Information Systems and Healthcare XVII: Operational Stakeholder Relationships in the Deployment of a Data Storage Grid for Clinical Image Backup and Recovery

Nelson E. King
American University of Beirut, nelson.king@aub.edu.lb

Follow this and additional works at: https://aisel.aisnet.org/cais

Recommended Citation
Available at: https://aisel.aisnet.org/cais/vol23/iss1/1

This material is brought to you by the Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Abstract:
A data storage grid (DSG) is under development for a federation of clinical sites to provide a cost-effective backup and recovery solution for their clinical images. Geographic separation provides fault-tolerance against localized disasters. Pooling of storage resources across organizations utilizes economies of scale associated with storage area networks. However, the control and administration of a DSG is now spread across multiple organizations increasing the complexity of deployment. Socio-technical issues specific to a DSG arise as there are now multiple stakeholders linked together in a network of new relationships. Agreement upon every relationship is necessary to determine service level agreements, security, and liability such as in the event of a security breach. Implications of socio-technical networks and stakeholder analysis on the operators, rather than the users, of an interorganizational DSG are discussed.

Keywords: grid computing, picture archiving and communication system (PACS), data storage and retrieval, disaster planning, socio-technical networks, stakeholder analysis
I. INTRODUCTION

Filmless medical imaging is no longer confined to the radiology department of hospitals. Digital imaging is now integral to the operation of trauma centers, image-assisted surgery, radiation oncology, and other minimally invasive procedures relying upon clinical images. Picture archiving and communication system (PACS) implementations are extending imaging services across the healthcare organization [Bucsko 2007]. Clinical operations are now dependent on digital storage devices.

An unexpected downtime event with PACS and its associated DICOM archive storage can now place patients at risk due to the inability to access or even the loss of historical image exams. DICOM is the digital imaging and communications in medicine standard employed in all digital imaging devices. PACS handled around 273 of 611 million radiology procedures performed in the United States in 2005 [Frost and Sullivan 2006]. Primary PACS storage was over 13,000 Terabytes and volume expected to grow over 25 percent annually from 2005 to 2012. Keeping up with this growth rate in storage capacity will compete for resources necessary to fund adequate backup and recovery for PACS and DICOM storage archives.

A solution has been proposed for the challenge of providing an inexpensive but high availability fault-tolerant clinical image backup and recovery in the form of a DICOM-compliant data storage grid (DSG) [Liu et al. 2005]. A DSG is a variant of grid computing which is the integrated use of geographically distributed computers, networks, and storage systems to create a virtual computing system [Foster 2002]. A data grid creates a virtual storage system that intentionally crosses organizational and geographic boundaries to share resources and facilitate collaboration. A data grid can be built upon heterogeneous platforms and is easily scalable.

When multiple organizations contribute to a grid and use it, questions are raised from an operational perspective such as: “Who manages the data grid?” “Who owns the grid and therefore is responsible for maintaining the grid?” “Who repairs a failed component of the grid?” The grid is by design amorphous which means there is no single system sitting in a specific location. Elements of the grid are spread across all of the facilities in the virtual federation. The benefit of sharing storage space across organizations does introduce ambiguity into operation of the grid. But perhaps a more important issue is the question of liability. What happens if protected health information is compromised at the site of a federation partner which results in a Health Insurance Portability and Accountability Act (HIPAA) violation? Who is liable, and what are the safeguards?

These questions arise when the infrastructure for an information system extends beyond the control of a single organization and location. This paper explores some of the socio-technical ramifications of deploying a DSG across organizational and geographic boundaries. Rationale for using a framework of stakeholders in a social network follows. Section III describes the DSG concept, methods of stakeholder analysis, and the kinds of relationships identified in this DSG. Section IV identifies the stakeholder relationships necessary to implement and operate an inter-organizational DSG and some design implications of these relationships. Finally, the design implications of these interrelationships framed as a socio-technical network are discussed in Section V.

II. BACKGROUND

A radiology PACS integrates the digital technologies that acquires, transmits, stores, retrieves, and displays digital images and related patient information from a variety of imaging sources to provide a complete solution for radiology image management. Forty-one percent of nonprofit hospitals have adopted radiology PACS although only 18 percent of for-profit hospitals have done so in a 2004 survey [Fonkych and Taylor 2005]. The prospect of backing up terabytes of imaging data annually and making them available under different disaster scenarios is a daunting task.

Storage system availability is essential in protecting the health and safety of patients while avoiding disruption of clinical activities and the loss of revenue from cancelled procedures [Anderson 2002]. This section describes the use of grid technology and the utility of a distributed data storage grid. The underlying framework used for analyzing stakeholder relationships in the DSG as a social network is also explained.
Grid Technology in Healthcare

Grid technologies are found in various healthcare applications. These include national health information sharing in Canada [Bilykha et al. 2003], medical imaging analysis [Germain et al. 2005], eDiaMoND for breast cancer screening and a federated database of annotated mammograms [Jirotka et al. 2005], biomedical modeling [Breton et al. 2005]. Alternative conceptualizations of using grids for data storage include the byCast grid for the BC Cancer Agency [Silk et al. 2003] and Globus MEDICUS [Erberich et al. 2007].

Most grid implementation papers focus on the technology and ignore the day-to-day implementation and operational issues. One of the few papers to address operations in a grid focuses only on data management [Hoschek et al. 2000]. There is no mention of interorganizational relationships or social aspects of operations. The only grid article to consider socio-technical issues is by Jirotka [2005], who explored trust and ethical issues of physicians who have access to sensitive health information in eDiaMoND.

Data Storage Grids

In a typical clinical implementation of PACS, maintenance and support for backup and recovery of clinical images is performed by the PACS administrator or the clinical IT operations staff. PACS and storage hardware are often placed in a single data center. Backup copies of images are sometimes placed on tape and less frequently sent to a remote storage site offered by an application service provider. Under typical conditions, any local disaster such as loss of utilities in the neighborhood or loss of air-conditioning to the servers results in PACS downtime. If the disaster causes physical damage to storage resources, the files must be restored. Recovery from tape backup is known to be slow and failure-prone resulting in additional delays [Staimer 2005]. More robust solutions are available such as clustered servers or a second data center, but these solutions are expensive and the recovery process is tedious. These are some of the reasons an alternative low-cost storage solution for backup and recovery, like a DSG, are attractive.

A data storage grid (DSG) adopts the principles of distributed and heterogeneous networks used in grid computing. Storage resources at participating sites are networked through middleware to function as one storage infrastructure [King et al. 2005]. The networked infrastructure approach is transparent (e.g., grid virtualization) to the radiologists and clinicians who are the users of these storage resources. The users see the DSG solely as just another DICOM application entity device in which clinical images are stored. The clinical images continue to be retrieved through software applications that are compliant with the Digital Imaging and Communications in Medicine (DICOM) standard. After receiving a DICOM image request, the provisioning aspect of grid middleware determines which storage location provides the fastest retrieval among the participating sites. This is in contrast to a DICOM only solution where the user would explicitly request a retrieval from a specific storage location whose availability is not known until attempted.

DSG as a Social Network

Systems, especially interorganizational ones, do not exist in social or technological isolation. A socio-technical network describes the arrangement of interactions between people, organizations, institutions, and a range of technologies [Lamb et al. 2000]. Social network analysis has been used to frame and describe organizational forms and structures [Zack 2000]. Pouloudi et al [2004] follow in this mindset and argue that health informatics is a “pluralistic discipline . . . that of necessity integrates both social/human and scientific/technical disciplines” (p. 8).

Grid systems involve diverse stakeholders, not just the end users who might be brought together because of shared data. Whenever a system is distributed across organizational boundaries, there needs to be some way to have all parties act for common objectives and coordinate across these boundaries. The coordination mechanisms must be arranged during design and implementation, so the effects of the socio-technical network on the social and technical entities are understood [Hansen, 2004]. A means to exercise “collaborative coordination” during implementation is also required [Kirsch, 2004]. This coordination must continue into day-to-day operations.

The technologies that link grid member organizations together include the standards and protocols embedded in the grid middleware. However the underlying organizations must also be linked socially. Stakeholders are presumably believed by grid developers to have their goals and priorities aligned with the relationships dictated by the transfer of data within grid middleware. Technology trust evolves into relationship trust [Ratnasingam 2005]. This line of thought says that transmitting the data reliably and safely equates to the data being used and protected in the same way.

This paper analyzes the social network relationships of stakeholders linked by grid technology. Social networks are analyzed on the basis of transactional content, nature of the links, and structural characteristics of the relationships [Tichy et al. 1979]. Since grid technologies are interorganizational, such entities also can be characterized by the
degree of differentiation between units to be coordinated, intensity of interfirm interdependence, and number of units to be coordinated [Grandori and Soda 1995].

Stakeholders in a Social Network
The goal of stakeholder analysis is to identify and understand the organizational roles present in the deployment of a new system. These roles extend beyond the common knowledge that users and developers are key participants in system development. The roles of many parties were ignored when the scope of systems were broadened across intra- and interorganizational boundaries [Pouloudi and Whitley 1997].

Few information systems are designed with an understanding of users and the ways that users interact with information or expectations of the embedded workflow. Information systems that incorporate different opinions, different focus and different conceptions of clinical administrators are rare [Anderson 2002]. A grid architecture is different from the single site storage model and its unfamiliarity will lead to questions. What happens when the introduction of a DSG forces administrators to wrestle with questions of IT infrastructure that many do not value in the first place? In the case of a DSG, the infrastructure is amorphous, belonging to everyone, but at the same time no one. How does one explain grid computing to clinical administrators with little IT familiarity?

This paper introduces the notion that design choices can in fact alter the power and relationships of stakeholders. Keating [2000] says that patterns of structure, although not known precisely in advance, emerge through operation of a system (e.g., the DSG in this case). An operational DSG does not have the luxury of waiting for structure to evolve. The relationships and interactions between partners of the federation must be known in advance because the technical implementation choices required to configure a DSG drive the linkages and relationships of the operational system. For example, technological connections of networked (grid) computing impose some form of structure on organizational relationships. In the simplest grid configuration, clinical site A is exclusively connected to sites B and C until other partners join the federation. Images of site A are stored on sites B and C. The need to store protected health information outside of one’s organization limits the connections of the DSG to those can be trusted. Administrators and IT operations staff at each site must interact with each other because their sphere of influence and responsibility for stored images cross organizational boundaries. These relationships are embodied in the technological connection between each of the sites. Emergence of a post-implementation change in the relationship, such as withdrawal of a site, puts the entire technological configuration at risk.

III. SYSTEM DESCRIPTION AND METHOD

System Description
A DICOM-compliant data storage grid (DSG) prototype that increases storage system availability while also providing redundant access to clinical images during recovery from unexpected PACS failures or other disaster scenarios has been fielded at the University of Southern California Image Processing and Informatics Laboratory (USC-IPI) [Liu et al. 2005].

This DSG concept leverages grid computing and the incremental costs of adding additional storage capacity in a hospital’s storage area network (SAN). Three or more clinical operations form a federation and swap storage space among themselves creating a low-cost distributed and networked storage system with high availability and fault tolerance as shown in Figure 1. The participating clinical sites benefit by avoiding the cost of a dedicated backup system or redundant data centers for their clinical images. The DSG also allows multiple paths to retrieve images if there is local congestion.

Grid computing is based upon utilization of heterogeneous resources with the grid middleware providing connectivity between organizations. Resources such as storage and network bandwidth are dynamically allocated from a pool of resources [Berstis 2002]. While grid computing allows nearly unlimited pooling of resources, there are practical limits to the number of sites that can participate in a DSG federation. The limit, discussed later, reflects the need to build and maintain social relationships with partners whom one can trust one's sensitive clinical image data.

The DSG concept of Liu et al [2005] is illustrated with three clinical sites for the purpose of simplicity. A university research hospital asks a community hospital and an outpatient center to join together and trade extra data storage capacity for the purpose of creating a DSG. Each clinical operation has a SAN with short-term and long-term storage. Typically a PACS first place images in short-term storage (“S” on the figure). Over time, those images in short-term storage are migrated to long-term storage (“L” on the figure). Unassigned space on the SAN of one partner is allocated to the other two partners. The university hospital hosts storage for the community hospital as well as the outpatient center. The community hospital and the outpatient center provide equivalent space to the
The choice of federation partners includes considerations such as the likelihood that both alternative locations would be disabled by the same disaster.

**Legend**
- RH: Research Hospital
- CH: Community Hospital
- OC: Outpatient
- SAN: Storage Area Network
- S: Short-term Storage
- L: Long-term Storage
- Grid Assets
- Query/Retrieve
- Data Storage Grid
- Research Hospital
- Community Hospital
- Outpatient

**Figure 1. Backup and Recovery with the Data Grid**

The "grid" consists of all the shared storage space assets at each of the clinical sites and the connecting hardware and middleware. The SAN of a federation partner physically hosts both storage allocated for the DSG and its own PACS short- and long-term storage. Measures to provide privacy and authenticity are necessary for any data storage such as access control. In a DSG setting, preventing the inter-mixing of data should be an operational concern to prevent inadvertent access and simplifying the termination of a DSG agreement. Physical security measures available on a typical SAN include separate partitions in the SAN, IP addresses, and even physical disks. Separate IP addresses (network cards) minimizes packets “sharing the wire.” A separate IP address is also easier to protect on the network from unauthorized or unintended access. At the extreme, the network connection could be from a separate ISP to provide additional protection for data access in case of a local disaster.

The specific security measures chosen by a federation partner are dictated somewhat by the particular capabilities of a storage system and the expectations of the owner of the data. For example, the prototype did not enable secure tunnels or encryption of data because the physical equipment was old and slow, and the anonymized test data never left the laboratory. In addition to concerns about privacy, an additional aspect of data security is image integrity. The images themselves should be protected from any alteration in the unlikely event that someone gains access to the physical images at both the local site or of federation partners. One method is Lossless Digital Signature Embedding (LDSE) [Zhou 2007; Lee et al. 2008].

For all federation partners, having an easy way to purge the data reduces their exposure to liability. Termination of participation is simplified by physically removing the disks so that all parties know there are no residual images left at a federation partner site. Data in a partition can be deleted but some information security officers might want the re-formatting of the physical media before restoring to service or even the physical removal of the media. This is the reason for storing images on a separate physical disks. The actual number would depend on the RAID level and the capabilities of the SAN software. This separation of data would also have to be applied to the backups of the SAN.

The backup and subsequent retrieval of images in a DSG is different from a backup system for a single site. The routing of images from the perspective of the research hospital is shown in Figure 1. The research hospital generates an image exam from the modality (e.g., MRI or CT imaging devices) which is sent to the PACS (not shown in the figure). The PACS sends the image exam to the short-term archive and simultaneously sends two additional copies of the image exam to the DSG. The first one goes to the storage space reserved for the research hospital that is physically located at the community hospital. The second copy goes to the outpatient center.
there are now copies of this image exam at three geographically separated locations, the likelihood of being able to retrieve an image is very high. Ordinarily the research hospital retrieves images from its own PACS short or long term storage archive which is typically faster. However, if a disaster occurs, there are two more copies of the image exam available at the other two sites on the DSG.

When an image exam needs to be retrieved during a disaster situation at the research hospital or a local outage, the DSG processes the DICOM query and retrieve command from the site on the DSG that has the best performance. In this scenario, the DSG offers a level of business continuity assuming the hospital still has connectivity to the internet. The physical location of the image is transparent to the user who requested the image exam. As an example shown in Figure 1, the image exam is sent back to the requesting party at the research hospital from the DSG portion of the SAN located at the outpatient center. While the query and retrieve commands are DICOM, the routing and transporting of the data is carried out by the grid middleware.

The shared storage resources of the DSG are tied together by USC IPI integration software that builds upon open-source grid middleware software from The Globus Alliance (e.g., Globus Toolkit) [King et al. 2005; Liu et al. 2005]. The integration includes connections to DICOM compliant devices on the clinical side, a meta-data catalog for the DSG, and integration of the SAN devices with Globus services. The pre-prototype used GT 3.2 Core, Security Infrastructure, System level services, GridFTP, Reliable File Transfer Service (RFT), and Replica Location Service (RLS). The follow-on prototype moved to the equivalent GT 4.x services.

The typical clinical implementation of PACS places responsibility for both system maintenance and the backup and recovery of clinical images upon the PACS administrator (e.g., operator) or the clinical IT operations group. Operations personnel must justify expenditures on backup equipment and provide the level of service expected by clinical administrators. The level of service is dictated by the users (e.g., system too slow or not available). Issues such as image security are handled by the entire clinical organization. Compliance with the Administrative Simplification provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) require careful attention to security and the protection of patient health information.

Data Collection Method

More than a year of active participation in every architectural discussion and code review is reflected in this article. The focus is on the architectural tradeoffs that influenced the DSG design rather than actual implementation details that would vary between sites. Data consisted of a mix of participant-observation notes and team developer notes. The author led the team that demonstrated the DICOM Data Storage Grid (DSG) concept at the Radiological Society of North America (RSNA) 2004 meeting. After this demonstration, the author moved into a mentoring role for the leader of the development team who took over as team leader. The author conducted several follow-up interviews to clarify what had taken place after the author was no longer involved. The development team leader was interviewed in December at RSNA 2005 during the demonstration of the subsequent DSG prototype version. One of the federation team members was interviewed in June 2006 followed by another interview of the development team leader in August 2006.

Stakeholder Analysis

The goal of studying stakeholders is to understand their different perspectives which can then be incorporated into the design of the system. Stakeholder analysis is a first step towards understanding the interorganizational relationships contained in a DSG. Information system projects are often abandoned because of stakeholder issues [Pan and Flynn 2003; Berghout and Remeny 2005]. Stakeholder representation [Pouloudi et al. 2004] leads to the identification of non-obvious interested parties. Stakeholder analysis shows how stakeholders are varied in their power to affect the use of the system, and in their interest toward its use [Boonstra 2006]. The effect of stakeholders on a firm depends upon the network of stakeholders surrounding the relationship [Rowley 1997].

The analysis in this design implementation followed the four principles of stakeholder analysis developed by Pouloudi and Whitley [1997] in their study of an interorganizational drug management system. These principles are: stakeholders depend on the specific context and time frame, stakeholders cannot be viewed in isolation, position of the stakeholder may change over time, and feasible options may differ from the stakeholders' wishes. Stakeholders are viewed as a network of relationships rather than isolated groups. The identities and roles of stakeholders are also not static. Finally, the analysis recognizes that all stakeholder wishes cannot be satisfied.

When looking at design from a socio-technical perspective, new design issues arise. Mere identification of stakeholder groups is insufficient. Analysis should also consider stakeholder motivation and needs, evaluation of power held by each stakeholder group, linkages between stakeholders, and how these groups influence strategic decisions in their respective organizations. The typical stakeholders in the deployment of a traditional clinical image
information system are the hospital administrators, the information technology (IT) staff who administer and operate an information system, and the clinical users. Note these are roles not necessarily individuals (see Sharp et al. 1999).

The boundary of this study was limited to the operation and management of a DSG that provides backup and recovery to a PACS. The PACS has direct contact with clinical users but the DSG does not. As a consequence, end user (e.g., radiologist) demands for fast response and high availability are treated as inputs to the design process rather than stakeholders.

IV. ANALYSIS AND DESIGN IMPLICATIONS

The adoption of a DSG means stakeholders in various organizations are now linked together over issues such as security, service-level agreements, and liability. Most organizations have a component-centric view of managing the data center [Strong 2005]. In addition, human and nonhuman stakeholders as well as the dynamism of the networks must be considered [Pouloudi et al. 2004]. There are human issues such as allocating responsibility for managing and operating the hardware and software of the grid. There are policy issues with regards to how much storage will be shared and how much bandwidth can be used by the federation. There will be administrative issues with regards to allocation of costs and protection against liability. There will also be issues as the DSG grows requiring new capacity and perhaps the addition of new federation partners. Grid computing adopted for use in a clinical DSG requires a more systemic and holistic approach to management. Therefore a DSG results in a combinatorial increase in networked stakeholders that must be considered as well as a change in orientation of managing IT resources.

DSG Stakeholder Relationship Analysis

The analysis of this three-clinical site DSG showed the intertwined nature of stakeholder relationships that arise when an information system is deployed across organizational boundaries. The analysis draws upon the network theory of stakeholder influences [Rowley 1997]. Three kinds of relationships exist which include the existing relationships that exist within an organization. New primary relationships arise from hosting the DSG across three organizations. These relationships are needed for the federation to function. There is also the potential for new secondary relationships across the three organizations depending upon design choices made. Adopting a three-site DSG adds nine new primary relationships and the potential for six optional secondary relationships as shown in Figure 2. Adding additional sites result in a combinatorial increase in the number of relationships that have to be managed.

Existing Relationships

Each site of the DSG federation already has clinical administrators and operations personnel. The typical relationship revolves around clinical administrators wanting faster delivery of images at a reduced cost while preserving the safety and security of the images. Operations personnel argue that faster and safer delivery of images requires an investment in computing infrastructure. These existing relationships are marked by the double line and the hexagons labeled 1-1, 2-2, and 3-3 in Figure 2.

Primary Collective Relationships

All the operations staff in the grid federation, shown as circle 1-2-3, represents the first primary relationship. They must come up with a collective position on the overall operations aspects of the DSG federation: security, service level agreements, hardware and software configurations, acceptable equipment, etc. The capacity and performance of shared resources from each site are likely to be different requiring negotiations on expected capabilities. For example, site 1 might desire image retrieval speeds equivalent to spinning disk, but site 2 may be using slower optical media. This collective set of operators must also agree on a minimum level of security and design the necessary security infrastructure. The agreement typically involves sharing the details of each sites’ security architecture so that potential vulnerabilities to the DSG federation can be examined. In this DSG, the collective operations staff must also come to agreement with the collective administrators on overall service levels and operational policies. One such agreement is to trade equal amounts of storage space.

All administrators in the grid federation, shown as rhombus 1-2-3, represent the second of nine primary relationships. They must develop a collective position on the overall administrative aspects of the DSG federation: security, service level agreements, mutual assistance, and liability. These agreements will most likely establish the minimum or common denominator expectations of the DSG. The collective set of administrators set the basic terms of the operations of a DSG drawing heavily upon their own operations staff. Each site administrator must then decide that sufficient benefit accrues to their own organization to warrant the effort, resources, and risks necessary to participate in the DSG. This collective body determines the allocation of roles and responsibilities for operating and maintaining the DSG among the various sites.
The collective relationship between DSG administrators and DSG operations personnel represented by hexagon 1-2-3 is the third new relationship. The very nature of a DSG is that middleware connects together three or more existing (and heterogeneous) computing infrastructures. Just as administrators and operations personnel of a single site must reconcile differences between administrative goals and operational realities, so must the DSG administrators and DSG operators come to agreement on similar policies. This relationship would be the place to discuss the collective administrators desire to have parity in response times. For example, one site might provide millisecond response times from spinning disks but another site might have slower optical media or even magnetic tapes. This means comparisons must be on a relative basis as there are unlikely to be identical computing infrastructures. In contrast to a single site or organization, these collective discussions may revolve around what are equivalent levels of performance and security.

Primary Site-to-Site Relationships

The specific agreements between administrators of each pair of clinical sites must also be worked out (e.g., three pairs of relationships shown as rhombus 1-2, 2-3, 1-3). The tone of the relationships and subsequent policies between specific members of a DSG federation will likely reflect prior relationships between these clinical sites. Trust and equitable power relationships are assumed to pre-exist as a prior condition to joining the federation (e.g., [Bunduchi 2007]). The agreements between administrators of different sites do not have to be the same. For example, the research hospital may already have a formalized relationship with the outpatient center while the basis of the relationship with the community hospital could be prior collegial professional activities that results in a memorandum of understanding.

Interactions between operations staffs in each organization account for the last three of nine primary relationships. The specific agreements between operations staff of each pair of clinical sites must also be developed (e.g., three pairs of relationships shown as circle 1-2, 1-3, and 2-3). These agreements are refinements of the collective position which can vary in detail among specific partners. One site may be comfortable with the security provided by the SAN software. Another site may wish the partner to go a step further and configure the SAN so that there is no mixing of data on physical hard drives.

Potential Primary Relationships

Optional relationships are theoretically possible but can be avoided by certain choices in design. Minimizing the number of new relationships reduces the coupling between organizations and complexity in communications. Not shown in this figure are relationships to stakeholders who are external to the operation of the DSG such as users (e.g., radiologists) and those parties concerned with regulatory compliance (e.g., HIPAA). These stakeholders
would then need to bring up their issues with the newly created collective stakeholders. For example, radiologists from site 1 demanding fast response times from the backup sites must now deal with the collective group of operators to see what is technically feasible. In addition, the collective group of administrators and operators would have to agree upon an operational policy. Only then can specific agreements be reached between site 1 and their federation partners. For example, site 1 administrators might have to pay site 2 administrators to upgrade the storage area network.

These potential primary relationships between end users and operators were avoided in the early design work by designating this DSG for backup and recovery purposes, rather than as an alternate site for active storage. Additional negotiations over operational service level agreements and the impact on each other's SAN and internet resources would have been necessary. A DSG used as active storage would further complicate the negotiations as the impact of operations Any operational capability (e.g., continuity) that could be gained from the DSG during a local disaster, such as a storage outage, would then be seen as a bonus not an expected service.

None of these potential primary relationships had to be considered in the design of this DSG prototype.

Potential Secondary Relationships

A secondary relationship reflects the desire of one site to interact with a member from another site even though the collective groups have already established the basis of a DSG agreement. The potential for secondary relationships are illustrated with a scenario likely to be encountered once a clinical DSG goes into operation. The DSG takes advantage of economies of scale by trading excess storage capacity at each site. However, storage needs for clinical images often grows faster than expected. Trading space reduces the complexity of relationships. However, one site is likely to have greater storage needs over time than the other sites. For example, five terabytes were traded when the DSG federation was established. However, site 1 now needs 10 terabytes but sites 2 and 3 do not need a larger allocation.

Sites 2 and 3 will want compensation to host additional storage space for site 1. When storage space was traded equally, the relationship between operation staffs focused on connectivity and sufficient performance. Administrator involvement was previously limited due to the agreement that storage space was being traded equally.

Now that site 1 needs more space two optional secondary relationships are activated. The administrators at site 1 must now interact at least indirectly with operations staff at site 2 and 3 on either compensation for additional storage space or purchase of additional equipment. Since there will be a financial outlay, site 1 will want to influence the purchasing decisions. The interaction may be indirect since the administrators at site 2 might want all communication between site 1 administrators and their operations staff to pass through them. However, the interaction still occurs.

The previous scenario of adding additional capacity is possible when sites 2 and 3 can easily increase additional storage space. Space might be added by re-allocating existing space, adding hard disks to empty slots in the SAN or adding another SAN controller module when vacant rack space exists. The cost of implementing these solutions is easily quantifiable. The DSG is still operating with a "trading space" mentality and marginal cost economics are still in play until reaching the maximum capacity of the SAN.

An Unstable Federation?

Just looking at the intertwined stakeholder relationships in a DSG may at first glance make deploying a DSG appear prohibitive. However, the benefits to each of the clinical sites through participating in a DSG have to be weighed. The cost of a wholly-owned backup and recovery system, such as a second site, would be prohibitively expensive. This level of expenditure for a second site is typically incurred only by firms whose data is not only mission critical but also requiring highly synchronous transactions. For example, financial firms after the 9/11 experiences in downtown New York built second data centers nearby but outside of New York. While clinical images are mission critical and fast retrieval is necessary, there are options to achieve the redundancy and speed in DICOM image transfer not available to sophisticated and highly synchronous infrastructure needed for real-time business transactions. A clinical site on a tight budget would not have the resources to adopt the backup practices of a financial institution. By partnering in a DSG, each member of the federation benefits from higher availability, fault-tolerance, and geographic separation of storage resources. The entry costs of a DSG are expected to be much lower and thus attractive. The premise of a DSG is the availability of unused capacity in a SAN which can be added for just the cost of additional storage media.

The DSG federation comes under pressure when the need for additional space requires major upgrades to either the SAN or other storage devices. What happens when additional (expensive) rack space must be procured? Procurement of a rack and controller is no longer a marginal cost situation. Federation partners may or may not
want to contribute to an upgrade of SAN storage capacity. For example, site 3 needs to purchase an additional rack and SAN controller modules to satisfy the request of site 1 for more space. What proportion of costs does site 1 pay site 3 for an expensive rack of greater benefit to site 3?

The negotiations must come from a holistic mindset: a degree of mutual aid may be necessary to keep the DSG federation healthy. Sites with growing needs may need to provide subsidies or purchase equipment for a federation partner with less ability to expand storage capacity. Each site has a vested interest in making the DSG attractive to the other sites so that the federation can continue.

A second situation where relationships in the DSG federation will be tested is in the situation of a security breach. There could be a breach at one site that changes the level of trust predicated in the original negotiations. Agreements governing relationships might have to be renegotiated such as allocating time for IT staff to check on the security of the DSG. Additional safeguards might need to be imposed. Federation partners with more security expertise may need to contribute resources to strengthen this federation partner so the overall DSG is more secure.

These examples of additional storage space and a security breach show that the interrelationships required for the DSG to function smoothly are in a state of equilibrium that can be thrown out of balance. There is a mutual dependence in the relationships that govern the DSG. Extra planning and analysis is necessary during the design phase to examine the potential cost and design implications associated with mutual aid.

**Design Example—Limit to Number of Sites**

There are practical limits to the number of sites that can participate in a DSG. The key considerations are the number of backup archive copies sent to the DSG and the available data streams for subsequent recovery for either a disaster or temporary overload of local services. These in turn determine the number of sites for which relationships must be established to build a DSG federation. An archive in medical imaging parlance is the storage repository to which actively used clinical images are committed after modality post-processing, not off-line archival as in storage management usage of the term.

<table>
<thead>
<tr>
<th>Table 1. Trading Space for Two Copies (5 Terabytes each)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage (TB)</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td><strong>3 Site</strong></td>
</tr>
<tr>
<td>DSG Space Contributed by Site 1</td>
</tr>
<tr>
<td>Site 1 images on Site 2</td>
</tr>
<tr>
<td>Site 1 images on Site 3</td>
</tr>
<tr>
<td><strong>5 Site</strong></td>
</tr>
<tr>
<td>DSG Space Contributed by Site 1</td>
</tr>
<tr>
<td>Site 1 images on Site 2</td>
</tr>
<tr>
<td>Site 1 images on Site 3</td>
</tr>
<tr>
<td>Site 1 images on Site 4</td>
</tr>
<tr>
<td>Site 1 images on Site 5</td>
</tr>
<tr>
<td><strong>7 Site</strong></td>
</tr>
<tr>
<td>DSG Space Contributed by Site 1</td>
</tr>
<tr>
<td>Site 1 images on Site 2</td>
</tr>
<tr>
<td>Site 1 images on Site 3</td>
</tr>
<tr>
<td>Site 1 images on Site 4</td>
</tr>
<tr>
<td>Site 1 images on Site 5</td>
</tr>
<tr>
<td>Site 1 images on Site 6</td>
</tr>
<tr>
<td>Site 1 images on Site 7</td>
</tr>
</tbody>
</table>

The DSG prototype was built for the minimum number of sites that could demonstrate grid functionality. For illustrative purposes, five terabytes are needed for each archive copy sent by site 1. In a three-site configuration, each site sets aside 10 terabytes of traded space for use by the DSG as shown in Table 1. Each archive copy is solely hosted by one site in the minimum DSG configuration. If greater protection against an outage is needed, additional sites can be added so only a portion of the archive is on any one site. A five-site configuration places one-half of an archive copy on each site while seven sites permits only one-third of an archive copy to be located on any one site. The net contribution to the DSG remains the same (e.g., 10 terabytes). However the number of potential stakeholder relationships that have to be managed increases significantly.

When a third archive copy needs to be held on the DSG, a minimum of four sites is necessary as shown in Table 2. The contribution of each site is now 15 terabytes. A further degree of safety can be added by placing one-half of an archive copy on each site as shown in the seven-site configuration.
Table 2. Trading Space for Three Copies (5 Terabytes each)

<table>
<thead>
<tr>
<th></th>
<th>Storage (TB)</th>
<th>% of Site 1 Copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Site</td>
<td>DSG Space Contributed by Site 1</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 4</td>
<td>5</td>
</tr>
<tr>
<td>7 Site</td>
<td>DSG Space Contributed by Site 1</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 2</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 3</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 4</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 6</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Site 1 images on Site 7</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Going beyond three archive copies or splitting an archive copy further among additional sites has increasingly limited payoffs in terms of data retrieval safety at the cost of increasing the number of relationships to be managed. While these examples show equal sharing among the sites, there is no technical constraint on different storage allocations among federation partners. The amount of storage traded between sites is individually negotiated.

**Design Example—Meta-Data Catalog**

The insights developed about organizational relationships from this stakeholder analysis provide more than an operations strategy. The analysis raised several architectural challenges. Grid computing implies a distributed architecture. Since no one member of the federation owns the grid, who is responsible for developing and operating common resources? One example in particular illustrates these challenges that have to be considered. Where would the meta-data catalog be physically located in this operational DSG?

The design considerations for the DSG meta-data catalog was influenced by examining stakeholder relationships in a DSG. The DSG is independent of a clinical PACS (i.e., separate DICOM application entity) which means a DICOM-compliant DSG must have an independent mechanism to query and retrieve based upon the DICOM meta-data. A DSG that receives a DICOM image must catalog two kinds of information. First, the DSG uses the replica location service to catalog the physical location of a particular image file. Second, the DSG parses the DICOM header and stores the necessary image attributes in a database so the image is searchable by DICOM compliant devices independent of a PACS.

The meta-data database for a DSG could be implemented as a centrally managed database. The database could be backed up using conventional techniques such as mirroring or replication. However socio-technical analysis asks the question: Who manages the meta-data database? If the meta-data database is hosted and maintained by the research hospital member of the federation, the two other partners (e.g., outpatient center and community hospital) now require service level agreements with the research hospital. These negotiations could become complicated if there is concern on the backup plan of the research hospital. For example, the research hospital might be situated in an earthquake-prone area. Mirroring the database at the same site does not provide geographic separation. Replication across one or more distant sites is possible but adds complexity and requires identical hardware and software to minimize operational problems.

Socio-technical analysis suggested that “trading” storage space reduced the complexity of the interrelationships among the federation partners. In the same manner, trying to reduce the complexity of relationships of a hosted meta-data database suggest that everyone hosts a copy of the database. In essence, each site has a database that contains the entire meta-data content of the DSG. Each federation partner could invest as much resources to make this database fast enough for the needs of the site. At the same time, the database hosts all the meta-data of the other sites. When the inevitable periods of downtime occur, DICOM queries from a site are routed by the DSG to the meta-data catalogs at other sites. Redundancy, geographic separation and pooled resources are now possible without a centralized database.

The architectural approach is straightforward. When the DSG parses the information from an image, the meta-data is sent as a message asynchronously to each of the separately hosted meta-data catalogs. In such a configuration, three identical messages are sent out. Once the message is processed at each meta-database there will be three catalogs with identical information though not identical databases (i.e., transactions will not be recorded in every
Repetition of the header information is achieved by duplicate transaction detail messages rather than processing the transaction log of a master database.

The uncoupling of the meta-data catalogs means that each site could store the information on their preferred database. If a traditional replication strategy were employed, all sites would have to use the same software vendor. It would be unlikely that all sites would agree on the same database vendor. Implementing such a strategy was beyond the resources of the prototype effort since the Globus Toolkit had no meta-data catalog service to use. Further consideration of the meta-data catalog is needed for a clinical DSG.

V. DISCUSSION

The DSG represents a socio-technical network consisting of interactions between people, organizations, institutions, and a range of technologies. People directly involved with the DSG are in roles as clinical administrators or IT operations staffs. They work for different clinical organizations with their own priorities and resource limitations. The technologies of grid computing are state-of-the-art and virtualized network services are a paradigm bound to be a stumbling block between clinical organizations. In addition, the hardware and software used at each site is likely to be different consistent with the philosophy of heterogeneity in grid computing. The deployment of this DSG across three organizations vastly increases the complexity of the network with respect to a single organization data center. Going beyond the minimum three DSG partners would further increase the complexity of the network.

How these designers found harmony amongst the disparity of this network was the motivation for a retrospective analysis of this stakeholder network from a socio-technical perspective. At first glance, this DSG seemed to add a great deal of complexity in stakeholder relationships (Figure 2). Renting storage space from a provider of backup systems would certainly be the simplest in terms of relationships that have to be managed. Looking beyond the typical solutions allowed for the consideration of the theoretical benefits of a DICOM-compliant distributed and fault-tolerant storage resource.

“Trading” Space Metaphor

What became evident from analyzing the creation of this DSG is the way that the design team implicitly configured the architecture to avoid overbearing relationships and strengthened natural coalitions of stakeholder groups. Healthcare information systems could benefit from the computer supported cooperative work literature which takes into account incentive structures, technology fit into work processes, and awareness to the coordination of work [Pratt et al. 2004]. The decisions incorporated into the architecture and design of the DSG illustrate how these CSCW considerations can be incorporated.

Incentives

Incentive structures are at the heart of this DSG. The designers took a normative view in which all stakeholder interests have intrinsic value [Donaldson and Preston 1995]. Each member of the federation must see sufficient benefit to take on the challenge of operating an inter-organizational storage infrastructure. The semantics of “trading (storage) space” implies collegiality and reciprocity versus “buying space” or “providing space.” The trading space metaphor embodies the pluralistic discipline of Pouloudi et al. [2004] in which the social/human is integrated with the scientific/technical disciplines. When space is traded, each organization has a vested interest to overcome the challenges of implementing a DSG. Trading is socialization of the storage space technologies.

The incentive of reciprocity is the realization that the typical healthcare organization cannot provide business continuity and disaster recovery on its own because of budget constraints. There needs to be a realistic assessment of current operational availability in the face of a multitude of downtime scenarios. These scenarios range from a localized air conditioning failure in the server room to power disruption in the neighborhood to a citywide disaster such as an earthquake. Geographic separation provides many advantages but, in the past, at the cost of a second site. The DSG offers both an operational and cost incentive. The networked aspects of DSG technology offer the financial incentives to take a serious look despite the organizational complexity.

Once the notion of sharing mutual benefits is accepted socially at the level of each organization, assistance at the operational level can take place. The question can be asked: “What else in the IT infrastructure could be shared to increase performance at a marginal cost (for one’s benefit so one can contribute more to the coalition)?”

The question of who owns and operates the DSG can now be answered in a collective sense. The federation owns the DSG but that alone is not actionable. Who really owns the DSG are the federation partners who enact the DSG just as much for themselves as for the federation. Without the federation, the ability to benefit from the marginal cost of traded space evaporates. This means there is mutual aid to help out in times of crisis. Collaboration within the
federation helps a weaker member become stronger. This might be technical assistance to improve infrastructure security or even pay for some of the infrastructure itself.

Workflow
The DSG introduces grid technology that changes workflow. A traditional IT infrastructure is designed to be isolated for the purpose of protecting digital assets. Backup and recovery is done within the organizational IT fortress. At the opposite spectrum is grid computing which at its heart is sharing of data and computational resources. A DSG for protected health information, such as clinical images, recasts the concept of sharing and protection. The DSG virtualizes the storage infrastructure so private data is moved among organizations for protection, but no sharing of data takes place. Only storage space is exchanged. Underlying the DSG is an unspoken social contract that my data can be better protected (at least from availability sense) by letting you keep a copy. Furthermore, it says I trust you because you trust me because your data is under my care.

To successfully deploy a DSG means that it is possible to protect networked assets without a “fortress” mentality. The federation partners of a DSG become allies who work as a coalition to protect each other. Protecting the DSG infrastructure is more than encrypting communication links between sites. The federation partners must also reinforce each others “fortress” so that the DSG can be more effectively defended against outside attack.

Using the DSG merely for backup, replicating images to external locations in near real time, does not take full advantage of the provisioning and virtualization of grid computing. The value of a DSG comes to play during operational emergencies, whether it is a recovery from a local disaster or temporary overload of local storage resources. The DSG acts as a “hot” backup site or provides load balancing. Images are still available for retrieval by end users (e.g., radiologists). The end users don’t know where the images come from although they may notice an increase in transfer time. The operators can then focus on restoring local services rather than trying to provide temporary emergency access to images.

Awareness
The DSG introduces awareness among administrators and operations staff to common problems faced by the federation. Solving the challenges of protecting health information in a networked DSG means that operational and security best practices will have been exchanged in the interest of developing a policy that represents a common denominator across all organizations of the federation. It is necessary to look beyond the operational benefits that a DSG brings to clinical image backup and recovery. The DSG may well be an example of a technology that can reinvent and transform existing organizational information systems and work practices [Scacchi 2004]. Once clinical images can be safely stored and protected in a DSG, other types of clinical evidence may be candidates. This in turn may lead to fresh ways of networking healthcare resources, especially those related to informatics, which can flatten the rising costs of healthcare which is detrimental to society as a whole.

DSG as a Socio-Technical Network
The traditional context of stakeholders is the firm or focal organization. Pictorially, stakeholders are viewed as spokes radiating from a hub with the firm, or lower level entity of the firm, being at the center [Freeman 1984]. The link between stakeholder and the firm was traditionally seen as dyadic. Rowley [1997] argues that instead of dyadic relationships, firms answer the simultaneous demands of multiple stakeholders. In the case of this DSG, there are multiple ensembles of stakeholders conversing at once. For example, operators from all three federation partners may sit together to discuss collective issues but interspersed in the conversation are quick sidebars between only two members to address a specific technological issue.

The stakeholders in a DSG cannot be viewed as being related to a single organization because they now have dual interests. They must protect the interests of their own organization but also promote the inter-operability of the federation. The storage space contained in the DSG only exists if the federation exists. The designers took advantage of natural coalitions that arose [Weiss 2003], such as their own as operators, or appealing to the collective interests of others (e.g. administrators).

The DSG analysis also shows that stakeholders belong to an extra-organizational entity (e.g. federation). This is the hallmark of a socio-technical network. There are interactions between the people (stakeholder roles), organizations (federation partners), and the underlying technologies and policies. While the DSG certainly operates with formal agreements among the partners, the federation was not formalized until the technical feasibility and the collaborative social fabric was established. The relationships built a socio-technical network to enable the creation of a federation. The DSG federation interconnects organizations (e.g. a grid) but not in a formal sense.
The stakeholder relationships of a DSG are best explained as a socio-technical network. These relationships make it hard to isolate the social from technical [Lamb et al., 2000]. The design metaphor trading (storage) space” is both technical but also embodies the social fabric that holds the DSG together. There is indeed a cultural and institutional context to this DSG, namely the pre-existing relationships between the federation partners. Whether the DSG could function in a different context or technical configuration is central to the characterization of an information system as a socio-technical network [Lamb et al. 2000].

Further Research
The analysis of this early DSG effort is only a starting point. Given the operational nature of this project within an IT infrastructure (e.g. backup system), drawing an arbitrary boundary around the two roles—administrators and operators—simplified the analysis. A weakness in stakeholder analysis is drawing the boundaries [Rowley 1997]. The reality is that technologies more visible to end users requires extending the boundaries around these stakeholders. A combinatorial increase in complexity of the analysis results. In this DSG, the views of operators and administrators were aligned toward a low-cost backup and recovery solution. When the analysis includes clinical end users, greater attention must be paid to mapping the stakeholder coalitions that arise [Friedman and Miles 2006]. For example, radiologists (end users) and operators jointly push for faster storage media so that response times are faster. The DSG may be pushed beyond backup and recovery towards use as active storage.

An intriguing question from this prototype DSG is the apparent absence of interorganizational conflict in the DSG federation. Conflicts exist in electronic networks such as those for business-to-business (B2B) operations [Sarkar and Cybulski 2002]. This raises the question about the network relationships in B2B being different from the interconnected but virtual federation of a DSG. One difference is that B2B is the linkage between two focal organizations, while the DSG connections are in some sense distributed within the fabric of the federation. Another difference, pointed out by the reviewer, is the collegial nature of information systems professionals such as database administrators. A collaborative effort is at the heart of a DSG since each federation partner must trust each other enough to place protected health information outside of their direct control. Organizations most likely to participate in a DSG federation would be related entities, such as affiliated hospitals, rather than newly formed partnerships of necessity. An examination of the network relationships between traditional and grid-based partnerships would provide insights on the degree of intertwining is necessary to establish a collaborative rather than strictly a transactional relationship.

VI. CONCLUSIONS
The deployment of a DSG expands the boundaries of stakeholder analysis to include federation partners and their existing IT infrastructures. An interorganizational information system, such as a DSG, must be viewed as a socio-technical network and designed accordingly. The concept of a DSG that employs grid networked computing principles requires a holistic view towards operating networked resources across multiple organizations. New organizational interdependencies need to be managed. Our traditional management and security tools for operating isolated IT systems must be reconsidered since a DSG creates virtualization across organizations and their unique IT infrastructures. Socio-technical analysis explains how key design decisions are embodied in the socio-technical network of the DSG. The most significant decision was incorporating an operational strategy of trading storage resources among federation partners.

REFERENCES


**ACKNOWLEDGMENTS**

This work was partially supported by NIH Grant No. R01-LM06270 from the National Library of Medicine while the author was with the University of Southern California Image Processing and Informatics Laboratory, Los Angeles, California. An earlier version of this paper was presented to the Second Middle East Conference on Healthcare Informatics. The current work was completed with the support of an American University of Beirut Junior Faculty Research Leave. The input of Vance Wilson, CAIS associate editor, and an anonymous reviewer strengthened this work.

**ABOUT THE AUTHOR**

**Nelson E. King** has a Ph.D. from the University of Southern California (2001), M.S. from the University of Arizona (1985), and B.S. from Columbia University (1979). His post-doctoral research in medical imaging informatics was with the Image Processing and Informatics Laboratory, Department of Radiology, University of Southern California. Dr. King led a research team that received a Certificate of Merit for a data grid infoRAD exhibit at the 2004 Radiological Society of North American meeting. He is currently Assistant Professor of Information Systems with the Olayan School of Business, American University of Beirut, Lebanon. The focus of his research is on the social and technical implications of process change in the introduction of healthcare information systems such as e-prescribing (E-service Journal), indoor positioning systems, and medication dispensing systems (*European Journal of Information Systems*).

Copyright © 2008 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from ais@aisnet.org