



Published in final edited form as:

*Conf Proc IEEE Eng Med Biol Soc.* 2010 ; 2010: 3475–3478. doi:10.1109/IEMBS.2010.5627838.

## Regulatory Compliance Requirements for an Open Source Electronic Image Trial Management System

**Colin Rhodes,**

VirtualScopics Inc., Rochester, NY 14625 USA

**Steve Moore [Member, IEEE],**

Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110 USA

**Ken Clark,**

Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110 USA

**David Maffitt,**

Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110 USA

**John Perry,**

Independent consultant in Hampshire, IL 60140 USA

**Toni Handzel, and**

VirtualScopics Inc., Rochester, NY 14625 USA

**Fred Prior [Senior Member, IEEE]**

Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110 USA (phone: 314-747-0331; fax: 314-362-6971)

Colin Rhodes: Colin\_Rhodes@virtualscopics.com; Steve Moore: moores@mir.wustl.edu; Ken Clark: clarkk@mir.wustl.edu; David Maffitt: maffitt@mir.wustl.edu; John Perry: johnperry@dls.net; Toni Handzel: Toni\_Handzel@virtualscopics.com; Fred Prior: priorf@mir.wustl.edu

### Abstract

There is a global need for software to manage imaging based clinical trials to speed basic research and drug development. Such a system must comply with regulatory requirements. The U.S. Food and Drug Administration (FDA) has regulations regarding software development process controls and data provenance tracking. A key unanswered problem is the identification of which data changes are significant given a workflow model for image trial management. We report on the results of our study of provenance tracking requirements and define an architecture and software development process that meets U.S. regulatory requirements using open source software components.

### I. INTRODUCTION

REDUCING time and costs for the development of new drugs is a major healthcare initiative and is a major concern of pharmaceutical companies and regulatory agencies. Because FDA approval is mandatory before marketing a medically related drug or product

in the United States, developers must meet FDA standards and processes which are designed to validate product safety and efficacy. A key part of this process is a progression of clinical trials which frequently are lengthy and expensive. If these trials could be shortened and performed for less cost, lives could be saved as superior drugs could reach the market sooner and new drugs developed that might otherwise be cost prohibitive [1]–[3].

It has been argued that imaging biomarkers can significantly shorten the time and cost of clinical trials [4], [5]. The use of qualified biomarkers for drug-development is expected to increase the probability of program success and reduce drug development and healthcare costs. Increased reliance on imaging biomarkers means an increased reliance on digital image management systems optimized to meet the unique requirements of clinical trials [6]. Multiple such systems exist based on secure DICOM image transfer [7], [8] or grid technologies [9], [10]. Thus far, however, no systems exist in the public domain that incorporate provenance tracking [11] required to meet U.S. government regulations governing data submission to the FDA.

Part 11 of Title 21 of the Code of Federal Regulations [12] applies to records in electronic form that are: created, modified, maintained, archived, retrieved or transmitted under any records requirement set forth in FDA regulations [13]. The goal of this regulation is to ensure that electronic records (including medical images) are trustworthy and reliable. The regulations include requirements for digital signatures to certify specific data modifications.

Trial management software must facilitate control of all data and insure that system integrity is maintained. All changes to data must be logged and traceable. The system must enforce user access management and change control, and provide an immutable provenance record that can only be accessed by authorized parties, including the FDA. The critical requirements arising from 21 CFR Part 11 relate to electronic signatures and auditing. System developers must determine and document what types of events require electronic signature, how that signature is to be implemented, and what additional data must be recorded in an auditable log.

This paper describes a process for developing and validating an open source software system that can be used in FDA regulated clinical trials. It then presents a proven image management workflow and a detailed analysis of provenance tracking requirements based on that workflow. Finally we introduce an architecture for an Electronic Image Trial Management System (EITMS) that is designed for use by pharmaceutical companies and contract research organizations (CRO) to manage digital image submission.

## II. SOFTWARE DEVELOPMENT PROCESS

FDA regulations require that organizations (a) use good manufacturing practice (GMP) [14] when developing systems that are part of the drug validation process and (b) manage the provenance of all software including both components that are external to the project and components that are developed by the organization.

A controlled development process starts with requirements analysis and a risk assessment. Our software process follows GMP and uses a set of commonly available tools (Rally

Software, <http://www.rallydev.com>) that provide a web-based system for controlling development using Agile software methods [15]. Requirements are identified, recorded, reviewed, and, once accepted, drive the development of design documents, test plans and ultimately software modules. Not all design documents need to be completed before the coding process can begin. A software iteration has goals that are clearly defined relative to the requirements listed in Rally, implementation that meets the design documents and unit tests to verify the outcome.

Software may be categorized in terms of its origin and each category treated differently:

1. Commercial off the shelf (COTS) or open source products that are intended for a larger market but are useful in the context our application.
2. Commercial or open source products that are specific to the domain, but are not controlled by our group.
3. The software that we develop and manage.

General purpose software products may or may not provide access to source code. They may or may not provide access to requirements, design or test documents. We incorporate these products into our applications as follows:

1. Take snapshots and archive released versions so we can reproduce a release at any time.
2. As part of our risk analysis, assume that the general functionality has been tested adequately for our application. For example, we would not specifically test the TCP stack or file system in the operating system.
3. If there are any features that are specific to our application, document appropriate regression tests and include these tests as part of our testing/release process.
4. Only release versions of the software with accompanying tested versions of these products.

Domain specific software products are directly related to our application and provide toolkits or applications that are directly used by our system. These products may or may not have been constructed according to GMP although they are widely used in the industry. They are internally and externally tested, but may not have documented regression tests. We incorporate these products into our applications as follows:

1. Take snapshots and archive released versions so we can reproduce a release at any time.
2. Identify features that are specifically used by our application and develop regression tests that are run as part of our test/release process.
3. Only release versions of our system with accompanying tested versions of these products.
4. Document and report bugs to the developer.

Locally developed software modules are written in response to requirements listed in our requirements documentation. We develop this software according our documented, controlled software quality process.

### III. WORKFLOW MODEL

Our analysis is based on a commonly observed (in our experience) model relating the trial sponsor, CRO and the clinical imaging centers (CIC). The CICs recruit participants, maintain all protected health information (PHI) for those participants, and submit clinical data to the CRO in support of the trial. All data submitted by the CICs have PHI removed and study-specific participant identifiers inserted.

The FDA requires that the system (clinical sites, CRO, process, documentation) has these characteristics:

- Data submitted to the CRO are limited to only what is required to run the trial. Data submitted to the CRO cannot contain PHI.
- Data submitted to the CRO must be sufficient to link the clinical data to the appropriate (anonymized) participant. That is, the system should not allow participant A's clinical data to be mixed with participant B's.
- Data submitted to the CRO can be reviewed at any time and checked for correctness. In practice, that means that sufficient information is maintained to allow the FDA to follow the clinical data including changes from the time it is collected at the clinical site until it is analyzed at the CRO and presented to the sponsor.

Figure 1 illustrates the flow of imaging data from the CICs to the CRO using the software we are developing. This workflow makes a worst case assumption that CICs may not have internet access and may have little or no local IT support. We therefore assume all communications are asynchronous and a software deployment model based on a preconfigured, low cost laptop PC.

1. Imaging data are first collected at a Clinical Imaging Center (CIC) and then transmitted to the Clinical Studies Anonymization Workstation (CSAW) which functions as an edge device. A site may transfer the data using a DICOM network transmission (1a, 1c, 1e), a DICOM CD (1d, 1f) or possibly images copied to a portable USB drive (1b, 1f). Figure 1 shows a Satellite Imaging Center attached to a CIC. It is common that one CIC collects images and also recruits participants from affiliated imaging centers that are considered remote to the CIC.
2. PHI is removed from the imaging data and replaced with a trial identifier (for that participant) at the CSAW. A digital signature is applied so that a user can attest to the changes made in the data. The imaging data and a manifest are encrypted and transmitted over the Internet, passing through outbound and inbound firewalls. The CSAW also supports web-based data transmission forms to capture imaging or other information that is not recorded in the DICOM image headers.

3. Data pass through the inbound firewall to the Image Check-in and QA Management System (ICMS). The network protocols on this link and the Internet link (2) include DICOM and HTTPS. The open source software suite includes ICMS software, but we expect CRO's may substitute a proprietary system at this point.
4. Images collected by the ICMS are stored in quarantine until QC is successfully completed then they are released to the trial archive.
5. Images collected by the ICMS are transmitted to a QA workstation for visual inspection. Though we plan to automate the QC process as much as possible, some tasks are best performed by humans.
6. Depending on the trial, images may be presented to radiologists for review, measurement or scoring or routed to an analysis system for automated processing. The IRMS manages this process and controls access to images and analysis results.
7. Analysis results are sent to the Sponsor. Potentially the Sponsor may also retrieve images from the IRMS for review.

This flow of data and workflow model are based on these premises:

- Mapping of PHI to Trial Id (TID) occurs at the imaging site, is maintained at the imaging site and is not known to the CRO.
- A study participant must be identified consistently across time within a given trial (same TID) but not necessarily across trials.
- While study date and time are PHI, the relative time interval between sequential acquisition events must be retained.
- The CSAW must automatically identify repeat subjects and flag them to the user.
- The CSAW must support a mechanism for distribution (or consistent local creation) of TIDs.
- The CSAW must be remotely serviceable by CRO tech support and minimize exposure to PHI in this process.

#### IV. ACQUISITION WORKFLOW AND PROVENANCE TRACKING

Our analysis of 21 CFR Part 11 provenance tracking included the entire workflow described in the previous section and included all named system components. To illustrate the results of this analysis only the CSAW edge device will be examined in detail.

The CSAW is a collection of three related applications. First, a DICOM Storage Service SCP is always available to accept inbound images over a standard DICOM network connection. Images submitted to the CSAW are stored in an imaging cache with full PHI and made available to other CSAW applications.

Second, a user application with a graphical user interface supports de-identification, management and data submission. This same user application supports the import of images

from a DICOM Part 10 CD or collection of images on a USB drive. Imported images are stored in the same imaging cache.

The user application presents the DICOM data sets captured in the staging area for management by the user. This management typically consists of four functions:

- Automated and manual evaluation of imaging data for study protocol adherence.
- Automated and manual mapping of PHI data to TID.
- Digital signature applied to data attesting to user actions and modifications to data.
- Submission of a work item in a queue for CSAW processing that will perform the de-identification (PHI removed from DICOM header), re-identification (TID inserted into DICOM header), and image transmission.

Third, the CSAW application is an export service that processes items deposited in a queue that define the procedures for modifying and transmitting the data to the CRO. The queue entries are specific enough to define the anonymization parameters and destination for the data.

Figure 2 shows the flow of data through the CSAW. Images are received and stored in the Image Cache. Input from the user creates the PHI to TID map. This map is stored in a table so it can be applied when an imaging study is received in the future. The user provides a digital signature that is the user's attestation to the changes made in the data prior to submission. The user input and de-identification rules are applied by the export service to produce the image data sent to the CRO. At points along the processing chain, audit entries (date-time stamp, user, action performed) are generated and stored in an immutable Audit Log in the CSAW. This Audit Log can be used by FDA auditors to track the provenance of data through the CSAW to the CRO.

The Part 11 Audit Log in the CSAW must log all changed information including pre and post change values. The log is hierarchical such that imaging-study level changes are inherited at the series level and both study and series changes are inherited at the image level.

The study coordinator or other imaging site authority must digitally sign that the following things have occurred at the edge device:

1. Correct subject data were received;
2. PHI to TID mapping is correct;
3. De-identification was correctly applied (CRO shares responsibility for correctness of de-identification by providing template);
4. All acquired data were processed and sent to the CRO (along with image manifest/data transmittal form).

## V. AUDIT STORAGE AND ACCESS REQUIREMENTS

Users at the CIC will know the participants and therefore already have access to PHI. The CSAW application is written with that assumption. The Audit Log is not maintained to satisfy HIPAA requirements, but is used to track the history of changes made to the data once received by the CSAW. The Audit Log must be immutable and only accessible by the FDA, CIC administrators, and CRO Tech Support. Access by CRO data analysts is not permitted so as not to influence any measurements or analysis.

The Audit Log is maintained in a secure DBMS at each site. The IHE Audit Trail and Node Authentication (ATNA) [16] Integration Profile is used to format and record auditable events. Extensions are required for research events that are not part of the clinical requirements addressed by ATNA. All audit information is stored locally and securely.

## VI. CONCLUSION

While it is impossible to create a turnkey system that is Part 11 compliant [17] we believe it is possible to construct an open source software system that will meet the technical requirements of a compliant system. The software is developed according to FDA requirements and includes proper requirements and design documents. A critical issue for other organizations is the manual for Installation and Site Verification. These are instructions and end user tests that a CRO would use to validate that the software is operating properly and meets the requirements of the trial. Other users should be able to take our system, package it with their enhancements, and meet the FDA requirements, supported by the validation tests provided. Of course this open source tool will have to be combined with proper procedural and administrative controls in order to be utilized in an actual clinical trial.

## Acknowledgments

This work was supported in part by the U.S. National Institute of Health under Grant 1 R41 CA132790-01A1.

## References

1. U.S. Food and Drug Administration. Critical Path Opportunities Initiated During 2006, Topic 2: Streamlining Clinical Trials, Propose Regulations to Require Electronic Submission of Study Data. 2006
2. FDA Review.org. The Independent Institute, Theory, Evidence and Examples of FDA Harm. Available: <http://www.fdaireview.org/harm.shtml>
3. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ.* 2003; 22:151–186. [PubMed: 12606142]
4. Hehenberger M, Chatterjee A, Reddy U, et al. IT solutions for imaging biomarkers in biopharmaceutical research and development. *IBM Systems Journal.* 2007; 46:183–98.
5. Chen J, Wong S, Chang J, Chung P, Li H, Koc U, Prior F, Newcomb R. Wake-up Call for the Engineering and Biomedical Science Communities – Major Challenges in Biomarker Development and Application. *IEEE Circuits and Systems.* 2009; 9:69–77.
6. Steiger P. Use of Imaging Biomarkers for Regulatory Studies. *J Bone Joint Surg Am.* 2009; 91:132–136. [PubMed: 19182040]
7. Vendt B, McKinstry R, Ball W, Kraut M, Prior F, DeBaun M. Silent Cerebral Infarct Transfusion (SIT) Trial Imaging Core: Application of Novel Imaging Information Technology for Rapid and

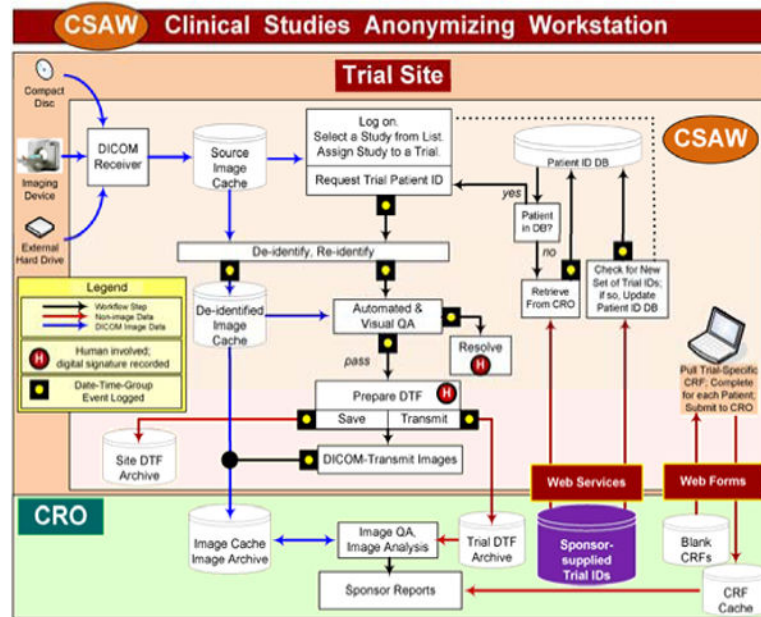


Central Review of MRI of the Brain. Journal of Digital Imaging. 2009; 22:326–343. [PubMed: 18398653]

8. Fitzgerald, TJ. Development of a Queriable Database for Oncology Outcome Analysis. In: Rubin, P.; Constine, L.; Marks, L.; Okunieff, P., editors. Cured II – LENT Cancer Survivorship Research And Education: Late Effects on Normal Tissues. Vol. 2. New York: Springer; 2008. p. 55-66.
9. Erberich SG, Silverstein JC, Chervenak A, Schuler R, Nelson MD, Kesselman C. Globus MEDICUS — federation of DICOM medical imaging devices into healthcare grids. Stud Health Technol Inform. 2007; 126:269–278. [PubMed: 17476069]
10. El-Ghatta SB, Cladé T, Snyder J. Integrating Clinical Trial Imaging Data Resources Using Service-Oriented Architecture and Grid Computing. Neuroinform. 2010;10.1007/s12021-010-9072-z
11. MacKenzie-Graham AJ, Van Horn J, Woods R, Crawford K, Toga A. Provenance in neuroimaging. NeuroImage. 2008; 42:178–195. [PubMed: 18519166]
12. Code of Federal Regulations, Title 21. 2009. Available: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=11&showFR=1>
13. U.S. Food and Drug Administration. Guidance for Industry Computerized Systems Used in Clinical Investigations. 2007. Available: <http://www.fda.gov/OHRMS/DOCKETS/98fr/04d-0440-gdl0002.PDF>
14. Code of Federal Regulations, Title 21 parts 820. 2009. Available: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=820&showFR=1>
15. Martin, R. Agile Software Development: Principles, Patterns, and Practices. Upper Saddle River, NJ: Prentice Hall PTR; 2003.
16. Audit trail and Node Authentication. 2010 Mar. Available: [http://wiki.ihe.net/index.php?title=Audit\\_Trail\\_and\\_Node\\_Authentication](http://wiki.ihe.net/index.php?title=Audit_Trail_and_Node_Authentication)
17. 21CFRPart11.com. Can a vendor guarantee compliant software for Part 11. Available: <http://www.21cfrpart11.com/pages/faq/index.htm>







**Fig. 2.**  
Imaging Acquisition and QA Workflow and Auditing Points.