

# Whitepapers on Imaging Infrastructure for Research

## Paper 1. General Workflow Considerations

Bradley J Erickson, Tony Pan, Daniel J Marcus, CTSA Imaging Informatics Working Group

### *Introduction*

The use of in vivo imaging data in clinical research can provide enormous scientific benefits, but it can also entail substantial complexity. There is a general process for developing a clinical research project. Figure 1 shows a high-level description of this process. Clinical research that involves imaging follows the same general workflow. However, the inclusion of imaging into a research protocol has additional workflow considerations and complexities. By implementing standard procedures and enforcing them through software and policy, many of these complexities can be mitigated and imaging can be successfully integrated into a variety of clinical research applications. The purpose of this set of three papers is to document some of the additional workflow considerations related to imaging that is used as part of a clinical trial.

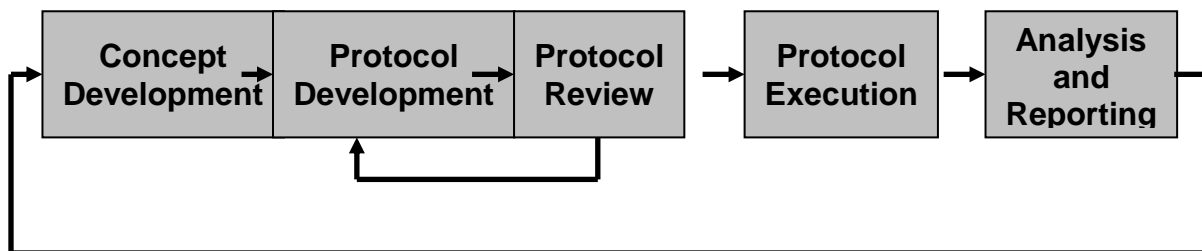


Figure 1. High level process diagram for development of a research project.

An important point that the authors wish to make is that imaging information is nearly always not simply an additional data point. The inclusion of imaging data results in many additional complexities that can lead to unintended or unrecognized risks, biases and errors. For that reason, it is critical that imaging experts be involved in studies that rely on imaging data. That involvement is required during the conception and design of the experiment, the data collection phase, and the data analysis phase. Some of those challenges that are present when imaging data are used in research will be further defined in this paper. The proper collection of imaging data for research use demands monitoring that can be made better and more efficient than current manual

methods. How measurements are extracted from the raw images and represented is a critical step that entails its own set of challenges. Finally, there are some unique and common security issues when images are used for research.

## **Paper 1. General workflow considerations for imaging-based clinical research**

We begin here by defining the general workflow paradigm for imaging-based clinical research. We intend to embellish the clinical research workflow illustrated above with constructs for the use of imaging in clinical research. One of the earliest steps in developing a research protocol is defining the hypothesis or aim of the research. As details of the hypothesis or aims are identified, one begins to better clarify the steps that must be undertaken to prove or disprove the hypothesis. Those steps can be considered a protocol workflow. In general, such a workflow is likely to include the following steps.

### *1. Development and Distribution of the imaging protocol.*

-Selection of sites and devices. The protocol selection process should involve scientific personnel who understand the aims of the project and imaging experts who understand the limitations of the imaging devices available. It is often the case that the imaging protocol will be driven by the weakest of the devices or alternatively that a site must be dropped from the study if it can't complete the required scans. For example, a study may require data to be acquired on a 3T scanner or that a scanner can acquire data with acceleration in order to complete on time. This phase of study implementation is best completed as early as possible so that sites can be selected before large amounts of time and money have been invested. However, it's inevitable that the final protocol will be developed after commitments have been made to sites. Therefore study investigators should be prepared to make difficult decisions about acquisition protocol and site inclusion as the startup phase proceeds.

-Selection of image types. Once a group of participating sites has been identified, the scanning protocols must be shared among sites. By scanning protocol, we mean the precise acquisition protocols and associated image reconstruction techniques that are used to produce the images that will be analyzed for the research protocol. While this

may seem trivial, it is often necessary to have unique scanning protocols for each software release of a given imaging device, let alone for each manufacturer.

-Collection of a device inventory. The inventory should include vendor, model, software version, coils, availability of vendor research agreement, and more. Contact list for scanner technician, physicist, and imaging PI. Maintaining this inventory in a table by site and field is invaluable for managing site startup.

-Harmonizing & calibrating protocols. The different device vendors have different programmable capabilities that define how an imaging sequence is executed on their devices. An example is that in MRI, the way that the RF pulses are created results in different shapes, which in turn, produces slightly different excitation and ultimately a slightly different signal. These differences may or may not be visible, but they usually are measureable. It is therefore necessary for each scanner platform that will be included in a study to work with a physicist with deep experience on that specific platform to develop the sequences that will be used for the study. It is typically the case that the sequences deployed for clinical use on scanners are either not adequate or not equivalent and therefore must be modified to harmonize the acquisitions across the study sites.

-Electronic protocol distribution. Ideally, the scanning protocol would be represented in an electronic form that could be consumed directly by the imaging device. This would reduce the chance of a human error entering incorrect scanning parameters into the imaging device. Standards for this are currently lacking, but are well within the capabilities of current technology.

-Acquire test data to validate protocol. Human data should be acquired on each platform and assessed using the measurements, including both manual and automated techniques that will be used in the actual study. Ideally images of the same person would be acquired on every platform, to ensure that measurements are consistent and is well worth the investment. When this is not possible, whether due to number of systems, radiation exposure, or other constraints, studies from each platform should still be acquired and compared for qualitative similarity of image characteristics. Acquisitions of physical phantoms can also be helpful for quantitative and qualitative assessment of cross-platform harmonization.

## *2. Site qualification acquisitions.*

In this step, humans, animals, or phantoms are imaged in a way that allows the researcher to be confident that a given site is able to perform the imaging examination. There are important differences that exist not only between devices from different manufacturers and different software revisions for a given model of an imaging device. There can also be degradations of image quality due to local site modifications as well as differences in quality control procedures. The images may be perfectly acceptable for clinical interpretation but precise quantitative measures that are often employed in medical research may be affected by these variations.

The site qualification procedure is often a pre-cursor to a large multi-site trial to assure those responsible for the conduct of the trial that they will be able to acquire the images envisioned by the protocol developers, and that the images are of sufficient quality and similarity to produce useful measurements. The pre-qualification images may also allow further demonstration of the data extraction and analysis tools to be applied to the imaging data.

For studies that are conducted over a long period of time, it is likely that imaging devices will be upgraded to new software releases. As a consequence, it is critical that participating sites coordinate their software upgrades with the central analysis site so that appropriate scanning protocols are provided as these updates occur.

## *3. Scanner scheduling*

Once the scanning protocol has been developed and the sites have demonstrated an ability to perform the scanning protocol, data acquisition can begin on research subjects. Of course, other steps like patient recruitment and scheduling of the examination must be performed, but this is a common need that exists for non-imaging components. While this seems obvious and straightforward, the acquisition of images on clinical devices (and possibly non-imaging tests) can produce challenges in the setting of highly integrated electronic medical systems. Depending on the radiology information system and ordering system that are used, it may be difficult to clearly identify examinations as research examinations. The consequence may be that a

routine clinical scanning protocol is employed rather than the desired research protocol. The scheduling system may also cause complications when more than one likely equivalent imaging device is available. It is well-documented[1] that imaging devices of the same make, model, and software revision can produce quantitatively different images. For that reason, many studies require that all images acquired for a specific research protocol be acquired on a single imaging device. This requirement is often difficult to support with clinical scheduling systems.

#### *4. De-identification.*

Once a study has been acquired, most institutions require that information that identifies the patient (Protected health Information or PHI) must be removed before it is used for research. In most cases the identifying information consists of the 18 identifiers outlined within the health insurance portability and accountability act (HIPAA). In some cases it may be desirable to encrypt some component of the original patient identifier in order to assure that any clinically significant findings can be mapped back to the patient. Another option is to provide an audit log in which the patient information and the research subject identifier are both stored in a file that never leaves the originating institution. There is significant variability in the demands placed on researchers. It also appears that Institutional Review Boards (IRBs) are much more stringent on privacy matters when network transfer is used than when other electronic media (e.g. CD ROMs) are used. While this generally is performed at the originating site, it is recommended that central sites repeat the anonymization process to assure that all Protected Health Information is removed.

#### *5. Insertion of Research Identifiers*

In most cases, it is necessary to attach research identifiers to imaging studies. The process of replacing a clinical medical record number with a research identifier is usually manual. This can lead to errors, and such errors can be difficult to detect and correct. At the same time, there are few clinical trials management systems that provide services allowing direct mapping of clinical record numbers to research identifiers. This latter task would be particularly complex if a patient was participating in more than one trial, as one could not simply return a study ID for a given medical record number. For studies involving more than one examination on a given patient, it would be desirable

for the process to recall the previously used research subject identifier and suggest it to the user, though again, one could not guarantee that a patient was not participating in more than one trial, and automatically insert the study identifier.

There are clinical systems that insert medical record numbers into image headers, and it is feasible to implement a clinical trials system that could perform confident insertion of study identifiers into a research imaging examination. The technology for accomplishing this is well understood and could be an important step in both improving efficiency as well as reducing errors. Based on experience with the clinical equivalent (the Modality Worklist), the error rate could drop from 2% to less than 0.1% while simultaneously eliminating any need for humans to do this step.

Note that, in clinical trials that employ blinding schemes, de-identification and assignment of research identifiers are an important requirement of the study design itself. In these studies, additional considerations specific to study design may need to be incorporated into the de-identification and research identifier assignment methods.

## *6. Image Transfer*

Once patient information has been removed and study identifiers inserted, the images are ready for transfer to the central analysis site. This is often accomplished by writing the images to physical media such as CDROMs. These physical media have several disadvantages including cost of production (labor as well as the media), time delays for physical shipment, and the possibility of losing data due to physical corruption. For these reasons, it is desirable to use network transfer between institutions. Most institutions have a firewall that provides a degree of separation between the network of the institution and the rest of the world. In some cases, moving data out from the institution across the firewall can be a challenge, though most institutions allow port 80 to be used for outgoing traffic, since that is frequently used for Web browser traffic. The greater challenge is then traversing into the central data center, as most firewalls are configured to not allow unsolicited inbound traffic on any port. In most cases, it is necessary to have a server outside the firewall to receive the images from contributing sites and have a second computer inside the firewall of the central site that periodically transfers the images in from the first server. Provisioning a server outside the firewall and assuring a high level of security can be a challenge and

is an area where security professionals should be consulted. We discuss security and privacy considerations as the subject of paper number three. An alternative is to set up a virtual private network (VPN) which connects the networks of the two organizations together. This potentially opens up information of one organization to another, and therefore is generally not a preferred mechanism, but sometimes is the best option available.

### *7. Automated QC Checks*

In many cases, it is possible to algorithmically determine if images are not acceptable. For instance, in the Alzheimer's Disease Neuroimaging Initiative (ADNI) trial, an algorithm was developed that identified image sets with unacceptably high patient motion[1]. Ideally, this would be done on the imaging device to provide immediate feedback to the local technologist that images need to be repeated, rather than relying on their judgment.

### *8. Study Reminders*

For studies that employ periodic re-imaging of patients based on time rather than some clinical event, it may be desirable to have the system generate reminders to the contributing sites. The failure to perform an imaging examination within a specified window can be a significant problem. It can lead to disqualification of subjects, and the loss of significant data. It is feasible to send notifications to both central and contributing sites indicating that an expected study was not received. In some cases, there is a legitimate reason for this study not being acquired, but in many cases such a notification can help to quickly identify an otherwise unrecognized error.

### *9. Scanning Protocol Adherence Checks*

Adherence to a study protocol is a critical element of using images for quantitative analysis. Failure to adhere to a protocol is a more insidious cause for disqualification of examinations and subjects than failure to perform a study or corruption by artifacts. Consequently, it is critical that adherence to a scanning protocol be verified. Rapid feedback might allow capture of that time point within the required time frame, and so is valuable to many researchers. Rapid feedback indicating protocol adherence can also motivate contributing sites to execute a protocol properly. Increasingly, much of the information required to document protocol adherence can be

found within the DICOM header. This can allow an automated process either at the contributing site or central analysis site to provide immediate feedback that a protocol was or was not followed. Further development of tools to analyze both the information within the header as well as the more complex task of extracting information from the actual image pixels could be valuable in improving the quality of images used in research protocols and increasing the adherence rate. Many researchers have developed tools to support their specific research project. Much like image collection and transfer, this is an infrastructure that should be developed once rather than many times. We note here that while DICOM is widely used, DICOM cannot keep up with the very latest imaging techniques. Since research often wants to leverage cutting edge imaging methods, it is not always possible to use DICOM, or at least publicly standardized DICOM fields, for computerized adherence checks.

### *10. Information Archiving*

Once a study has been received it should be written to an archive. Even if it does not pass quality control, most researchers would want to have access to the images “just in case”. The archive should also provide storage for any further products of the analysis process, including both human generated annotations and markup as well as intermediate or final images from a computational method. The user as well as a description of the software used should also be a part of the archive. Archive technology for images is now widely available, though archive technology that integrates images with the associated information (annotations, mark-up, and audit logs) is less common.

The specifications for a research archive are also very different from a clinical archive. The pattern of storage and retrieval are much more like a cache while the image analysis is done, and then becomes much more like archive, with much less frequent access. Like a clinical archive, altering image and associated data is usually not allowed, but mechanisms to reflect corrections must be supported.

### *11. Information integration with clinical information*

Once the image information has been extracted, it is necessary to integrate the image derived information with the other clinical information. This typically would occur in a clinical trial management system but may occur in a less formal setting. This is the point where it is critical that the correct research identifier for the image data be



associated with that same research identifier for non-image information. We will only briefly note at this point that the measurements should be produced in a form that is computable to allow for efficient data reduction. This topic is more thoroughly covered in paper number two.

## *12. Management Reports*

It is critical for the central site to be able to create and view management or status reports. These reports would reflect such information as the number of imaging studies and research subjects that have been contributed from each contributing sites, the status of each study that has been contributed including the analysis steps that have been conducted and those which remain. In many cases it is critical to manage the appropriate level of access to this information, which is a more complex security model than that employed for clinical purposes. A discussion of the security aspects is provided in paper three.

The problems and issues we describe above are not new or novel. However, we believe there is an opportunity because each study typically develops its own infrastructure for addressing these problems. If there was one open source and freely available infrastructure, it would allow researchers to focus on the scientific issues, and would reduce duplicative efforts. We suspect the actual amount of duplication is much larger because nearly every multi-site study that utilizes images requires this infrastructure.

## 13. Software to support imaging in clinical trials

There are some existing packages that are available that address pieces of the issues listed above. Some of those examples are listed in[2]. Components or concepts from these packages could likely be incorporated into a more general solution that could be provided to the research community. Packages such as XNAT (<http://xnat.org>) provide tools for accepting images and associated data, moving those data through defined quality control steps, search and query tools, and provide viewing tools for common data formats.

Executing any research project involves a number of steps. Research involving images is no exception. The fact that images are digital and that measurements likely

also exist in digital format creates an opportunity for a digital workflow. There are widely accepted languages for describing workflow, and there are commercial and free workflow engines that can assist in the execution of a workflow.

Creating a description of a research process using an established workflow description language could be beneficial by encouraging a standard description of how research is to be done, leveraging workflow engines that might assist in the execution of the workflow, and providing a documentation method that might be valuable when data is to be submitted to other agencies such as the FDA.

### **Conclusion**

Creating an efficient and flexible workflow model for imaging-based clinical research is critical. Without it, each laboratory will attempt to build its own tools and infrastructure to address workflow needs. Designing and building it well is the most efficient way to use resources and maximizes the ability of researchers to share research methods and for approval agencies like the FDA to understand how research was conducted.

1. Mortamet, B., et al., *Automatic quality assessment in structural brain magnetic resonance imaging*. Magn Reson Imaging, 2009. **62**(2): p. 365-72.
2. Cancer Imaging Program. *Image archive software solutions*. 2011; Available from: <https://wiki.nci.nih.gov/display/CIP/CIP+Survey+of+Biomedical+Imaging+Archives>.