
PAM Squared

**University of Michigan
Center for Sustainable Systems**

Consulting Group: PAM Squared

Members: Alexis Ashby, Maryam Seifeldin, & Priyanka Shanmugasundaram

Contact Information: pam-squared@umich.edu

Client: Center for Sustainable Systems

POC: Helaine Hunscher & Geoff Lewis

Contact Information: hunscher@umich.edu & glewis@umich.edu

Table of Contents

Executive Summary.....	2
Introduction.....	3
Methodological Overview.....	4
Finding & Recommendation 1.....	7
Finding & Recommendation 2.....	10
Finding & Recommendation 3.....	11
Conclusion.....	13
References.....	14
Appendix.....	14

Executive Summary

At any given moment, the Center for Sustainable Systems employs up to 25 researchers, most of which are graduate students. This means that the center has a high turnover rate, and depends on kept up student files and data, which is currently housed in the center's Q-drive.

However, permanent staff have trouble accessing components within the shared drive and streamlining their archiving process. Some problems with the current system are already known such as the poor labeling system and lack of structure. PAM Squared was called in to assess what else may be influencing how the Q-drive is maintained and provide some suggestions to improve the current system.

Our interviews with three permanent staff members and three student researchers revealed the following findings:

1. Student researchers are unaware of staff's needs and expectations, leading to disorganized research folders from the get-go.
2. The Q-drive's organization system does not promote findability and may not fit student researchers' needs.
3. The VPN system doesn't gracefully integrate researcher file management habits and can be cumbersome.

After careful consideration, we made recommendations on how the Q-drive can be improved in both the short and long term. Some of these recommendations focus on technology and specific tools while others focus on building new components to the Center for Sustainable System's culture:

1. **Standardization from Onboarding Onwards.** Provide documentation on file maintenance practices upon arrival, and make sure that current researchers know where to find this documentation. We included a possible naming system and suggestions for how to remind student researchers about expectations.
2. **An Offline Integrated Data Management Software to Add Tags to Files.** This would increase findability of files within the Q-drive and ease the process of adding metadata to files. We included a few software recommendations to get the center started.
3. **Use Git to Integrate the Functionality of Cloud Drives and the Continued Use of the Q-Drive System.** As such, the system would meet the needs of permanent staff and student researchers. We included a run down of what Git is and how to integrate it with the Q-drive.

Introduction

In any research institution, particularly in recent years, the storage and organization of collected data and resources is a problem that researchers, administrators, and students must grapple with. The performance of experiments, data collection, or information gathering on their own do not reach the goals of research; in order to benefit society or to be built upon, research results must be shareable and accessible to others. Traditionally, researchers would publish their results to books or peer-reviewed journals in order to communicate their findings. However, with the recent prevalence of data-centered research, simply publishing results does not sufficiently distribute information in a useful way. Furthermore, internal data management becomes more and more of a problem as large, multifaceted datasets are collected and long-term research goals are distributed across multiple individual researchers.

Problem:

The Center for Sustainable Systems (CSS) is an interdisciplinary research center within the University of Michigan's School for Environment and Sustainability. The CSS conducts research on various sustainability topics to include: energy, mobility, urban environments, infrastructure, agriculture, and consumer products. At the CSS there are usually several ongoing projects, with upwards of 25 student researchers working at the center at any one moment. The CSS has accumulated a lot of data and resources from their research over the years. Currently, all of this information is stored on a local hard drive. This hard drive is comprised of both a protected archive, accessible to a limited number of permanent staff members and an open shared drive open to all at the center. The drive is accessible via vpn connection.

Permanent staff expect that students maintain their research in the following folder structure on the the CSS's shared drive throughout their time at the center: research data, grant writing, models, presentations, and publications. Ideally, upon leaving the center the information that the student has maintained from the shared drive is incorporated into the protected archive. However, under the current system CSS staff are having trouble accessing components within the shared drive because information is poorly labeled, difficult to find, or not included in the shared drive system at all. The lack of structure on the shared drive, makes the archiving process difficult. There is currently a large backlog of data for the archive that has not been incorporated yet. In addition to current CSS staff, the lack of structure impacts new researchers entering the center. New researchers are often unaware of and unable to locate the plethora of resources contained within the shared drive.

Methodological Overview

PAM Squared used qualitative research methodology based on human-centered design and contextual inquiry to further investigate the CSS' issues with the shared drive.

“The core premise of Contextual Inquiry is very simple: go where the customer works, observe the customer as he or she works, and talk to the Customer about the work. Do that, and you can't help but gain a better understanding of your customer.”

-Beyer & Holtzblatt (1998)

We began our process by engaging in an initial client meeting to. As a result of PAM Squared's initial client meeting with the CSS, three themes emerged:

- Storage solutions, Accessibility, Interfaces
- Employee Entrance/Exit Practices
- Data Management Systems/Research Data Management

These themes helped to guide PAM squared's exploration into the problem. Prior to conducting interviews of permanent staff and researchers at the CSS, PAM squared conducted an extensive background research report on the best practices and emerging trends for the three areas outlined above.

Interviews

After completing their research, PAM squared interviewed pertinent individuals at the CSS. Individuals that were interviewed were suggested by senior staff at the CSS. PAM squared interviewed the following individuals at the CSS: three permanent staff members and three graduate researchers. Questions during the interview were derived from information obtained from the initial client meeting and inspired by information from the background research report. During each interview, PAM squared members took notes and were granted permission to record the interview.

Sample interview questions indicated below:

Entrance/Exit Practices:

- What is your role for providing instructions to new researchers about the Q drive?
- Can you describe the on-boarding experience for the most recent incoming researcher?
- Can you describe the most recent close-out procedure for a departing researcher?
- Can you describe your on-boarding experience at CSS?
- What did you enjoy most about your onboarding experience?
- How could your on-boarding experience have been improved?

Storage Solutions, Accessibility, and Interfaces:

- Describe the last time you were reminded of or explained to about where to store research files.
- What methods do you use to keep your files organized within/outside of the Q-drive?
- Can you walk me through the last time you access files such as presentations from the Q drive?
- When you remind students about archiving what do you make sure to point out about the file structure and naming process?
- What do you think is a strength/weakness of this system?

Data management systems/research data management

- What is the scope of your interaction with the current data management system?
- Ideally, how would this system work?
- Currently, working files and archived files are kept on separate drives. What is the reason for or thinking behind this?
- Describe how in the past research projects have been passed between student researchers here who don't ever meet in person.
- Have you worked with "inherited" data or research during your time here?
- How often do you use the Q drive compared to other local or personal data storage systems for your research?
- What was your experience getting a hold of the existing research?

Data Analysis

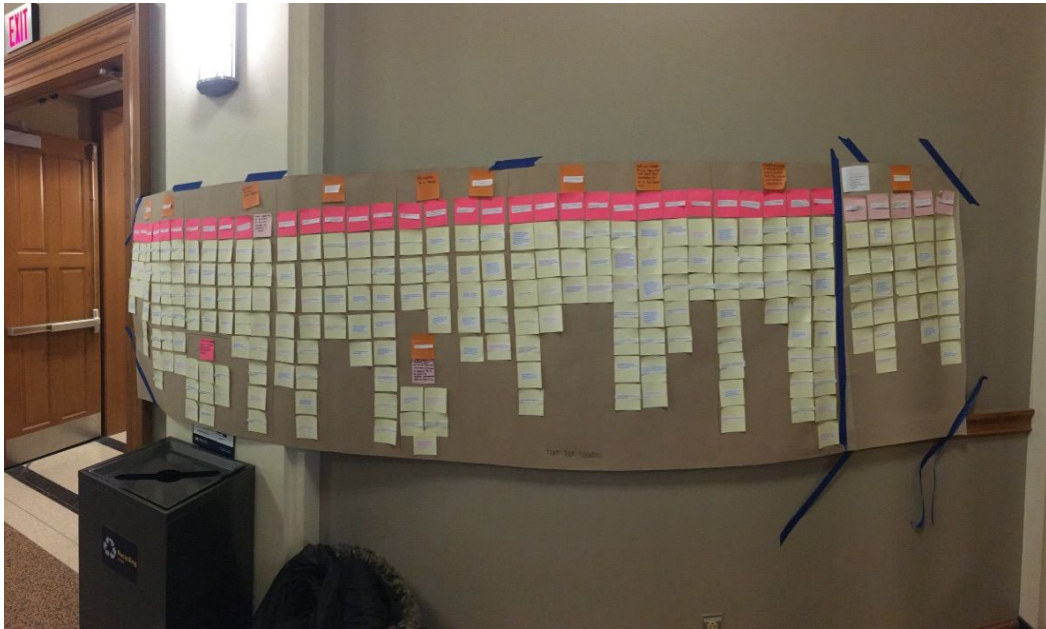
At the conclusion of each interview, PAM squared members met to review their notes. For each minute of interview that was recorded, PAM squared captured an insight in the form of a one sentence summary. These one sentence summaries were utilized to were then used to create a hierarchical series of clusters based on their relationship and themes from the interview notes.

Information from the highest level notes included (**orange sticky notes**):

- Staff and students do not communicate about their expectations for use of the core drive.
- Institutional knowledge is critical to the current structure and is often inconsistent with student understanding
- Researchers utilize cloud drives more than Q drive to obtain functionality that the Q drive does not provide.
- The Q drives ability to function as a data management system depend on the work and time put in by the director and admin staff
- The website is an open resource that is underutilized.
- Current naming system could benefit from standardization and a shift from researcher's name being folder names to project folders.

- Current onboarding practices are fast and minimal which students enjoy,
- Staff already consider the Q drive as a resource, while researchers do not know of the resources on the Q drive.

The analysis of the higher level theme items were used to guide recommendations and findings contained in this report.



Affinity wall diagram: yellow-interview notes, pink-higher level themes, orange- highest level clustering

Current Strengths:

Through our analysis, we were able to determine a number of strengths within the current system that have been either preserved or encompassed within our recommendations. Relating to the current onboarding practices, it was revealed that researchers most enjoy the fast and minimal process. Best practices in research on orientation programs indicated that orientation programs should be used to establish social connections for employees. Hacker (2004) suggests introducing newcomers to all employees within an organization and creating a mentor/mentee relationship between new and current employees (p. 173). Many of the student researchers that were interviewed indicated that this was done in practice, as they often met with their research teams upon arrival at the CSS.

Specific to the structure of the data, the open accessibility, central location, and thought out folder structure were all considered to be an important asset. This also coincided with the research found on data management practice. A study by Shen (2016) found that while researchers prefer to use personal storage devices for their research, roughly 27% of researchers thought that their data had the potential to be

used in cross-disciplinary research. With the CSS being an interdisciplinary research lab, this is an important consideration. Lastly, having the ability to access the hard drive remotely is a critical component. This was something that was also highlighted by the individuals interviewed as something that was enjoyable for all that used it.

Finding & Recommendation 1:

Finding 1: There is a lack of communication between administrative staff and student researchers

Transferring CSS data from the shared drive to the archive drive has been a major problem that many professional staff members have pointed out. Part of that problem stems from how the current organization and structure is not followed by student researchers. Instead, student files and work are updated irregularly and organized in the way they see fit. The drive ends up being organized and dependent on institutional knowledge. As a result, professional staff members put more effort than should be expected when utilizing the Q-drive and transferring files from the shared section to the archive section. In fact, one staff member mentioned that “as time goes on and on, it is scary to look at that folder” when the shared folder was brought up in interviews. A task as simple as pulling slides from a presentation or referring a student to previous work can become overwhelming with how the Q-drive is organized at the moment.

However, this could be a result of some mismatch in expectations between professional staff and student researchers. For example, professional staff expect that student researchers are working with inherited data and have a problem with the process of transferring that data. When we asked students about this issue, each student researcher denied using any inherited data and did not consider it a problem regardless since students on the same project generally experience some overlap between incoming and exiting student researchers. In addition, it seems that students prefer to work off site rather than in the center, contrary to what professional staff assumes.

These are relatively minor examples in comparison to the lack of awareness that student researchers have about professional staff’s needs and work. Students do not know expectations with regards to the shared drive - and look at it as an optional tool rather than a package that needs to be submitted upon exit. During interviews, students mentioned having no knowledge of the purpose of keeping things organized on the Q-drive. Because students are unaware that their work will eventually be used for other purposes at the CSS, the motivation to keep a package that fits the needs of professional staff is just not there.

On a related note, students also talked about how the Q-drive is not a priority for

them. While the students we interviewed have not exited the CSS yet, many expressed that they could imagine how organizing and uploading their files onto the Q-drive could be easily overlooked.

When looking at the difference between professional staff and student researcher perspectives, we noticed that professional staff and student researchers have different priorities. This could be a product of what was mentioned above - a lack of communication about the archive's purpose and how to maintain the data the CSS needs. As such, professional staff and staff members that have been at the CSS longer are left with a larger burden to alleviate the chaos of the Q-drive because of the institutional knowledge they hold.

Recommendation to address Finding 1: Communication, standardization, and tools to help facilitate both

What Needs to be Communicated:

The Q-drive would benefit from standardization, meaning that the tags, folder structure, and file naming conventions should follow a simple formula that is enforced at the CSS. Below are some suggestions for how, specifically, researchers can keep their work findable:

1. Standardized Metadata

More on this will be discussed in later sections, but if you choose to have a tagging system for your files, we recommend requiring some standard tags such as: project name, P.I./first author, categories of research, publication or presentation year, conference or journal name, and research method employed.

2. Folder Structure

This is a strength of the Q-drive. There already is a standardized folder structure that makes sense, especially on the archive side. However, the CSS may want to consider organizing folders by project or P.I. rather than student researcher.

3. File Naming

To ensure version control, unfinished but regularly updated files could be named in the following manner: FILENAME_DATE_INITIALS where the date and initials depend on when the file was last updated and who updated it. Other important information you may want to include may be the level of progress. For example, a file may be named "CSSsuggestions_12102018_PAM_working.docx" or "CSSsuggestions_12102018_PAM_SUBMITTED.docx" depending on if the author plans to edit the file further in the future or not.

Implementing Standardization:

To start a cultural shift towards standardization at the CSS, we recommend the CSS move forward with the following:

1. Make user friendly onboarding documentation. This can include how to use and maintain the Q-drive, best practices for data management at the CSS, the purpose behind the Q-drive for professional staff, incentives tied to keeping the Q-drive semi-regularly updated, and what a student's Q-drive folder should look like upon exit. Students also expressed interest in documentation regarding research resources, which should be included with the package but separate from the information regarding the Q-drive.
2. Distribute that documentation among current researchers. This will set the tone for future student researchers and help students who might not be aware of resources available at the CSS. If appropriate, perhaps distributing a copy to research advisors could be a helpful in holding students accountable.
3. Maintain and update these documents regularly. When changes are made, redistribute the documentation electronically to inform researchers at the CSS and serve as a reminder about expectations.

Tools to Encourage Communication and Maintain Standardization:

Our understanding is that a major strength of the CSS is that there is often in-person help available at the office. To enhance that presence and support for student researchers, we believe some online tools could be of use, in addition to regular email reminders. However, some of these tools cost a significant amount of money, so we understand that it may not be feasible to implement them. For example, Slack is a useful tool made for communication within project groups. In addition to reminders, documentation can be uploaded onto Slack for additional ease of access. Slack's channel features also allow researchers to group themselves based on projects and encourage collaboration within the CSS. While Microsoft Teams is a project management tool, Teams through Microsoft Office Suite has a similar communication component while also the ability to integrate Word, Excel, etc. and other Microsoft applications. This again would be useful for collaborations and can be a cloud-based alternative for the many researchers who use Google Drive.

Less technology-oriented solutions to establishing and maintaining the organization of the Q-drive are to provide physical reminders in the spaces that researchers work in and use recommendation letters as incentive. This could look like having a flyer or two up about the required naming conventions of certain files, dates of the month or semester personal folders should be up to date by, and where to find resources in the Q-drive. These flyers near computers and desks would establish that this maintenance is a researcher's responsibility and part of the culture at CSS. During the onboarding process, student researchers could then be informed that the timeliness

and quality recommendation letters would reflect how close to complete or maintained their personal folders look. This may serve as a more serious incentive for researchers to manage their data and files.

Finding & Recommendation 2

Finding 2: The Q drive organization is ill suited to the center's needs

Partially as a result of haphazard communication between researchers and administrators, there is no standardized system of naming or filing across the the shared side of the Q drive. Even in the archive, a successful search comes down to institutional knowledge. For instance, knowing who works or worked on which project, or maybe what the file you're looking for might be named so you can search for the right term.

Furthermore, while giving each researcher their own folder to organize to their liking makes it easy for researchers to know where to put their work, there are some serious drawbacks. Not only does it contribute to the confusingness of the drive's structure overall, but researchers are less willing to use or even search for the resources available in others' "personal" files because they feel like they are intruding on other researchers' privacy.

Recommendation to address Finding 2: Short-term and long-term strategies

The simple, short-term recommendation:

Some of these issues could be addressed with an Excel sheet "database" that lives at the top of each side of the Q drive. Given the dynamic state of the shared side, this will likely be a more useful recommendation for the archive, but with vigilant maintenance it could be relatively effective even on the shared side of the drive.

A single excel file would be housed at the top level of each side of the drive. There would be a file in each row, and pertinent information about each of these files in the corresponding columns. Critically, there would be a column that indicates the file path to the file. Staff could then add other columns to include other information about each file, like the type of file, what project it is a part of, or who created it.

The more complex, comprehensive, long-term recommendation:

Ultimately, any given organizational scheme for the Q drive (i.e. going by person, date, project, or file type) will be arbitrary on some level. It will also require researchers to adapt to a system that likely clashes with their preferences and won't necessarily help faculty and administrators find what they need. From our interviews with permanent staff, however, it seems more important that files are easily findable or searchable than it is that files are in any particular organizational hierarchy.

Taking a cue from the organization of the website, we recommend an offline integrated data management software that can add “tags” to existing files the Q drive. These tags are then searchable and create a powerful, ad-hoc system that attaches intuitive metadata to files in a helpful way. We found two examples of such software: Tabbles and TagSpaces. We will focus on the Tabbles infrastructure only here but the two systems are very similar. The system is very lightweight and meshes very simply with their existing infrastructure. The system is essentially an offline plugin for user-specific custom metadata. The metadata travels with the files regardless of where they are moved or stored as well, meaning if CSS staff decide to change how to structure things in the future they can still easily transition existing metadata. In addition, it means less structure in the archive drive is even necessary for effective management. It does, however, have its drawbacks; administrators will have to be very careful and proactive about applying the appropriate metadata manually.

Metadata tags can include project content, like it does on the website, but because tags are unlimited the number of content types can be more than 17. Tags can also indicate anything from file type (whether it’s a presentation/paper/data file) to researcher name to the project the file is a part of. Administrators can decide how broad or detailed they want the metadata attached to any given file should be so that they can search for it effectively in the future.

Finding & Recommendation 3

Finding 3: The VPN doesn’t work well for researchers’ needs

Although it doesn’t directly hinder researchers’ efforts, we identified the VPN as another point of miscommunication that has the potential to create further issues in the future if it goes unaddressed. Researchers appreciate that they are able to connect to the Q drive remotely; however, most of them use local resources and cloud-based storage for their research and either regularly or occasionally back up to the Q drive.

This means the Q drive is often not up to date with researchers’ current progress and that researchers are not accustomed to the idea that their research must “live” on the Q drive for administrative purposes in addition to collaborative facilitation. According to the information we gathered during interviews with researchers, they appreciate that remote access to center files and resources is available and would not be happy if that were to change; however, they are accustomed to cloud-based resources that are supported with robust infrastructure, which a VPN system lacks. As technology progresses, this issue will only become more of a frustration for new researchers who come in accustomed to robust proprietary file storage resources.

Recommendation to address Finding 3: Short-term and long-term strategies

The simple, short-term recommendation:

Feeding into our general recommendation for improved communication protocol at the center, we advise in the short term to start sending monthly, or at least once-a-term email reminders to researchers to organize and back up their local research files to the shared Q drive.

Although this will not resolve researchers' lack of motivation to use the Q drive as their primary research data storage system, it will encourage a pattern of behavior amongst them that will both keep the shared drive relatively up to date and keep them in the habit of including

The more complex, comprehensive, long-term recommendation:

Git is a free version control system that works for local or remotely stored files. It is a powerful way to keep access to previous versions of documents and to keep different parts of big projects synchronized across multiple computing devices and servers. Although it is most powerful when managing code and data-centered projects, it can be very useful for use with any digital files. Users can interface with the software either with their computer's terminal/command line or with a number of open source graphic interfaces. GitHub is a remote host where users can create file repositories that can be added to, edited, duplicated, and copied onto a different machine amongst other functions all using git software. There is a learning curve with the software, but once the basics are mastered it could make for very efficient file management for a research institution like the CSS.

Because git repositories can be housed and copied almost anywhere, it can be integrated with whatever local file management system researchers choose to use. Put simply, these tools can be used to emulate the functionality of a Google Drive-like system (which researchers seem to prefer) without abandoning the Q drive (which administrators seem to prefer).

For the center's purposes, there are a several setup schemes that might work, but we recommend the following: step one would be to install git on center computers and create a center account on GitHub. Next, to have researchers and project leaders start using git and private GitHub repositories. Then, for researchers to maintain their files and to add the CSS account as a collaborator on their research repositories and regularly back their work up from their local drives to their corresponding GitHub repositories. Researchers will then be able to use git both to keep their files backed up to GitHub and to have easy access to older versions of their work. Administrators, using

the CSS GitHub account, can then download and regularly update the Q drive to reflect researchers' changes to their repositories. Git makes this a very fast and efficient process.

Ultimately, implementing this system would take substantial effort and research on the part of CSS staff. However, we believe in the long term it offers the most flexible and accommodating resolution to the disconnect between researchers' and administrators' data management preferences.

Conclusion

Based on the findings and recommendation contained in this report, PAM squared suggests the the CSS utilize the strategies to help address the problem with the current Q drive system. It is also important to note that the recommendations outlined consider the current strengths of the CSS operations in relation to the Q drive system. Both staff and students at the CSS indicated that they would receptive to changes to the system if positive improvements were made to deficient areas. Given this, any changes to the system should strive to preserve the current strengths of the Q drive system.

To address the lack of communication between permanent staff and researchers, PAM squared has suggested a number of strategies and tools to help improve upon the communication flow at the CSS. This will help to ensure that both staff and researchers are aware of the expectations and the importance of incorporating data into the archive.

For the issue with data storage, recommendations have highlighted the importance of changing the current culture with the Q drive. Organizing the data by project component instead of researcher names and by utilizing a tagging system, will reinforce the idea that the Q drive can be used and relied upon as a resource to both permanent staff and researchers alike.

With the vpn concerns, there may also be opportunity for the CSS to utilize a web based cloud server to backup information from the Q drive. This will help to provide researchers with a reliable network to which to store their information should the CSS feel comfortable with storing their information via an online service.

PAM Squared would like thank the Center for Sustainable Systems for the opportunity to work with them. We would also like to thank all of the individuals we interviewed for allowing us to gain a better understanding of CSS operations so that we could better guide our recommendations.

References

Beyer, H., & Holtzblatt, K. (1998) .Contextual Design: Defining Customer -Centered Systems. San Francisco, Calif: Morgan Kaufmann.

Hacker, C. (2004). New employee orientation:Make it pay dividends for years to come. *Information Systems Management*, 21(1), 89-92. doi: 10.1201/1078/43877.21.1.20041201/78989.9

Shen, Y. (2016). Strategic planning for a data-driven, shared-access research enterprise: Virginia tech research data assessment and landscape study. *College & Research Libraries*, 77(4), 500-519. doi:<http://dx.doi.org.proxy.lib.umich.edu/10.5860/crl.77.4.50>

Appendix

For more information on recommended technology tools:

- ★ Slack: <https://slack.com/features>
- ★ Microsoft Teams: <https://products.office.com/en-us/microsoft-teams/group-chat-software>
- ★ Tabbles: <https://tabbles.net/>
- ★ TagSpaces: <https://www.tagspaces.org/>
- ★ Git: <https://git-scm.com/>
- ★ GitHub: <https://github.com/>
- ★ Open Source graphic Git interfaces: <https://git-scm.com/download/gui/windows>

Link to Client Presentation:

<https://docs.google.com/presentation/d/1fo4ldBEiW9euJ47Jaf6O1kFEMXkefSfyk3XBn04t61g/edit?usp=sharing>