J Digit Imaging (2012) 25:692–702
DOI 10.1007/s10278-012-9493-3

Whitepapers on Imaging Infrastructure for Research Part Three: Security and Privacy

Tony Pan · Bradley J. Erickson · Daniel S. Marcus · CTSA Imaging Informatics Project Group

Published online: 8 June 2012

Introduction

This is the third part of three describing an extension of the process for developing a clinical research project including the use of in vivo imaging data. The high-level process diagram for development of a research project including images is shown in Fig. 1. In part two of this series, data management requirements and practices for including images were described. This part describes the security and privacy requirements for supporting images used in research.

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 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 is 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.

T. Pan
Center for Comprehensive Informatics, Emory University,
201 Dowman Drive,
Atlanta, GA 30322, USA
e-mail: tony.pan@emory.edu

B. J. Erickson (✉)
Department of Radiology, Mayo Clinic-Rochester,
200 First Street SW,
Rochester, MN 55905, USA
e-mail: bje@mayo.edu

D. S. Marcus
Department of Radiology,
Washington University School of Medicine,
4525 Scott Ave., Campus Box 8225, St. Louis, MO 63110, USA
e-mail: dmarcus@wustl.edu

Security and Privacy Issues

Research consortiums and multi-site trials are increasingly a reality in the healthcare and research enterprise. In addition, there is increasing incentive and/or requirement for federally funded research projects to share the data collected and generated as part of the research plan. As the amount of data being collected increases in volume and complexity, the necessity of data-sharing infrastructure becomes obvious.

With the implementation of formal data-sharing mechanisms comes the requirement of secure data sharing, both to protect subject privacy, as well as to provide investigator interests. Research image sharing scenarios face the same challenges as those for the generic data-sharing scenario. At the current time, however, there exists a gap in a researcher, an institution, or a research project's ability to provide common infrastructure to support secure image sharing in a research scenario.

In this part, the authors investigate the security requirements of the research imaging workflow outlined in part 1. While a significant emphasis will be placed on the security requirements of image sharing, it should be noted that data acquisition as well as analysis and result management also have significant security requirements and implications. The objective of summarizing these security requirements is to provide general guidance in the design and implementation of a secure image sharing system and to motivate research

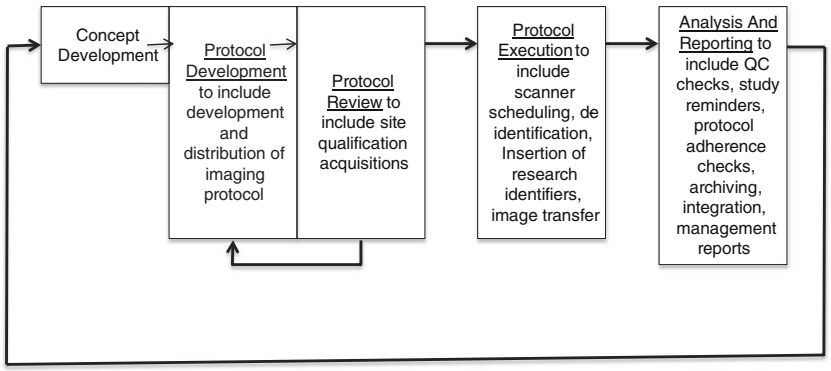

Fig. 1 Process diagram including the workflow steps required for imaging infrastructure for research

organizations and funding agencies in their consideration of data security issues in collaborative research.

The paper is organized as follows. The first section reviews the research workflow outlined in part 1 and identifies the specific component of the research workflow whose security concerns this paper focuses on. The next section outlines relevant regulations and policies that impacts secure image data sharing. Next, a high-level overview of the specific capabilities required to support secure data sharing is presented. The next section discusses the security requirements of different general data-sharing scenarios. The final section lists the current available standards and technologies and gaps that might be applied to address some of the requirements we outline.

Research Workflow

There is a clear flow of steps when imaging is used in clinical research. These are outlined in part 1. It begins with the preparation steps of defining what and how imaging will be used (protocol definition), assuring that the involved clinical sites will be able to acquire the images (protocol sharing and site qualification), actually collecting images on subjects and sending the image data in a compliant way to the central analysis site (image acquisition and collection), and doing the data analysis (data analysis, quality control, and reporting). Security is a major consideration. This includes protecting private information about patients/subjects. This is well known to those in clinical practice. Clinical research adds another consideration—the need to keep the data private for the proper people conducting the research, and to have acceptable audit trails for any modifications or annotations to the data. It is critical that these issues be properly addressed up front—changing processes and systems to accommodate security requirements is usually a painful and ineffective strategy.

This paper is motivated by data-sharing activities including image acquisition, collection, analysis, and reporting. These involve primary data, which are the images and clinical data, and derived data, including generated images, analysis results, and reports. Within the research protocol, the principle primary and secondary data-sharing activities include moving data from the data collection sites (from image acquisition device and image archive such as PACS, and from clinical data repository) into the study's data repositories as well as retrieving images from the image and clinical data repositories for reporting, quality control, data analysis, data sharing, and reuse.

The central pattern of these data-sharing activities is a transmission of data between the data provider and data recipient. Security becomes a concern when the data provider and recipient are in different security jurisdictions and the data moves across jurisdiction boundaries. Jurisdictions in this context are defined and governed by a set of policies, which may be formally or informally stated at different granularities. Examples of jurisdictions include patient care in a healthcare enterprise, a single IRB approval, a research protocol, and data-sharing practices of a single researcher. In the case of clinical research, the jurisdictions align primarily along two different dimensions of concern: patient privacy and intellectual property. Along the patient privacy dimension, a data-sharing activity needs to consider the presence and visibility of protected health information (PHI). A researcher may wish to control access and visibility of the data based on the stage of the research and collaboration and data-sharing agreement in order to protect intellectual property. Each of these dimensions is also governed by regulations and policies such as HIPAA, IRB, and data-sharing agreements.

The discussions in the remainder of this paper are framed by the general idea of data movement across jurisdictional domains. Specifically, this paper will focus on the security requirements and capabilities needed in the research data-sharing environment and the research image data repository.

Regulations and Policies

Image sharing in clinical research settings may involve the use of human and animal data. There are several laws, regulations, and policies at the governmental and institutional levels. In this section, we briefly describe some of the

relevant regulations, categorized into data centric, infrastructure centric, and process centric regulations.

Data Centric Considerations

The US Department of Health and Human Services publishes an International Compilation of Human Research Protections every year which includes legislations, regulations and guidelines for data and privacy protection in almost 100 countries. Clinical trials and research that involve multiple countries should consult this document regarding rules and regulations that are applicable in participating countries. In the USA, the major legislations include the Privacy Act of 1974, the Health Insurance Portability and Accountability Act (HIPAA) of 1996, and the Confidential Information Protection and Statistical Efficiency Act of 2002. The most current regulations are detailed in the HIPAA Privacy Rule: Standards for Privacy of Individually Identifiable Health Information and the HIPAA Security Rule: Security Standards for the Protection of Electronic Protected Health Information.

In the context of multi-center clinical trials and research, the Privacy Rule allows three types of data to be exchanged for research: identified data with patient authorization, limited data with data use agreement, and de-identified data. The Privacy Rule recognizes two ways to de-identify PHI from data sets. The first approach, the "safe-harbor" method, is to remove all 18 identifiers as enumerated in the regulation:

1. Names.
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, except for the initial three digits of a ZIP Code if, according to the current publicly available data from the Bureau of the Census:
   (a) The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people.
   (b) The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers.
5. Facsimile numbers.
6. Electronic mail addresses.
7. Social security numbers.
8. Medical record numbers.
9. Health plan beneficiary numbers.
10. Account numbers.
11. Certificate/license numbers.
12. Vehicle identifiers and serial numbers, including license plate numbers.
13. Device identifiers and serial numbers.
14. Web universal resource locators.
15. Internet protocol address numbers.
16. Biometric identifiers, including fingerprints and voiceprints.
17. Full-face photographic images and any comparable images.
18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.

The second approach requires a qualified statistician to determine, using generally accepted statistical and scientific principles and methods, that the risk of identifying study subjects is very small. The Limited Data Sets contain some useful identifying information such as zip codes and all elements of dates, but are stripped of 16 categories of direct identifiers. Note that among the direct identifiers that must be removed in both de-identified and limited data sets are the "Full-face photographic images and any comparable images." This specific requirement should be addressed in systems and studies that involve three-dimensional images of the head. On the other hand, it is also important to note that fully identified data may be used and exchanged within a multi-center study as long as written authorization is properly obtained from study subjects.

The HIPAA Security Rule is applicable when identified and limited data sets are used. It stipulates specific requirements for administrative safeguards, physical safeguards and technical safeguards. Of special relevance are requirements for access management, authentication, audit control, encryption and integrity, transmission security, backup and storage, and disaster recovery. In particular, access control should enable authorized users to access the minimum necessary information needed to perform job functions.

An Institutional Review Board (IRB) governs the use of confidential and protected data in research. Researchers are required to obtain IRB approval or exemption for each research study when working with PHI laden data. As each IRB considers its own institutional requirements, the policies may be inconsistent between institutions. It is often useful for a multi-institutional research study to establish common language for IRB approval applications to be submitted to the individual IRBs.

Patient consents form a component of the research protocol and governs how patient data may be used in

research. Consent may be obtained for a single research study, or may have blanket coverage for all current and future research studies. In the second scenario, the data are protected by HIPAA and IRB approval as well and may require de-identification before release to the researcher.

NIH outlines its general policies on data sharing for NIH-funded grants at http://grants.nih.gov/grants/policy/data_sharing/. Specifically, the policies identify privacy and confidentiality issues, as governed by HIPAA and IRB, and intellectual properties issue such as timing of data sharing and proprietary data release.

The Federal eAuthentication Guideline establishes four levels of assurances for authentication processes involved in electronic transactions. The level of assurances can be met using different mechanisms for authentication, ranging from username- password pair to biometrics and hardware tokens. The full guideline is available at http://csrc.nist.gov/publications/nistpubs/800-63/SP800-63V1_0_2.pdf.

The Federal Information Security Management Act of 2002 (FISMA) requires federal agencies to establish security practices to protect information and information systems from unauthorized access, modification, and disruption. The National Institute for Standards and Technology (NIST), under FISMA, provides guidance and policies how to assess security needs based on information and information system categorization (FIPS PUB 199, NIST Special Publication 800–60) and recommendation on minimum security requirements and implementation (FIPS 200, NIST Special Publication 800-53). FISMA and NIST guidance and policies serve two roles in the context of image sharing security requirements. Firstly, the FISMA and NIST policies inform the basic security requirements of the image sharing environment. Secondly, as the image-sharing infrastructure matures and used in interaction with federal information systems, the security requirements as outlined by FISMA and NIST will need to be supported by the image sharing infrastructure.

21 CRF Part 11 specifies the criteria for establishing the trustworthiness and reliability of electronic records. It establishes guidelines for the change management and documentation of the electronic system and audit and signature of its data content. Systems supporting clinical trials that must comply with FDA regulations should also consider additional requirements specified in 21 CFR Part 11. Main system level requirements include source data documentation and retention; audit trail, time stamp formats, and data entry controls. In addition, the development and deployment of the image sharing system itself will need to provide adequate documentation and change management.

Securing Image Data Sharing: Capabilities and Requirements

As described previously, security is an important consideration when data and image are shared across jurisdictional boundaries. To support secure data sharing, a number of capabilities must be implemented and choreographed. In this section, we provide a high-level overview of each of these principle capabilities and their requirements. These capabilities are broadly categorized into those that relate to "Data Transformation for Security," "Infrastructure Administration," and "Data Access Control."

Data Transformation for Security

Data may be transformed to protect access of sensitive data. These may include fundamental changes to the data content, as in the case of de-identification, and utilization of information hiding techniques, such as encryption.

De-identification of medical imaging data involves complete removal of identifying information without tracking the mapping between the resulting data and the patient.

For medical imaging data such as DICOM, de-identification involves the DICOM header and associated images. De-identification of the DICOM header is usually straightforward. The DICOM standard specifies the data elements that store patient names, IDs, study date, and other PHIs. A complete de-identification simply replaces all PHI data elements with pseudo values that are constant, random, or assigned by the coordinating center. It can be more complex because the standard allows for free text entry as well as private elements that might contain identifying information. Consequently, it is best practice to remove all the descriptive elements such as the "Series Description" or "Patient Comment" field and all private elements inserted by imaging manufacturers. The problem with routinely removing all private information is that "cutting edge" sequences usually put technical information about the sequence in these private elements until standards bodies describe where and how to place the information in public elements. Some clinical studies de-identify DICOM images to a Limited Data Set by maintaining all the dates in the DICOM header or by shifting the dates by a random offset. Another argument for a Limited Data Set is that, strictly speaking, the various DICOM Unique Identifiers (UIDs) in the DICOM header can be considered as identifiers in the eighteenth category of the HIPAA safe harbor standard, but their removal or modification is problematic due to the nature of how DICOM UIDs are assigned.

Frequently, imaging studies contain images that contain patient information embedded in them (e.g., secondary capture images). If these images must be included as a part of the data set, then imbedded text that contains PHI

must be removed by cropping, pixel replacement or other more advanced image processing techniques. For three-dimensional images, such as CT or MRI, of the head, we must also address the concern of one's ability to reconstruct a subject's face using three dimensional rendering techniques. Again, simple cropping of the images may be sufficient to de-identify them. But, if facial features must be preserved in these images, the final dataset may need to be treated as identified data sets even if the headers are properly de-identified.

Some DICOM studies may contain a special type of DICOM objects called DICOM Structured Reports in which image observations, measurements, and diagnostic findings are reported in both structured and narrative text elements. For these objects, de-identification must also take care to remove potential PHI in the narrative text elements manually or with the assistance of natural language processing techniques.

In some cases, anonymization or partial de-identification may be required by the research protocol. In anonymization, the mapping is kept by a trusted third party such as an honest broker, which may be used to re-identify the data in cases such as when a research finding critical to patient health requires immediate notification to the patient. Partial de-identification may be required in other situations, for example public health studies may require the use of actual zip codes.

Closely related to the de-identification and anonymization is the need for a research identifier. A research identifier is an identifier that is unique within a research study and is used to confidently identify a subject or sample. It may or may not be linkable back to the original patient identifier.

To enable this linkage, there is often a need for privacy preserving transformation. Such a transform preserves information utility while hiding specific confidential information. An example is to statistically model the zip code distribution in a data set and reassign new zip codes based on this distribution in order to maintain utility of the data for population studies, without associating individual subjects to their actual zip codes. Other approaches include statistical aggregation, for example replacing subject age with age range. The transformation to be used should be chosen during research protocol design so as not to adversely affect the data analysis process.

The image data may be encrypted to provide a layer of protection in addition to the data access control layer in the image data-sharing infrastructure and in the image repository. Encryption protects the data in case the data access control layer is bypassed. The encryption may be performed for the entire storage media or for the specific data files. It is important to consider how data encryption and communication encryption, discussed in "Data Access Control," interact. It is also important to consider the interaction between the data access layer of the infrastructure in the encryption and decryption of the data and the management of the encryption key.

A "signature" is used to ensure that the data are original or alternatively have gone through modifications by known users or systems. Signatures do not change the content or representation of the data, but rather, provide assurance that the data are produced by a trusted source. This typically performed by computing a checksum or a one-way hash. The collection of signatures representing all users or systems that have touched the data may be tracked. In addition, redaction may be used when signing a partial subset of data without affecting the originating signature. It is important to consider the how data signature and message signature during communication, discussed in "Data Access Control," interact.

An honest broker is a trusted third party that can perform the above listed tasks, including de-identification, reidentification, management of research identifiers, data transformation, signature, and encryption. The honest broker resides in one jurisdiction and manages the policy-driven data transformation and mapping to allow data access from a different jurisdiction. An honest broker is a common method for obtaining data in one clinical jurisdiction from another clinical research jurisdiction. Brokers may be human or may be computer systems that can perform policy-based transformations.

Infrastructure Administration

The data-sharing infrastructure provides critical capabilities to support the image sharing activities. Specifically, "Policy Management," "Authentication," "Audit Logging," and "Trust Management" are important components in the Security Infrastructure.

Policy Management

A research image-sharing infrastructure consists of multiple components and actors governed by different jurisdictional policies. The management of this infrastructure and the policies is an important aspect of the image sharing security. In addition, the infrastructure should provide fundamental operations such as user and host authentication. As data sharing across jurisdictions is based on policies, a critical component of the infrastructure is the policy management system in each jurisdiction, and the policy mapping between jurisdictions. Some major policy and regulation categories have been outlined above. It is important to note that often this is a difficult problem as policies vary in content between jurisdictions, and format of the policy statement may be explicit but in non-computable form (e.g., word documents), or even implicit (individual researcher's decisions). Policy

management in the infrastructure therefore needs to have the flexibility of support computable policies as well as human decisions (based on implicit or non-computable explicit policies).

While systems for managing policies and policy mapping may not be standardized, it is useful for a research image sharing system to implement mechanisms to support the formulation and evaluation of standardized, computable policies such as research protocols, IRB approval, and patient consent, which, together with other explicitly defined policies, form the authorization policies and the corresponding jurisdictional domains.

Authorization policies are typically created by an administrative user to reflect and support the policies outlined above. In some instances, the policies may act as authorization policies directly (e.g., patient consent can be viewed as authorization policy based on research study and researcher role). Often, the authorization policies represent components of the broader policy, for example in IRB approvals and research protocols.

Authorization policy management can be configured to support different levels of distribution. Both the data to be secured as well as the authorization policies can be managed centrally. In this scenario, the collocation of the data and authorization policies allow efficient reference of the protected data in the authorization policy. The construction of the policies can take advantage of the fact that the policy administrator will have access to the most up-to-date view of the data. Alternatively, the data can remain in distributed storage locations while the authorization policies can be managed centrally. In this scenario, data references must allow unique identification of the data globally.

Distributed authorization policy management enables multiple policy repositories for different data repositories. There may be a "many-to-many" mapping between policy repositories and data repositories. A data service may also have a closely coupled policy repository that contains policies that are specific to the data type and attributes. When multiple policy repositories are associated to a single data repository, the need for policy synchronization and the possibility of inconsistent policies in different repositories for the same data adds a significant amount of complexity. However, multiple policy repositories allow policies to be managed at different granularity, for example for research study and for research institution.

In a multiple institutional data-sharing environment, user identity may be managed centrally or in a distributed fashion. Central user management establishes an organization-independent user domain but requires a separate identity for each user. Such a user management system may be administered by the data coordination center of a research consortium.

Distributed user management, on the other hand, rely on the identity management system of participating institutions, but requires standard interfaces for exchanging user information to be layered on the existing identity management systems such as LDAP and Active Directory.

A user management system must provide the ability to create, delete, and inactivate users. It may need to manage additional user attributes. The system needs to issue authentication tokens that are consumable by the other resources in the data-sharing environment. It also needs to support revocation and expiration of existing tokens, which affects the user's data access authorization to the data resources.

The user's roles within a study, for example "radiology reviewer" and the user's attributes, for example "RSNA member" are important factors in authorization policy. While some identity management systems may manage use roles and attributes, the use of these roles and attributes are not consistent between different management systems and across organizations or jurisdictional domains. It is therefore likely that the research image-sharing infrastructure may need to supply a role and attribute management system. A research study may also wish to define the roles and attributes, and mapping to participating user's roles and attributes if exist to provide consistency.

Authentication

Authentication is the process of asserting the identity of a user or a computational entity. Authentication therefore involves user management, the act of authentication, and the transmission of the authentication token for use in subsequent authorization and other security related activities.

User authentication is the act by which a user proves his or her identity to the system, using information such as username and password, software token, biometrics, or hardware tokens. The identity management system, based on the user provided information, asserts whether the user is known and has proven his or her identity. Once the user's identity has been confirmed, a software token can be issued and used by other resources in the system, such as authorization components to support authorization decisions.

Similarly, a secure image sharing infrastructure may require that the software and hardware components in the sharing environment authenticate themselves. The software or hardware service or component needs to prove its identity. This can be done via pre-assigned tokens such as a certificate, by prearranged passphrase, or by name, for example in DICOM. Web servers utilize signed certificates to prove their authenticity. A client web browser maybe required by a web server to furnish a certificate. The authentication supports assurances that the service a user interacts with is truly the one the user intends to interact with.

Audit Log Management

For any data access and data transformation that has been performed in the image-sharing infrastructure, it is important to log the access and transformation in order to support auditing. Auditing provides a way to detect and assess security anomalies and policy compliance. Audit log also allows a researcher to understand, to a limited degree, the provenance and the source of the data. For effective and trusted audit log analysis, the infrastructure needs to provide audit log management as a trusted service. In addition, components in the research image-sharing infrastructure wishing to leverage the audit log management service should conform to the audit log syntax.

Trust Management

In a distributed environment, it is important for the resources, such as a data repository, to trust other resources that it depends on, such as user authentication and audit logging services. Trust relationships enables one resource to more easily "accept" the authenticity and quality of the information being provide by the other party. This is particularly important for components that provide authentication service, authorization policy repository, and auditing. An example of how trust is managed involves certificate authorities on the internet such VeriSign. A web server may use a certificate that is signed by the VeriSign certificate authority, which establishes a level of assurance that the web server using the certificate is known and registered with VeriSign with traceable information such as payment method and confirmed address. Once the trust mechanism has been defined and implemented, a resource or user in the image sharing environment can make reasonable assessment of the quality and authenticity of the information exchanged, such as user credential and authorization decision.

Data Access Control

Secure data sharing using an image sharing infrastructure fundamentally is about access control. The previous two sections discuss data transformation to protect information content, and data-sharing infrastructure administration. In this section we discuss "Data Access Control" using the security related information captured in the infrastructure, to transport the transformed data. We discuss "Authorization" and "Delegation," "Audit Logging," "Non-repudiation," and "Transmission Protection."

Authorization

A proper research data-sharing system requires a service that integrates authorization capabilities that can allow, disallow, or partially allow a user to access the data and resources managed by the service, such as an image repository. An authorization decision needs to be made by accounting for user identity, roles and attributes, existing security policies, and data content.

Authorization Decision Engine

An authorization decision engine utilizes the user identity, role, or attribute to evaluate if he has access to the requested resources based on the existing authorization policies. The decision engine may be centralized or collocated with the data repository. When the decision engine is collocated with the data, privileged data are not transmitted over network unnecessarily. The authorization policies may be stored with the data or may be centrally managed. If the policies are collocated with the data, decision engine can determine authorization using local data and policies. If the policies are centrally managed, the authorization engine must first retrieve the policies, and care must be taken to minimize the overhead of policy retrieval.

If the authorization decision engine is centralized, a mechanism of notifying the image repository of the authorization decision is required. Security Assertion Markup Language is one such mechanism for transmitting the assertion that a user is authorized. Note that if the authorization needs to rely on content of the data, a local decision engine would still be necessary.

The tradeoff between central and local management of authorization policies and the authorization decision engine implementation affect how the system is optimized for performance. For example, policy databases collocated with the data allows optimized queries to interleave the authorization decision evaluation and the data retrieval.

The authorization decision may be formulated as a pre-data access filter, where the query is transformed to include conditions as outlined by the authorization policies. Alternatively, the decision may be made as a post-data-access filter, where the database query results are filtered to match the authorization policy requirements.

Delegation

It is common that a user may wish to delegate his or her data access authorization to another user or an automated agent. This is often required when a user invokes an automated analysis process that needs to access protected data or when an administrator may want another administrator to perform actions in his absence. Delegation may be one of two forms—user delegation or authorization delegation. In user delegation, the user credential is delegate to another user. This allows the second user to act in all capacity as the first user. In authorization delegation,

the first user delegates the authorization to perform a particular action on a particular resource to a second user. The authorization delegation model provides tighter control over authorization and allows more fine grain delegation and audit tracking. However, it may be harder to manage, especially when multiple policy sources and decision engines are involved.

Audit Logging

To detect non-compliance with security policies and to track data access by users, each service should provide or integrate audit logging capabilities. The service may wish to log who performed what action on what resources, authorized based on which policies. An important aspect of audit logging is the generation of standard log messages, so a consistent tool may be used to query, display, and analyze the audit logs within a research image sharing jurisdiction. The Audit Trail and Node Authentication profile of the Integrating the Healthcare Enterprise (IHE ATNA) defines one such standard for audit logs in the healthcare domain that should be strongly considered for this purpose.

Non-repudiation

Non-repudiation asserts that an authentication is genuine, and proves that data transmitted and its source is authentic. With non-repudiation, the system can be assured that a valid user actually made the request, and the data in the response is authentic and from the requested source. This is strongly related to the authentication and digital signature requirements outlined earlier.

Transmission Protection

Transmission protection ensures that data movement across a distributed network is secure. Mechanisms like encryption and signing of data may be applied to data transmission in a transient way. The data may be encrypted prior to transmission and decrypted immediately upon receipt. Similarly signature may be applied to ensure data authenticity.

Alternatively, the communication channel, rather than the message, can be encrypted. For example, HTTPS uses SSL to encrypt the connection between a web server and a web browser.

Secure Data-Sharing Environments

Security requirements outlined above may apply differently in different image sharing environments. In this section we briefly inspect the requirements for different data-sharing environments that focus on two topological aspects of the image sharing environments: centralized vs. distributed data repositories, and centralized vs. distributed security infrastructure. These two aspects can be mapped to different real world data-sharing environments as shown in the table below (Table 1).

Network-Based Data Sharing

These security requirements are affected by these topological layouts of the image sharing environment. Some requirement may be less stringent in one environment when compared with another. The table below outlines some of the specific differences in requirements compared with the centralized repository with a centralized security scenario (Table 2).

Data Sharing with Physical Media

Where online transmission of data is not feasible, physical media may be used for data transmission. This is often the approach used when transmitting large amount of data or when the electronic system is not reachable from the data requester. The data sharing by physical media scenario requires different process of enforcement of the security requirements outlined above.

The process of discovery and retrieval of the data from the image repository depends on the availability of a secure search interface to the data requester. If a secure search interface is available, the requester may search for data present in the system, following the electronic authentication, authorization, and audit logging requirements outlined above. Once the data have been identified, a separate request is sent to a data administrator at the repository site for the actual data. If a secure search interface is not available, the data requester in essence delegates the search activity to the data administrator, who then performs the search on behalf of the data requester and must enforce the authentication, authorization, and audit logging requirements. This second case is often the approach used by honest broker systems in information warehouses. For the remainder of the discussion,

Table 1 Requirements for centralized vs. distributed data repositories and centralized vs. distributed security infrastructure

	Centralized security infrastructure	Distributed security infrastructure
Centralized data repository	Hospital	Central (public) repository
Distributed data repository	Research institution with multiple research groups	Grid (caGrid), web services, internet, and research consortium

Table 2 Security requirements for different image sharing topologies

	Requirements	Central repository and distributed security	Distributed repository and central security	Distributed repository and distributed security
Data transformation	De-identification		Need consistent de-identification approach	Need consistent de-identification approach
	Research identifier		Need global identifier	Need global identifier
	Privacy preserving transformation		Consistent transformation	Consistent transformation
	Encryption		Encryption by each repository using same mechanism	Encryption by each repository using same mechanism
	Signature		Signature for each repository	Signature for each repository
	Honest Broker		Consistency between brokers critical	Consistency between brokers critical
Infrastructure	Policy management	Consistent policies in different systems and coordinate different systems		Consistent policies in different systems and coordinate different systems
	User identity management	Each organization manages own users		Each organization manages own users
	User role and attribute management	Cross-institutional roles important		Cross-institutional roles important
	Authentication	Authentication against local identity provider. Security token needs to be acceptable by all		Authentication against local identity provider. Security token needs to be acceptable by all
	Audit log management	Log may need to be managed by and potentially replicated at multiple sites. Log mining would require accessing multiple log repositories		Log may need to be managed by and potentially replicated at multiple sites. Log mining would require accessing multiple log repositories
	Trust management	Critical to have well established trust fabric between security components		Critical to have well established trust fabric between security components
Data access and movement	Authorization	May need to combine multiple authorization policies	Each repository needs to enforce authorization	May need to combine multiple authorization policies. Each repository needs to enforce authorization
	Delegation			
	Audit logging	May need to log to multiple log management services		May need to log to multiple log management services
	Non-repudiation			
	Transmission protection			

we will focus on this second case, with the assumption that the data requester has no access to the electronic data repository system and the administrator manages all interactions with the repository.

Once the appropriate data have been located, the administrator must prepare the data for placement on physical media. To protect the privacy and integrity of the data, the administrator would then need to de-identify the data according to the appropriate policies, and encrypt and sign the data if the data are not public. The data can then be placed on physical media.

The data requester, on receipt of the physical media, may access the data. This may require decryption and validation of the data signature. However, since no electronic exchange of the user credential or data repository credential was available, the decryption and signature validation of the data requires separate transmission of the appropriate certificates. The requester's certificate may be sent as part of the request

Table 3 Security considerations and process changes that accompany the use of physical media

	Requirements	Applicability to physical media transmission
Data transformation	De-identification	Must be performed by administrator prior to placement on physical media
	Research identifier	Must be performed by administrator prior to placement on physical media
	Privacy preserving transformation	Must be performed by administrator prior to placement on physical media
	Encryption	Must be performed by administrator prior to placement on physical media
	Signature	Must be performed by administrator prior to placement on physical media
	Honest broker	Administrator acting as honest broker
Infrastructure	Policy management	Administrator needs to be able to read and enforce policies
	User identity management	Administrator may manage the user identity on behalf of the requester.
	User role and attribute management	Administrator must track requester's role and attributes
	Authentication	Administrator must authenticate the requester through human readable communication channels
	Audit log management	Administrator must keep audit log
	Trust management	Trust is established between administrator and requester non- electronically
Data access and movement	Authorization	Administrator must perform the authorization for data access
	Delegation	A request delegate access request to the administrator, who then must apply the appropriate level of access
	Audit logging	Administrator performs the audit logging
	Non-repudiation	Administrator must keep record of data request, access, physical media packaging, and shipment
	Transmission protection	Administrator encrypts and signs the data prior to data shipment

Data transformation

so that the administrator at the repository site may encrypt the data with the requester's certificate. Each data requester therefore will receive data that only he or she can decrypt. The alternative of using a common data repository certificate for encrypting and decrypting data is not as secure as any data sets from the repository may then be decrypted by anyone who has access to the repository certificate.

Table 4 Technologies and standards that can be leveraged in the development of a secure image-sharing infrastructure

	Requirements	Available technology or standards
Data transformation	De-identification	CTP DICOM anonymization
	Research identifier	IHE PIX
	Privacy preserving transformation	application specific
	Encryption	X509 certificate, XML encryption, and PGP
	Signature	X509 certificate, XML signature, MD5/SHA1 sum
	Honest broker	
Infrastructure	Policy management	XACML
	User identity management	LDAP, active directory, OpenID, caGrid Dorian
	User role and attribute management	LDAP, active directory, caGrid GridGrouper
	Authentication	SAML, WS-Trust, DICOM, caGid Dorian
	Audit log management	IHE ATNA schema
	Trust management	caGrid grid trust service
Data access and movement	Authorization	SAML, OAuth
	Delegation	caGrid certificate delegation service
	Audit logging	IHE ATNA profile
	Non-repudiation	
	Transmission protection	WS-Security, HTTPS

The following table illustrates some of the security considerations and process changes that accompany the use of physical media (Table 3).

Existing Technology, Systems, and Standards

Security considerations are common to many domains that have data management and access needs, ranging from financial systems to healthcare enterprises. There is a significant amount of existing technology that addresses the individual requirements of a secure data-sharing platform. The table below lists some of the technologies and standards that can be leveraged in the development of a secure image sharing infrastructure. Please note that this is not an exhaustive list, and there exist significant efforts in W3C, caBIG, DICOM, HL7, and IHE communities that can be leveraged here (Table 4).

Conclusions

Image sharing for research use is a complicated process. As image management systems have focused on clinical data management and therefore usually focus on data sharing within a single jurisdiction, the requirements of cross-jurisdictional data and image sharing are often not met.

In this paper, we outline common requirements for research image sharing from the security perspective: how to transform the data securely for research, what the infrastructure needs to provide to create a secure data-sharing environment, and how to secure data access and movement in this infrastructure. Specifically, we view the security requirements as a consequence of data movement across jurisdictional boundaries.

These requirements are not trivial. In a secure image-sharing infrastructure, we may need to de-identify the data, providing user authentication capabilities, manage access rights, create audit logs, and encrypting the data on transmission. These are requirements that may not necessarily be met by any existing systems. We hope that by outlining the requirements for secure research data sharing, we create a common language and thought framework in which to discuss information system integration for research data access. We also hope that these requirements will motivate the development of a system that can address these requirements to support research image sharing.