# Good Bones: Anthropological Scientific Collaboration around Computed Tomography Data

Andrea H. Tapia
College of Information Sciences and Technology
Penn State University
atapia@ist.psu.edu

Rosalie Ocker
College of Information Sciences and Technology
Penn State University
rocker@ist.psu.edu

Mary Beth Rosson
College of Information Sciences and Technology
Penn State University
mrosson@ist.psu.edu

Bridget Blodgett
College of Information Sciences and Technology
Penn State University
bward@ist.psu.edu

## ABSTRACT
We report preliminary results from a socio-technical analysis of scientific collaboration, specifically a loosely connected group of physical anthropology researchers. Working from a combination of interview data and artifact analysis, we identify current barriers to the scientists' collaboration as it relates to a valuable but scarce resource, a high-resolution computer tomography scanner. We analyze a two-layer structure of the collaboration, one that is loosely coupled through shared scanner access and use; and one that is tightly coupled through shared creative development of research questions, data analysis and interpretation. We conclude with implications for enhancements to the sociotechnical context and supporting infrastructure.

## Categories and Subject Descriptors
H5.m. Information interfaces and presentation

## Keywords
Scientific collaboratories, virtual organizations

## 1. INTRODUCTION
In science and engineering, virtual organizations have been formed and studied as scientific *collaboratories*, where researchers collaborate across geographic locations, working with colleagues, data, or instruments without location constraints [3, 27] . A collaboratory is a socio-technical system [11]; it consists of technology (e.g. scientific instruments and software, as well as communication and coordination software for working with distant colleagues), as well as social practices (e.g., the values, norms, and procedures shared by users of the collaboratory).

In the EVOSTA project (Examining Virtual Organizations through Socio-Technical Analysis), we are studying a scientific collaboratory with the goal of understanding how such structures function as virtual organizations. The collaboratory in question has emerged around a rare piece of scientific equipment, a High Resolution Computed Tomography (HRCT) scanner. While this collaboratory represents the important class of scientific collaboration structures that are "glued together" by a central critical resource [3], its goals are more complex than simply coordinating access to a tool.

The complexity of this particular collaboratory arises from its two-layer structure that is constantly changing shape. In one layer, it acts as a loosely-coupled organization providing *persistent support* for managing the scanning tasks and resulting data; in the other layer it serves as a source of *project-specific support* that requires more tightly-coupled communication and coordination for a cohort of distributed projects that emerge, coalesce, and follow their own trajectories. Within the project layer, each project develops its own organizational sub-structures and problem-specific relationships to the shared resource (e.g. type and timing of scanning; research prototcols that must be followed and documented; meta-data or archival requirements and so on). In this paper we report a preliminary analysis of this physical anthropology collaboratory. Using case study methods [28] that focus on a single ongoing research project (the Homind project), we show how some issues are associated with the persistent layer of the virtual collaboration, some with project-specific activities, and some with both.

In the balance of the paper we first summarize related work in scientific collaboratories, then describe our research setting and methods. We report the barriers that we identified in our analysis, followed by a synthetic discussion of how these issues map to the two-layer structure of the virtual organization. We conclude with implications for improved support for these scientists as well as a more general theoretical discussion

## 2. BACKGROUND
### Information and Communication Technology Support of Collaboratories
An assortment of information and communication technologies (ICTs) has been used to share information and create a sense of presence in a collaboratory [1]. These include email, chat, listservs, videoconferencing, wikis, blogs, and VOIP [20]. Many scientists use these ICTs as a way to carry on a "continuous conversation" that emulates the environment found when researchers are collocated [17]. By helping to maintain presence, ICTs can serve to keep distant collaborators focused, so that their attention is not overrun by more prominent demands of their local context [1].

Perhaps due to the inherent complexities associated with their work, scientists often prefer to use simple communication technologies such as email and instant messaging rather than more sophisticated and dedicated groupware applications [14]. However, the use of these spur-of-the-moment ICTs can impede the systematic storage and retrieval of project information; such technologies seem to encourage what some have termed throw-away communication [20]. At times researchers make use of broader technology services such as digital libraries and shared data repositories. By breaking down information and creating metadata, information retrieval tools make data more easily accessible [20].

There are many challenges associated with ICT use in collaboratories. Even with simple technologies, it can be difficult to get all members of a collaboratory to agree on a common platform or toolset, as members have different preferences and computing power and information speeds may vary greatly from one site to another [1, 13]. As for any virtual group, when synchronous meetings are needed, time zone differences and other scheduling constraints can pose major obstacles. Cross-institutional data archival, retrieval and transmission may entrain security risks [20].

## 3. RESEARCH DESIGN

The HRCT scanner is part of the CQI for Quantitative Imaging (CQI), an NSF-funded research facility that is part of the anthropology research infrastructure at a major U.S. university. It is a hub for scientists working on advanced imaging technologies. HRCT operates on the same general principles as medical CT, but uses higher energies. HRCT presents some challenges to researchers in need of the high-resolution images it can provide. Even with the increasing availability of fast computers with large amounts of memory, HRCT datasets often overwhelm a project's computational resources. Typical volumetric datasets range in size from 500MB to 2GB. The memory needed just to view the data is greater than current high-end desktop computers.

At the level of HRCT scanning, there is a regular flow of requests to the CQI, and the staff are responsible for responding to these in a timely way. To do this, they must understand the project requirements of each request (size, timing, deadlines), specimen-related concerns (ownership, lending policy, special handling), and data needs (format, media, delivery, checking). Thus any one request can involve considerable negotiation, clarification and follow-up.

One project that relies centrally on HRCT scanning is the Hominid Project (formally entitled Genetics of Craniofacial Variation and Human Origins). Hominid integrates studies of primate morphology and paleontology with gene mapping in baboons and mice. Skulls of baboons, mice and fossilized hominids are scanned and studied across academic fields and institutions. The project consists of a distributed team of senior (professors and research scientists) and junior (postdoctoral associates, graduate and undergraduate students) members located at three research sites located in the northeast, the southeast, and the southwest.

At the project-specific level, the collaborative activities are those typical of all research project: after a project vision is created and funding is obtained, the team plans and implements a set of interrelated activities that leverage the expertise and resources of each site. For instance in Homind, the data for baboons and mice (and related expertise) are obtained from different sites. Coordination involves the prioritization, scheduling, and transport of different specimens and resulting HRCT datasets, the co-creation and interpretation of the subsequent data analysis, and the shared development and publication of scientific findings.

### 3.1 Data Collection

Our case study relies on two forms of data – interviews and textual documentation from project records. We conducted 13 interviews encompassing all key project stakeholders including the PIs from the three research sites, postdoctoral fellows and graduate students. Interviews lasted 30-75 minutes; each was audio and/or videotaped and transcribed. The texts were original project documents pertaining to the Hominid NSF proposal, the intellectual property agreements, emailed correspondence and scanner documentation.

#### *3.1.1 Analysis Methods*

The transcribed interviews and textual data records were analyzed using *analytic induction*, a mixture of deductive and inductive approaches [10]. First, we developed and expanded four deductive code sets that were based on insights gained from the overall research context; sources for these codes included prior studies on virtual organizations, collaboratories, socio-technical analysis, and core questions posed during the interviews. In the analysis presented here, these four code sets were used to organize the inductive coding into four general barriers to virtual scientific collaboration.

## 4. FINDINGS: BARRIERS TO VIRTUAL SCIENTIFIC COLLABORATION

Several general barriers to the success of scientific collaboratories have been identified [3]. Of particular relevance to our research are barriers to collaboration that span multiple institutions. In addition to difficulties associated with distributed work contexts and different work cultures, these broad-reaching collaborations are subject to institutional layer policies (e.g. legal) set by high-ranking institutional officials outside the context of a given collaboratory.

Each of the next four sections discusses a general barrier related to scientific collaboratories: cross-institutional collaboration, collaboration readiness, collaboration technology readiness, and intellectual property. The discussions are organized by example barriers that were induced from the case study data. After presenting the findings within each barrier type, we position our findings with respect to the two-layer structure of this organization – the persistent loosely coupled sharing centered on the HRCT resource, and dynamic tightly coupled coordination of individual projects.

### 4.1 Barrier Type 1: Cross-Institutional Collaboration

Cross-institutional collaboration refers to scientific collaboration that occurs across multiple, distinct, and often geographically distant institutions. This form of collaboration requires scientists to invest resources and energy in establishing and maintaining relationships with other organizations. Although distance presents a variety of challenges to collaboration such as establishing common ground [5] and maintaining awareness of distant counterparts [26], crossing boundaries between research institutions increases project complexity and is frequently a greater barrier than distance alone [8].

Our data analysis revealed three examples of barriers under the general category of cross-institutional collaboration (Table 1): heterogeneous stakeholders, distance and distractions, and cooperative technology support and access.

| Table 1. Cross-Institutional Collaboration Barriers |
|---|
| Heterogeneous stakeholders |
| Distance and distractions |
| Cooperative technology support and access |

*Heterogeneous Stakeholders*

This barrier was problematic at both the persistent and project layers. Although the CQI "owns" the scanner and software and thus is a salient stakeholder in the persistent layer, some individuals participate in dual roles - as both HRCT service providers and Hominid investigators. This creates an inescapable interaction between the goals and concerns of the persistent layer and the project layer. For example, in their HRCT roles, the members hold organizational goals of ensuring quality service and attracting collaborators who can contribute to the CQI's research trajectory. As Hominid team members, they are more concerned with the traditional values of academia, including publications, funding, promotion and tenure, and so on. Other stakeholders contribute other tensions to the coordination of HRCT activities. For example, the university hosting the lab has its institutional concerns like visibility and long-term education or training. When a new request comes in from an important client (e.g., an internationally-known lab), the CQI staff might feel they need to give it priority, even if it means delaying the work of the Hominid project. Because some of the specimens used in Homind come from repositories in other institutions, the staff must carefully address issues related to transfer and access of specimens; these regulatory concerns can also get in the way of project-specific goals.

*Distance and Distractions*

Scientists at each institution work on multiple research projects, only one of which is the Hominid project, and have competing demands for their time. Multiple researchers working at different levels – including PIs, post-docs and graduate students – complained about the absence of face-to-face meetings where all investigators from the three sites were present. Indeed the failure to have such meetings, especially at the formulation stage of Hominid, was seen by one PI as a major contributor to the lack of clear protocols regarding data ownership and control, and norms for collaboration.

*Cooperative Technology Support and Access*

This barrier refers to the technology required to share the datasets that the HRCT scanner produces with the scientists requesting the scans. CQI has devised several work-arounds to circumvent these problems, but each has had unsatisfactory outcomes. For example, the CQI found that sending files as email attachments is error prone, because individual project members have different file size limits for attachments. As a result an email sent to multiple members may lead to some recipients receiving an attachment and others not, which in turn can lead to missed information, misunderstandings, or extra work. Another work-around has been to create DVDs and use postal mail. Understandably, this option is viewed with disdain by the scientists who have become used to downloading information on demand. Finally, a third-party transfer system has been tried, but it also proved to be slow and cumbersome. Together these impediments cause the transfer and sharing of data across institutions to be delayed or even completely stymied.

## 4.2 Barrier Type 2: Collaboration Readiness

Collaboration readiness refers to the extent that potential collaborators are motivated to work with each other [15]; see also [16, 19, 21]. It has several broad components: motivation to collaborate, shared principles of collaboration, and experience with the specific elements of collaboration. Success seems to require a positive orientation toward collaboration, either as a result of incentives or as a result of normative practice [6]. Because of the distributed nature of the Hominid project (much of the collaboration occurs virtually), collaboration readiness in this case study implies a strong positive orientation to collaborate, where ICT is used to span distances. Our analysis uncovered four examples of barriers in the general category of collaboration readiness (Table 2).

| Table 2. Collaboration Readiness Barriers |
|---|
| Materials and data sharing |
| Weak reward structure |
| Cost of sharing scarce resource (e.g. scanner) |
| Project decision making structure |

*Materials and Data Sharing*

This barrier arises from the need to share physical objects like computers and datasets. For example, several powerful desktop computers used for manipulating scanned data needed to be shared by both physically-present and distant members of Hominid; this at times caused conflicts because the team had not worked out sharing practices in advance. In the case of data sharing, some project team members claimed ownership of certain forms of data, and this would impeded data analysis efforts by others. Furthermore, data sharing with non-project researchers came up as an issue for the project team. They were forced to generate the protocols regarding what data could be shared, with whom and when.

*Weak Reward Structure*

This barrier refers to the absence of well-defined motivations and rewards for the sharing and coordination needed for smooth operation, at both the persistent level of HRCT and the Hominid project level. In the CQI, the staff receive regular requests for access to scanned data files. To respond, they must locate the scanned data and create a copy to share; this is a multi-step, tedious, and manually intensive process (see Manual Processes under Barrier 3). However there is no clearly defined. In the Hominid project, team members knew that their data, analytical tools and findings would eventually have to be shared with other physical anthropologists, but even so they were unsure of what rewards such sharing might have for them. They often referred to using or mining the data for their own needs first and then "dumping" it for others to use.

*Cost of Sharing*

This barrier refers to the cost of sharing data and scanner resources, for instance the time spent to set up the scanning process for an object. For example, the CQI staff told us that much of their time was not spent on scanning objects but rather on the negotiations with individuals and institutions about the scanning. They said that because of the mechanisms their university had defined to charge for scanner time, a significant amount of time was "wasted." They spent time managing individual scanner contracts and costing each process, which involves negotiating parameter settings for the scans to be done, and the delivery of scanned data.

At the project layer, Hominid team members have begun to receive requests for project data. Because these requests have been few thus far, each has been handled on an ad hoc basis by the overall PI through tele-meetings. However, the PIs know that as they create more data the number of requests for sharing is likely to rise. They have already discussed the need for a more formalized mechanism for approving and sharing data.

*Project Decision Making Structure*
This barrier refers to the mechanisms in place for making group decisions. The project PIs insist that there must be consensus between the PIs in order for any major decision to be made, but there is no well-defined process for making this happen. With PIs distributed across three institutions and team members in additional distant sites, the team finds that consensus is difficult and slow. A specific example of this barrier relates to research publications. A post-doctoral fellow working on Hominid wrote a paper based on the findings she had gathered from one piece of the project. However the consensus process (accomplished via email and tele-meetings across the three institutions) took months to arrange, just so she could garner the support needed to publish her paper.

## 4.3 Barrier Type 3: Collaboration Technology Readiness

Collaboration technology readiness refers to the presence of sufficient technology infrastructure, as well as the availability of local technology expertise, both explicit and implicit [15, 16, 19]. Research continues to demonstrate that the introduction of ICT that fails to complement, or is incompatible with, existing policies and practices will not increase scientific collaboration [22, 23]. Barriers to data sharing include competition and the large amount of work involved in making data reusable when compared to the benefits and risks of sharing [2].

In a general sense, both the CQI staff and the Hominid team were comfortable with the use of the web, email, word processing and spreadsheets. These were deemed necessary to the interaction between distant project members. However, the tools available to them were not always seen as functioning well in the support of collaboration. In our case study, many of the collaboration technology issues arise from the virtual aspects of the research interactions. The collaborators are spread across different institutions and rely on Internet technologies and postal services to meet their data sharing needs. In addition, their data varies in terms of size, type and location; at times it also raises complex access issues. For example the ownership and management of data often imply specific protocols for others to access the data. Different data may also be stored in slightly different formats in different locations, or on different devices. The databases used to store data are stored are proprietary to each PI and institution and are not linked in any systematic way. Thus data creation, storage and management can be time consuming, increasing the costs of collaboration for all parties. Our analysis uncovered three example barriers related to collaboration technology readiness, shown in Table 3.

| **Table 3. Collaboration Technology Readiness Barriers** |
|---|
| Access to data |
| Data storage, retrieval and transmission |
| Manual processes |

*Access to Data*
This barrier refers to one of the core technology issues for this scientific collaboratory. Raw scanner data, metadata used to document the scanning process, and “landmark” (the process of identifying key points in space on a 3D image, in these cases a skull) data all have their own access problems. The files of raw scanner data are very large and file size has become a significant barrier (for movement, for storing on any given device). Because of such issues, these files are currently stored in specialized storage arrays in the CQI and made available only via FTP or physically via a custom DVD. These files are protected by the university’s stringent firewalls that provide intrusion protection, but also reduce ease of access. This is an example of how something as simple as storage location can hinder access to data. Metadata documenting the scan process include both identifiers of the scanned object and the scanner settings when the object was scanned. These metadata are currently stored in several forms and in several locations. At the project layer, landmark data is produced concerning each scan. The scan is retrieved from the CQI’s database, converted to a viewable image and then landmarked. The coordinates of the landmarks are stored in a local database, which is proprietary to that institution. This database is not accessible to anyone outside the hosting institution. This is a problem that was noted by several project members. The local project database is created and maintained by a single IT project staff member who supports not only the PI and associated lab, but the entire project. None of the PIs or project team members have any access to the database. As a result, much of the job of this IT support person is finding ways to share information across institutions. In most cases the landmark data is burned onto a DVD and sent via mail to co-PIs at other institutions.

*Data Storage and Transmission*
Storage was an issue for both the loosely coupled activities of the CQI and the Hominid-specific interactions. A key aspect of this problem concerns the linking of different data storage locations. The scans are stored in one place; images built from the scans are stored in another place; landmark coordinates are stored in yet another place. And in fact, the metadata that join these various sets of data together via subject identifiers are stored in yet another location. The project team members understand that they need a mechanism for more seamless linking, but they are not there yet. In general, project team members do not know where these different datasets are stored or how to access them. At the CQI, storage of non-scan data is a problem. Despite the electronic format for scanned data and associated metadata, the investigator and project data for the CQI is often stored in paper notebooks or in the minds of the CQI staff. The staff are quite aware that this paper storage system has significant limitations: it is impossible to search, relies on staff memories, is inaccessible from a distance and is difficult to share or copy.

*Manual Processes*
A barrier related to the manual aspects of the research emerged at both layers of this collaboratory. The staff expressed frustration about the degree of manual activity required to create or manage the data. For example, the CQI staff have been considering the process of making taped backups of the scans, moving those backups to a long term storage area as they get older, and the very time consuming process of retrieving them when they were needed again. The CQI staff also relayed problems creating the metadata files to document the scan parameters for each episode. This is done manually and is also very time consuming. With respect to Hominid-specific activities, several team members complained about the manual labor in creating and managing the data.

## 4.4 Barrier Type 4: Intellectual Property

Universities often guard their intellectual property in ways that hinder multi-site collaboration [3]. Intellectual property issues in collaboratories vary depending on the focus of the collaboration. They can include participants’ claims to IP emerging from the collaboration; ownership and licensing of IP; dissemination of scientific data; and apportionment of liability with respect to violations of competition law, violations of human and animal subjects rights, and loss of reputation due to incompetent or unethical conduct[18]. Collaboratories can also run into institutional-related problems, intellectual property issues, especially legal issues, that cannot be resolved [24].

In the case of the Hominid project taking place in the anthropology collaboratory, large amounts of data are produced across three institutions and stored, shared and used virtually. The ownership, use and sharing of the data was never addressed in any formal agreements within the project. As a result the PIs and the project staff have a sense of unease; they are wrestling with conflicting ideas about the mandate to share, and are confused as to what and how data can or should be shared with others. Our analysis uncovered three examples of barriers related to intellectual property, shown in Table 4.

| **Table 4. Intellectual Property Barriers** |
|---|
| Data ownership and conflicting control |
| Different vetting processes |
| Broader sharing vs. publications |

*Data Ownership and Conflicting Control*
The ownership and control barrier affected activities at both the persistent layer and the project layer. While the HRCT CQI and scanner are clearly integral to the process of creating research data, this piece of the collaboratory rarely plays a role in debates and concerns over ownership. Instead a variety of stakeholders who have some role in the HRCT-related projects holds some claim to ownership of the specimens and the data produced from them.

For example, when a project focuses on scans of sculls or other specimens owned by a museum or institute, this owning organization often believes that any data produced from their specimen is owned by them. At the same time, the university owning the HRCT scanner also claims that scans produced via its equipment are owned by the university. Finally, the project researchers who request produce the scans and produce the landmark data also make claims to ownership of the data and findings.

Data ownership was not called into question when specimen, scanner, data and institution were all one. However, a virtual collaboratory enables research that is multi-site and multi-institution, and this places ownership at the center of many debates. This staff member believed that out of collegiality and tradition, agreements would be created for shared ownership of data. However, this was not always the case when it came to publication and use of the data.

With respect to the Hominid project, the question of data ownership has produced a lot of discussion that has cost time on the project. All project members expressed some confusion over ownership of data. Most believed there were some high-level agreements, at least initially between the three institutions and the NSF; none knew exactly what had been agreed. Most believed there were no operating agreements. When the project team members were asked more specifically about the landmark data that they generated (i.e., not the scan data or genetic data), their responses were more confused. Even more confusing was the ownership of data related to specimens. When discussing information related to specimens held by museums or other nations, there was also a strong question of ownership.

*Different Vetting Processes*
The vetting barrier was primarily a problem for Hominid-specific activities. The unclear ownership of data entails an unclear procedure for granting permission for outside scholars to use the data (this is also related to the access issues discussed earlier). The PIs had no procedure in place, and no prior agreements as to how the data would be shared.

*Broader Sharing vs. Publications*
The problems concerning sharing and publications were primarily an issue at the project layer. All project members knew that sharing project data with a wider audience was a mandate. For some it was put exactly in those terms, as required by the funding agency. For other team members, sharing data is something that will be done, but they still do not know the details of exactly what, when and with whom they will share.

# 5. DISCUSSION

Through our analysis we have identified thirteen barriers that impede the development and operation of the virtual HRCT collaboratory. We have categorized the barriers into four types – CI: Cross Institutional barriers, CR: Collaboration Readiness Barriers, TR: Technology Readiness Barriers, and IP: Intellectual Property Barriers. As implicit in our discussion thus far and summarized in Table 5, these barriers may be operating primarily at the level of the loosely coordinated HRCT activities, at the level of the Hominid project, or both. We turn now to a more general discussion of these barriers and how they are influencing the two different layers of the physical anthropology collaboratory.

**Table 5. Summary of Barriers Affecting the Two Layers**

| Barrier | Persistent | Project |
|---|---|---|
| **CI**: Heterogeneous stakeholders | √ | √ |
| **CI**: Distance and distractions | | √ |
| **CI**: Co-operative technology support and access | √ | |
| **CR**: Materials and data sharing | | √ |
| **CR**: Weak reward structure | √ | √ |
| **CR**: Cost of sharing scarce resource | √ | √ |
| **CR**: Project decision making structure | | √ |
| **CTR**: Access to data | √ | √ |
| **CTR**: Data storage and transmission; retrieval | √ | √ |
| **CTR**: Manual processes | √ | √ |
| **IP**: Data ownership and conflicting control | √ | √ |
| **IP**: Different vetting processes | | √ |
| **IP**: Broader sharing vs. publications | | √ |

## 5.1 Barrier Types

We observed that the Cross-Institutional barriers operated at a mix of levels. Distance and distractions was primarily a problem for the research team members, whereas co-operative technology support and access was an issue for the persistent activities of the CQI. The issues related to heterogeneous stakeholders affected both layers of the collaboratory. These Cross-Institutional barriers suggest that this virtual collaboration among physical anthropologists experienced significant costs to create and maintain their efforts across institutional boundaries and physical distance. The number and diversity of stakeholders at both levels presents an ongoing challenge to collaboration.

In terms of Collaboration Readiness, again we observed barriers in both layers. Two of the barriers operated at both levels, lack of potential benefits and cost and resource sharing. Materials and data sharing was an issue for the more tightly coordinated project activities, as team members had to struggle with who should have access to project resources at what time; similarly the problems in decision making were experienced when consensus decisions needed to be made. In contrast, the weak reward structure was an issue for both the ongoing activities of HRCT scanning (e.g., setting aside time for other projects even when Hominid needed attention) and for the team members themselves. The same pattern was true with respect to the cost of sharing the scarce resources. By definition scarce resources like an HRCT scanner, and valuable resources like a dataset must be coordinated among team members, but without explicit reward structures in place the coordination may occur in an *ad hoc* fashion or as an afterthought.

The barriers associated with Collaboration Technology Readiness are notably different in the scope of their impacts. All three problems – access, storage, and manual procedures – were influencing the activities taking place in both layers of the collaboratory. To some extent this is not surprising, as technology is the essential glue that makes a virtual collaboration possible. At the same time this pattern points to a technology mismatch between the technology now in place and what is needed.

Finally, we note that barriers in the Intellectual Property class were also mixed in the scope of their impacts. The uncertainty and developing notions about sharing of the project results, and the uncertain process of vetting others' use of the data, were of concern to the Hominid researchers. However, the more essential questions of who owns the data were of concern at both levels – the researchers must acknowledge ownership as relevant in their work, while the CQI has the general responsibility of managing ownership and access of the specimens and data that pass through their procedures. The expectations of outside agencies like NSF, museums, or even other nations figured prominently into these considerations at both levels. This particular barrier suggests a need for more formal agreements between the stakeholders and institutions as to how to manage data and the intellectual products produced from the data.

Thinking more synthetically about the two layers and how they interact, Table 5 suggests that in general more problems are influencing collaboration at the project level than at the level of the CQI and scanning. This is not surprising given the more tightly-coupled nature of research projects relative to the management of the scan process and results. The project level is also the locus of strong individual and sub-team motivations – PIs building or enhancing their own research reputations while also mentoring and training their junior collaborators. In contrast the HRCT scanner was obtained and positioned as a research facilitation activity, so most of its barriers relate to its responsibility for responsive and high-caliber management of the scan process.

To be more specific, the Hominid project experienced some problems that the CQI activities did not. Most importantly, the Hominid researchers had problems with sharing; sharing resources, sharing data and sharing materials. They also encountered problems related to sharing *outside* the scope of their own project. Lastly, the project level experienced problems managing distance. These project-level barriers can be partially explained by longevity, in that the HRCT CQI has had more time to discover and deal with problems and establish collaborative patterns. At the same time, each project should benefit from the CQI's strategies; as it refines its collaboration patterns, the individual projects should also enjoy the benefits of longevity.

These differences in layer may also be at least partially explained by the physicality of the CQI – it is tied to the HRCT scanner and surrounding offices. The one barrier that emerged only at the CQI level (cooperative technology support) is inherent to its facilitation role, namely finding ways to support custom project contexts. It is in this role that the CQI can evolve to better support the needs and resources of specific projects.

Focusing more on the persistent level of HRCT scanning, the Technology Readiness barriers were common. This emphasizes the persistent role of the CQI to find and provide appropriate tools and orientation for virtual collaboration. Some of these issues may have emerged because of changes in the role of the CQI. If the HRCT activities were originally seen more in a sub-contracting role (i.e., a place where data was produced, but the CQI staff were not full partners in the research), it is understandable that at the persistent level is still developing appropriate expectations and tools for collaboration. Currently the CQI staff act as PIs on the Hominid project, and thus take a more significant and active in role research activities. They have evolved from a role of technician to scientist, and expectations among all stakeholders must change to match.

## 5.2 Loose and Tight Coupling in Science Collaborations

We have represented the persistent layer of HRCT scanning activities as loosely coupled. Loosely coupled organizations are known for a lack of coordination, and an absence of regulations so as to promote organizational flexibility, adaption and sensitivity to the environment [25]. The HRCT scanner layer is persistent because it continues to exist through time, while projects that make use of the CQI and its resources come and go. The CQI is loosely coupled in its semi-autonomous nature, wherein it operates somewhat independently of its parent organization, the university. The loose coupling also enables it to interact in parallel with a variety of different research projects (i.e., in addition to Hominid). But also because of its loosely coupled operation, the CQI has developed procedures and policies that are idiosyncratic, ad hoc, and just-in-time to fulfill its operating needs. It has often operated without formal agreements or standard operating policies and procedures.

Now however, with the advent of virtual collaborations, the CQI, the scanner, its data and its staff have all been thrust into new roles. The CQI takes on a new role as a partner in a larger scientific effort. The CQI not only creates scanned data for its users, but also now is often asked to store and manage the data specifically for projects, and to also address such issues as access, sharing and other project-specific protocols.

These new requirements are particularly salient when the project is a multi-year, multi-institution endeavor. Several scholars have argued that in large collaboratories, "hierarchy of authority, written rules and regulations, formalized responsibilities, and a specialized division of labor" are essential to the success of their bureaucratic procedures [4]. When collaboration occurs at a distance, more formalized structures may be required for success [12, 20].

It is exactly this situation in which we find the Hominid project. It is a large collaboration (for Physical Anthropologists) and requires collaboration across three large institutions and with several other isolated researchers at a distance. The Hominid project has tightly coupled needs that result from joint pursuit of research questions and associated analyses. As a tightly coupled collaboration the project needs a more formalized structure for its operations, procedures and protocols. Hominid is experiencing several barriers to developing these elements that would increase its efficiency and success. Interestingly, Hominid is turning to the CQI, the more loosely coupled organization, for help in formulating is more structural and procedural elements.

# 6. SOCIOTECHNICAL DESIGN IMPLICATIONS AND CONCLUSIONS

Although we are just beginning our analysis of the physical anthropology collaboratory, we are already able to draw some implications for design or refinement of the social and technical supports in place. For example, many problems associated with ownership might be addressed by procedures that establish agreements and related policies in advance. The CQI can play a guiding role in this, perhaps by requiring such agreements as part of any participation in a new research project. As these procedures develop and are formalized, implementation support through a web-based form, with subsequent storage and access in a shared repository, are a natural evolutionary path.

Some technical problems have arisen because the project researchers tend to rely on email and attachments for many sharing activities. These behaviors bring along "knock-on" effects related to each institution's communication infrastructure and policies. If the sharing were shifted to a custom online system (e.g., using an open-source infrastructure like Drupal), data transfer problems might be eased, although there might still be issues about shared access to specialized high-power computers. Again, it seems likely that the CQI would be the one to promote changes in sharing practices of this sort.

Finally, the problem of needing consensus prior to any individual "usage" of the shared data and resources might be solved by finding technology other than tele-meetings to support the discussion and emergence of consensus. Discussion forums (with appropriate protection and authentication mechanisms) might support a more mixed-mode process of synchronous and asynchronous interaction, hopefully easing and speeding the processing of individual requests for consideration.

In this paper we have reported preliminary results from a sociotechnical analysis of scientific collaboration. Through our analysis of interview and document data and artifact analysis, we identified thirteen current barriers to the scientists' collaboration as it relates to a valuable but scarce resource, a high-resolution computer tomography scanner. Drawing on these barriers, we analyzed a two-layer structure of the collaboration, one that is loosely coupled at the scanner level; and one that is tightly coupled at the project level. We conclude that virtual collaboration, fundamentally dependent on technologies, needs to develop more appropriate technologies, which support the on-going two level nature of the collaboratory

# 7. ACKNOWLEDGMENTS

We would like to acknowledge the US National Science Foundation for the partial support of this research OCI – 0838400,*VOSS: HRCT Scanning as Glue: Sociotechnical Analysis and Support of a Loosely-Coupled Virtual Organization of Emergent Distributed Projects*.

# 8. REFERENCES

[1] Ackerman, M., E.C. Hofer, and R. Hanisch, *The National Virtual Observatory*, in *Scientific Collaboration on the Internet*, G. Olson, A. Zimmerman, and N. Bos, Editors. 2008, The MIT Press: Cambridge, Massachusetts. p. 135 - 142.

[2] Birnholtz, J. and M.J. Bietz. *Data at work: Supporting sharing in science and engineering*. in *2003 International ACM SIGGROUP*. 2003. New York, NY: ACM Press.

[3] Bos, N., et al., *From shared databases to communities of practice: A taxonomy of collaboratories*. Journal of Computer-Mediated Communication, 2007. 12(2): p. article 16.

[4] Chompalov, I., J. Genuth, and W. Shrum, *The organization of scientific collaborations*. Research Policy, 2002. 31(5).

[5] Clark, H.H. and S.E. Brennan, *Grounding in Communication*, in *Perspectives on Socially Shared Cognition*, J.M.L.a.S.D.T. L.B. Resnick, Editor. 1991, American Psychological Association. p. 127-149.

[6] Cohen, S.A., *A consultative and participative approach toward change management in a large insurance company*. Consulting Psychology Journal: Practice and Research, 2000. 52(2): p. 142-147.

[7] Corbin, J. and A. Strauss, *Grounded theory research: Procedures, canons, and evaluative criteria*. Qualitative Sociology, 1990. 13: p. 3-21.

[8] Cummings, J. and S. Kiesler, *Collaborative Research Across Disciplinary and Organizational Boundaries* Social Studies of Science, 2005. 35(5): p. 703-722.

[9] Eisenhardt, K.M., *Building Theories From Case Study Research*. Academy Of Management Review, 1989. 14(4): p. 532-550.

[10] Epstein, L. and A.D. Martin, *Coding Variables*, in *Encyclopedia of Social Measurement*, K. Kempf-Leonard, Editor. 2004, Academic Press.

[11] Kling, R. *Critical professional discourses about information and communications technologies and social life in the U.S.* 2002. Montreal, Quebec: Kluwer Academic Publishers.

[12] Maglaughlin, K.L. and D.H. Sonnenwald, *Factors that impact interdisciplinary scientific research collaboration: Focus on the natural sciences in academia*. 2005, University College of Bor√•s. Swedish School of Library and Information Science.

[13] Myer, J., *A National User Facility That Fits on Your Desk: The Evolution of Collaboratories at the Pacific Northwest National Labratory*, in *Scientific Collaboration on the Internet*, G. Olson, A. Zimmerman, and N. Bos, Editors. 2008, The MIT Press: Cambridge, Massachusetts. p. 121-134.

[14] Nentwich, M., *Cyberscience: The Age of Digitized Collaboration*, in *Scientific Collaboration on the Internet*, G.M. Olson, A. Zimmerman, and N. Bos, Editors. 2008, MIT Press: Cambridge, MA. p. 33 - 49.

[15] Olson, G., T.A. Finholt, and S. Teasley, *Behavioral Aspects of Collaboratories*, in *Electronic Collaboration in Science*, S. Koslow and M. Huerta, Editors. 2000.

[16] Olson, G.N. and J.S. Olson, *Technology Support for Collaborative Workgroups*, in *Coordination Theory and Collaboration Technology*, G.M. Olson, T.W. Malone, and J.B. Smith, Editors. 2001, Lawrence Erlbaum Associates: Mahwah, NJ. p. 559.

[17] Olson, J.S., et al., *A Theory of Remote Scientific Collaboration*, in *Scientific Collaboration on the Internet*, G.M. Olson, A. Zimmerman, and N. Bos, Editors. 2008.

[18] Reichman, J.H. and P.F. Uhlir, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*. Law & Contemporary Problems, 2003. 66(WtrSpr): p. 315-462.

[19] Sonnenwald, D.H., *Contested collaboration: A descriptive model of intergroup communication in information system design*. Information Processing and Management, 1995. 31(6): p. 859-877.

[20] Sonnenwald, D.H., *Scientific Collaboration*. Annual Review of Information Sciences and Technology, 2007(41): p. 643-681.

[21] Sonnenwald, D.H. and L.G. Pierce, *Information behavior in dynamic group work contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control*. Information Processing & Management, 2000. 36(3): p. 461-479.

[22] Sooryamoorthy, R., et al., *Scientific Collaboration and the Kerala Model: Does the Internet Make a Difference?* Journal of International Development On-Line 2007. 20.

[23] Star, S.L. and K. Ruhleder, *Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces*. Information Systems Research, 1996. 7(111-133).

[24] Stokols, D., et al., *In vivo studies of transdisciplinary scientific collaboration: Lessons learned and implications for active living research*. American Journal of Preventive Medicine, 2005. 28(2): p. 202-213.

[25] Weick, K.E., *Educational organizations as loosely coupled systems*. Administrative Science Quarterly, 1976. 21(1-19).

[26] Weisband, S., *Maintaining awareness in distributed team collaboration: Implications for leadership and performance*, in *Distributed Work*, P. Hinds and S. Kiesler, Editors. 2002, MIT Press: Cambridge, MA. p. 311-333.

[27] Wulf, W., *The National Collaboratory*, in *Towards a National Collaboratory*. 1989, Rockefeller University: New York.

[28] Yin, R.K., *Case Study Research: Design and Methods*. 3rd ed. 2003, Thousand Oaks, CA: Sage Publications Inc.