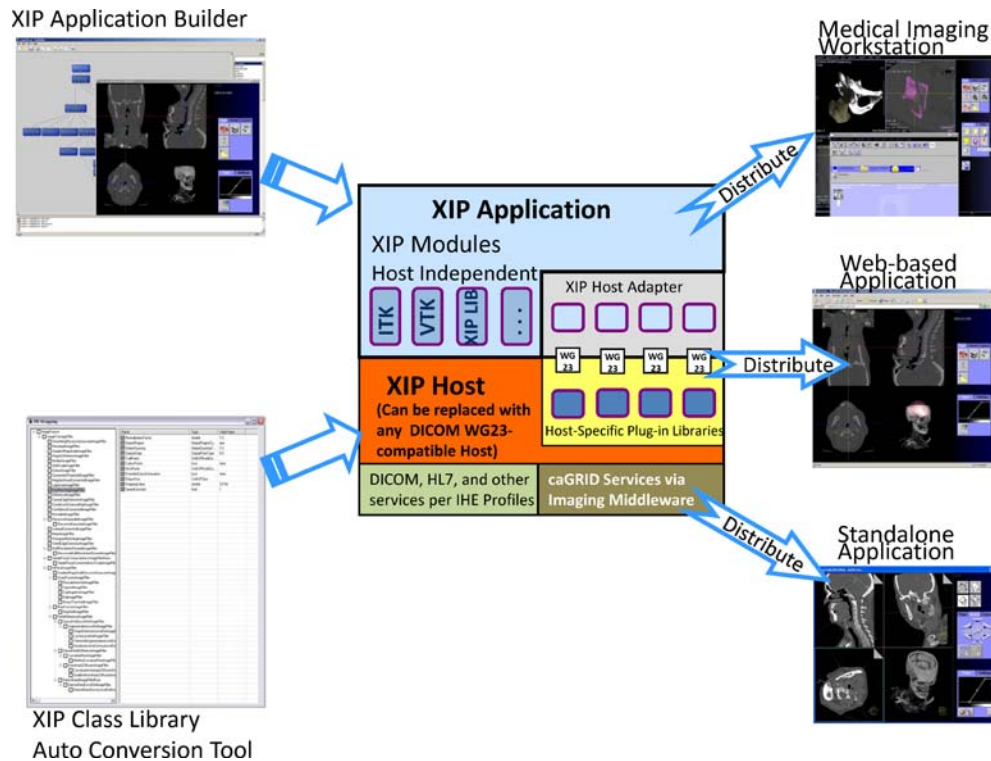


Fig 2. The XIP Application Builder is used to create XIP applications, which may run on any host that supports the DICOM Application Hosting Interface, and it may be deployed in a variety of configurations.

medical images and information. The reference applications delivered in the first year support a hypothetical clinical trial with multiple data collection centers, distributed analysis to characterize tumor size (eg, RECIST¹⁰ criteria), and adjudication of any conflicting information reported in the distributed analyses to create a final, consolidated result.

The initial phase of XIP development reached completion in September 2007. The open source software and project documentation are available from the caBIG™ GForge code management system and repository (<http://gforge.nci.nih.gov/plugins/scm cvs/cvsweb.php/xip/Developer/?cvsroot=xip>).

Change Analysis

To assess disease progression and response to treatment it is necessary to precisely and accurately detect, quantify, and characterize change in a

lesion or other image feature between imaging examinations obtained at different time points.¹¹ The traditional way to measure response has been to use the physical dimensions of the mass. Physiological and metabolic changes can also be measured using, for example, F-18 fluorodeoxyglucose positron emission tomography and may indicate response or nonresponse to therapy. Some modern therapy agents result in little or no change in size (at least initially), but decrease the growth rate or stop growth from continuing. Thus, a change in growth rate may be as important as a change in size. Greater precision in measurement may reduce cost of clinical trials—fewer subjects, shorter trial period. Traditional measure and subtract methods are often not precise. This may be because the boundaries of tumors are usually not sharply defined. Even for relatively well-defined lesions, variations in human measurements can be large. To help cope with defining appropriate standards for analyzing and tracking change, the SIG created the Algorithm Validation Tools (AVT) project.

Given the increasing use of imaging-based biomarkers, it is critical to determine the accuracy and reliability of the measurement process and of algorithms that attempt to detect change. Unfortunately, the problem of establishing “ground truth” is very difficult. The goal of the AVT project is to provide tools that may be used by efforts, such as the Reference Image Database to Evaluate Response (RIDER) and the Lung Imaging Database Consortium (LIDC),¹² to produce both validation data sets and processes for establishing a gold standard of truth for these data sets.

The AVT package is envisioned to include three major components:

1. An image analysis component that displays images and permits features to be identified and marked. This would be designed as an XIP application and incorporate annotation and mark-up functions to capture measurements and metadata relating to the measurements (eg, person who made the measurements, how they were made).
2. An assessment database schema for storing the observations and measurements produced by the image analysis component (or equivalent functions).
3. Tools to extract measurements placed in the assessment database and compute their variability. This component may be designed as a grid analytic service⁶ or as an XIP application and should encapsulate the open source R¹³ statistical package.

The Software SIG has completed its work on the AVT requirements and has delivered its report to NCI. This phase 2 project was launched in the summer of 2007 and will progress through 2008.

DISCUSSION

The projects of caBIG™ and, in particular, those of the In Vivo Imaging Workspace represent an experiment in team development of open-source software. A community identifies needs, defines requirements, and recruits an open/multi-institutional development team that creates an initial implementation. The process does not stop here, however. Tools like XIP are specifically designed to make new development and new

extensions easy for the research community as a whole, whereas the open-source software that comprises XIP can itself evolve through community contributions. It is hoped that groups such as the Software SIG will continue to support and develop applications such as XIP and AVT and to promote their adoption and evolution by the cancer research community.

The AVT package is specifically focused on the needs of the RIDER and LIDC programs. This choice was made as a way to limit the scope of the project and have a user community that provides requirements and guidance to direct the development to an immediately useful end. This is a different type of experiment in which scientific programs at NCI are coupled with caBIG™ development activities with the goal of productive synergy.

CONCLUSION

The success of caBIG™ will ultimately be measured by the impact this information technology initiative has on cancer research. The In Vivo Imaging Workspace and its SIGs will measure success more simply as the rate of adoption of tools such as XIP both within the cancer research community and in the imaging science world in general. Through the collaboration of caBIG™ and DICOM WG-23, it is hoped that a bridge can be successfully built between the research community and commercial vendors that support clinical imaging.

REFERENCES

1. Buetow K: Cyberinfrastructure empowering a “Third Way” in biomedical research. *Science* 308:821–824, 2005
2. National Cancer Institute: caBIG™ Primer an Introduction to caBIG™. <https://cabig.nci.nih.gov/overview/cabig-primer/>, 2006
3. National Cancer Institute: Introduction to Imaging in the caBIG™ Program. In Vivo Imaging (IMAG) Newcomer Information. https://cabig.nci.nih.gov/workspaces/Imaging/imaging_newcomer, 2007
4. Leuf B, Cunningham W: *The Wiki Way: Collaboration and Sharing on the Internet*. Reading: Addison Wesley, 2001
5. DICOM Committee: DICOM Digital Imaging and Communications in Medicine Strategic Document. <http://medical.nema.org/dicom/geninfo/Strategy.pdf>, 2007

6. Saltz J, Oster S, Hastings S, et al: caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics* 22:1910–1916, 2006
7. Wernecke J: *The Inventor Mentor: Programming Object-oriented 3D Graphics with Open Inventor, Release 2 (OpenGL)*. Boston: Addison-Wesley, 1994
8. Ibanex L, Schroeder W: *The ITK Software Guide 2.4*. Clifton Park: Kitware, 2005
9. Schroeder W, Martin K, Bill Lorensen B: *The Visualization Toolkit an Object-Oriented Approach to 3D Graphics*, 4th edition. Clifton Park: Kitware, 2006
10. Therasse P, Arbutk SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 92:205–216, 2000
11. Patriarche J, Erickson B: Change detection & characterization: A new tool for imaging informatics and cancer research. *Cancer Inform* 1:1–11, 2007
12. National Institute of Biomedical Imaging and Bioengineering: Image and Clinical Data References. <http://www.nibib.nih.gov/Research/Resources/ImageClinData#RIDER>, 2007
13. Dalgaard P: *Introductory statistics with R*. Berlin Heidelberg New York: Springer, 2002