

File Synchronization and Sharing: User Practices and Challenges

Robert Capra, Emily Vardell, Kathy Brennan
School of Information and Library Science
University of North Carolina at Chapel Hill
{rcapra, evardell, kbrennan}@unc.edu

ABSTRACT

In this paper, we present results of qualitative analysis of interviews with 16 university faculty regarding their use of file synchronization and sharing technologies in their work and personal lives. We identify key problems and considerations that influence file synchronization and sharing practices, and describe commonly used strategies. Our findings show that while behaviors have evolved as technology changes, users continue to make extensive use of familiar patterns of access, including automated and manual methods. We identified several factors that influence users' choice of methods for synchronization, including privacy and security concerns, data size, institutional policies, and knowledge of systems. We present insights about how boundaries and data separation affect synchronization practices, and about specific synchronization issues users faced when collaborating. Our findings extend research on personal and group information management practices and provide insight for the design of synchronization and sharing systems.

Keywords

File synchronization, file sharing, personal information management, group information management

INTRODUCTION

Information users today face many challenges in managing information across multiple electronic devices, storage systems, and collaboration tools. Information workers often have multiple computers and mobile devices both at work and at home. Cloud-based services provide new ways to store and share information across people and devices but also add complexity to information management.

A recent survey of 895 technology experts conducted by the Pew Internet project (Anderson & Rainie, 2010) found that many experts believe the next decade will bring increased

adoption of a hybrid computing model in which some work is done in the cloud and some is done on personal devices. This suggests additional challenges and opportunities for synchronization and sharing technologies. Users need to have confidence that data will be available when and where they need it, and that they will be able to limit access when desired. Many individuals and organizations are making increased use of tools such as Dropbox, Google Drive, and SharePoint to support collaboration and shared access, but these systems can increase data fragmentation and raise concerns about privacy, trust, and control.

Relatively little research has focused on how users synchronize data across devices, share data among collaborators, and what problems they encounter when doing so. Yet the cost of synchronization and sharing failures can be significant – users may not be able to access information when or where they need it and may spend considerable time and effort developing practices to work around limitations (Dearman & Pierce, 2008). Increasing use of cloud services is spurring practical interest in this topic to inform system designs (Tang, Brubaker, & Marshall, 2013), and more research is needed to help understand users' needs, practices, and challenges.

Our research focuses on how people handle the file synchronization and sharing aspects of “personal” information management (including both work and personal information under the person's control). We define synchronization broadly to include both automatic and manual processes and the associated issues of managing files that are shared across devices and among people. The primary questions that drove this research were: how do people make sure they have access to documents when and where they need them, what considerations do they have in choosing methods for synchronization, and what issues and problems do they encounter?

We identified 16 users (university faculty) of file synchronization and sharing technology and conducted in-depth, semi-structured interviews with them about their strategies and practices. We present findings that emerged from our qualitative analysis of the interview data, organized along four main topic dimensions:

1. How do people access the files that they need when and where they need them? How is this access maintained?

2. How do people synchronize files across devices, and what considerations and issues do they have in doing so?
3. What boundaries do people maintain to keep information separate, and how does this influence the methods used?
4. How do participants manage files and synchronization when collaborating with others?

We also report on several cross-cutting themes and discuss factors that influence participants' use of synchronization and sharing technologies.

RELATED WORK

Understanding and Use of Synchronization

Several recent studies have focused on users' understanding and use of synchronization technologies. Through surveys and interviews, Marshall and Tang (2012) identified a hierarchy of concepts that users need to understand to effectively use synchronization and sharing systems. The hierarchy included the concepts of personal versus shared repositories, replication, synchronization, triggers, and conflicts. Participants in their study had varying levels of understanding of synchronization and sharing services and used such services for a variety of purposes: transferring files among devices and people, backing up files, temporary shared access, synchronization, and collaboration. Participants also noted problems that arose in maintaining synchronized folders when collaborating, issues when collaborators had different levels of understanding about how synchronization systems worked, and concerns about privacy and data security. Dearman and Pierce (2008) also found that users may not fully utilize synchronization and sharing technologies because of issues with trust, visibility, and ease of use. Instead, they reported that users may make use of manual methods to transfer, synchronize, and backup files, a finding echoed by results from a study of music management practices by Brinegar and Capra (2011). In related work by Volda et al. (2006), they noted that users may use email instead of group repositories to share files.

The study by Dearman and Pierce (2008) focused on the multi-device information practices of academics and business people. They reported four main findings: 1) activities of a single user often span across devices, 2) devices often have roles defined by users based on circumstances, 3) separating work and personal activities is challenging despite users' efforts to do so, and 4) even though users have strategies for accessing information across devices, they still encounter many problems with it. Connected research on personal information management has investigated issues such as how people maintain access to information, how information can become fragmented across tools and systems, and the how people organize and manage files (Jones & Teevan, 2007).

Collaboration and Cloud-Based Services

Other recent studies have investigated how people are using cloud services and synchronization technologies such as Dropbox and Google Drive for collaboration. Massey,

Lenning, and Whittaker (2014) identified two main barriers that hindered effective use of group repositories through their interviews of 27 people from different workplaces – the first was disagreement about how information should be organized for group use and the second was a non-conforming behavior of some individuals who ignored the group process by emailing files instead of using the shared repository.

In a case study of a group information repository, Rader (2009) observed that users distinguished between their own files versus group members' files and were more knowledgeable about files they had added. Confusion about which were working files and which were older contributed to cluttered and poorly organized group shared folders. Her participants identified a desire to have notifications of changes and additions in group documents.

Volda, Olson, and Olson (2013) also reported on challenges that users faced in using cloud-based services and identified influencing constructs: the use of different digital identifiers to represent different real-life roles, the use of different services for different purposes, and users' participation in different collaborative groups that each had different norms and work practices. Their participants engaged in aggregation (e.g., trying to keep things in one service) to simplify complexity but also tried to use segmentation to keep identities separate. However, maintaining segmentation was often reported as a "losing battle," with aggregation sometimes winning out. Participants also noted problems with collaborators who did not have access to the shared space and collaborators who used different naming and organizational schemes. They reported that teams often decided which cloud service to use based on what a majority of the collaborators had access to or experience with rather than specific features. Other work has noted the difficulty that end-users have understanding access control systems and investigated ways to help users in managing access to shared information items (Brusoloni et al., 2008).

Managing Boundaries

In addition to keeping information coordinated among multiple devices, synchronization and sharing technologies also need to help users maintain boundaries and manage information in the context of segmentation. One example of such a system is Cimetric (Marshall et al., 2012), which was designed specifically to support segmentation in collaborative work. One of its main capabilities is "bi-level" synchronization that distinguishes between each individual's private working documents and a different set of files that are being shared among the group.

Social scientists have explored the ways in which people set and maintain boundaries among their identities and roles, and the effects that technology has on these boundaries (Clark, 2000). As reported by Volda, Olson, and Olson (2013), difficulties using technology to maintain boundaries can also impact boundary crossing in ways that users may wish to avoid. A prior study of university employees found

that email played a role as a boundary artifact but that there was also a fair amount of boundary permeation in email use (Capra, Khanova, & Ramdeen, 2013).

METHOD

Our objective was to gather information about participants' file synchronization and sharing practices and their impressions about how these processes worked. We used a semi-structured interview as our data collection method. We developed an interview guide with essential questions as well as suggestions for follow-up and probing. A sample of our questions is shown below:

1. What information do you synchronize between electronic devices and/or services?
2. Why do you synchronize this information and data?
3. How do you synchronize your data?
4. What synchronization tools do you use?
5. What level of confidence do you have in your synchronization processes (i.e., yourself, the software, the electronic devices, and websites)?
6. What problems have you encountered?

At the beginning of each interview, we asked participants to complete a "data ecology" diagram on which they indicated information about their devices (e.g., desktop, laptop, tablet, mobile smartphone), the location of the devices (e.g., home, office, lab, mobile), and their synchronization methods (e.g., USB flash drives, shared network drives, cloud applications). We tested the interview guide and ecology diagram with several faculty members in our own department and made revisions based on their feedback.

Participants

We recruited and interviewed 16 faculty members from a large US university. Recruiting was done via university and departmental listservs. Our goal was to interview people who had experience with and needs for synchronization and sharing technology. University faculty frequently have file synchronization and sharing needs, and those needs are similar to those of other mobile information workers (e.g., working from multiple locations, collaborative projects, use of multiple computers and devices). After the initial recruitment email, we used a screening questionnaire to help select a final set of participants who had experience with these technologies. Our final set of participants came from a wide variety of academic units and departments (Note: in the text, participant quotes are labelled P11-P27).

Interviews

Interviews were conducted at a location of each participant's choice (most chose their offices, but a few chose to be interviewed at our lab) during the spring of 2012. Interviews typically took 45-60 minutes. During the sessions, we collected notes, diagrams, and audio recordings that were used to produce full transcripts.

Since we were conducting interviews at our university, we informed participants that they were not required to disclose any data or practices they did not wish to for privacy or security reasons. We also let them know that while quotes from the interviews might be published, the quotes would be anonymized.

After obtaining signed consent, we began the interview by asking each participant to complete the data ecology diagram. This exercise typically took about 10 minutes and was used as a focal point in the semi-structured interview, often functioning as a point of common grounding. At the end of the interview, we offered participants a \$20 incentive and thanked them for their participation.

ANALYSIS

We used qualitative analysis techniques to investigate the main themes of the interviews. We used a hybrid approach in which we started with the main areas of interest that were used to design the semi-structured interview questions, and then, in our open-coding, we allowed themes to emerge from the participants' responses.

We randomly selected three interview transcripts to begin open-coding. Two authors independently analyzed these three transcripts, establishing codes that emerged from the interviewee comments. We then compared our emergent codes and identified an initial coding scheme. In this initial phase, we determined two major themes: strategies for file syncing and strategies for file sharing. Beneath these higher level codes, we established many sub-themes with additional aspects (e.g., separation of work/personal and the strategies used to address this). This scheme was used by one author to code the remaining transcripts, with some additional codes identified as a result of newly emerging themes. Once the author had coded all transcripts, a second author reviewed the coding for verification. Any discrepancies were resolved through author consensus. NVivo qualitative analysis software was used to facilitate the coding and summarization process.

RESULTS

We present results of our analysis in four main areas: 1) How are files accessed, 2) How are files synchronized among devices, 3) How do boundaries and data separation affect synchronization practices, and 4) What issues are encountered when collaborating. We also report on several cross-cutting themes and throughout the sections we discuss factors that influence participants' use of synchronization.

How Are Files Accessed?

Participants described a variety of methods for sharing, synchronizing, and maintaining access to files they need. In this section, we describe methods they used to maintain access to files when working on different computers and locations. In the next section, we discuss methods used for purposeful synchronization of files across different devices. These two activities are related – synchronization can be a way to maintain access – but not all access methods

involved keeping permanent synchronized copies. Often, the focus of maintaining access is on a smaller set of files that the person is currently working on and may be based on dynamic factors such as what files are needed at this particular time. On the other hand, synchronization typically involves a longer planning horizon and may focus on keeping exact replicas of an entire directory structure.

Methods of Maintaining Access

Participants described six primary methods for maintaining access to files: using a remote mounted drive (15 participants), USB drives (14), carrying a notebook computer (14), emailing files to themselves (12), the use of cloud storage (10), and using a remote desktop connection (8). Among these methods, participants noted portability, ease-of-use, and universality as advantages of USB drives, and several people described carrying them on their person so that they would always have access to files when they were needed. Emailing files leverages the familiarity and ubiquity of email, but some participants mentioned that this method was not practical when many or larger files were involved. Carrying a notebook computer was often described as a method used when traveling or in other situations in which they did not want to rely on a computer at their destination.

Many departments have networked file storage that can be mounted, and this was a common method discussed by 15 of our participants to maintain access to files. One participant even noted how an IT upgrade allowing him to mount department drives from home had greatly improved his workflow. Many participants (8) also made use of remote desktop software to provide direct access to their work computer when at home, noting its advantages as not having to manually transfer files and having access to programs as well as files. However, drawbacks were also noted. These included limits on screen resolution, desktop icons getting re-arranged due to screen size differences, and lag due to slow connections. Several participants described remote desktop as a “fallback” option if they forgot to use some other method providing access to a file from home.

Some participants worked on a small number of files at a time and leveraged this to maintain and control their access:

P26: “I have a little folder... it is just a repository for whatever I happen to be working [on].”

Several participants discussed ensuring access through redundancy:

P18: “I’ll have like a triple system. So, I’ll have it on my laptop, I actually save it to a flash drive just in case my laptop dies, and I also email it to myself.”

Participants also used cloud storage services to maintain access to files, which we will discuss in a later section.

Download-Edit-Upload

Many participants (9) described using an interesting pattern of access that we refer to as *download-edit-upload (DEU)*. In the DEU pattern, they would: 1) copy or download a file from its primary location (e.g., a desktop computer or file share) to the local computer that they were currently using (e.g., a computer at home), 2) work with the copy of the file on the local computer, and then 3) upload or copy the modified version back to the primary location. This pattern was described in a variety of situations, with several common characteristics. First, the user had a desire to work on the file locally, often for speed issues:

P17: “If I’m at home, that’s where the issue comes in of how fast a network you have. So, if I’m at home, I’ll generally download it, then work on it, and put it back.”

Second, participants who used this pattern were careful to keep track of the authoritative version of the file:

P22: “That is the master document when I’m on the road. When I’m here the master documents are always on this computer. If I take a document from this computer... [and] work on it someplace else... I will, at some point move it... back into a folder on my desktop here.”

Third, we note that participants undertook the DEU process manually rather than relying on synchronization software. In several cases, participants even described processes that involved “extra” steps to achieve this pattern. For example, P19 manually copied files to and from Dropbox in a DEU pattern rather than relying on its synchronization feature.

The download-edit-upload pattern played an important role in many participants’ daily work and offers several important benefits. DEU is inherently a manual process, so users retain complete control over it. It is effective without relying on automated synchronization software. In addition, DEU is a simple process that only requires the user to know how to copy files from one place to another. In contrast, many synchronization programs require at least some level of configuration and understanding of the synchronization model (Marshall & Tang, 2012), and many require synchronization at the folder level rather than single files. For daily work with just a few small files, DEU is a flexible, relatively fast process with a side benefit that a temporary “backup” copy of the file is maintained at the original source. The files to be copied can be decided dynamically at the time of need rather than having to be designated in advance for synchronization, making DEU well suited for temporary working needs.

Summary

Our participants employed multi-faceted approaches to maintaining access that involved factors such as the number and size of files to be copied, the types of computers and Internet access that would be available at the “other” location, and the costs of forgetting to manually make a copy. As illustrated by participants who viewed remote desktop as a “fallback” safety net for access, some

participants used multiple tiers of methods tailored to fit their particular needs. Many of these underlying methods of maintaining access have persisted for many years, being realized in similar, but evolving forms as technology changes (e.g., see examples presented in Jones, Bruce, & Dumais, 2001). In the Emerging Themes section, we discuss how people transition from old methods to new ones as technology changes.

How Files Are Synchronized

In the previous section, we considered how people maintained access to a current set of working files. In this section, we focus on synchronization of files for the purpose of creating on-going, purposeful replicas across devices. We include situations where the goal is to synchronize entire file systems, specific directories, or just subsets of files. However, we distinguish this type of synchronization from the download-edit-upload (DEU) pattern reported in the previous section. Here, the purpose of the users' synchronization is to maintain a permanent replica (or archive), whereas in the DEU pattern, the main focus was to maintain current access to working files on a different device for a temporary period of time.

Use of Automatic Synchronization Software

Ten of our participants discussed the use of automatic synchronization systems, including specialized software such as rsync and Unison, systems for archival and backup such as Apple's Time Capsule, and cloud-based services such as Dropbox. Participants noted the conveniences of automated systems (e.g., not needing to remember to sync, no extra steps) but also pointed out limitations:

P15: "It's one folder at a time. You can't just pick certain files within that folder. You have to synchronize the whole folder, which will also do all sub-folders."

Manual Processes

Eight participants described varying degrees of manual processes for synchronizing files. In some cases, *initiating* synchronization was done manually, but the process was then automated after it was started (e.g., backing up to an external hard drive or manually starting a program such as rsync). However, others described completely manual processes. For example, one participant described a process she used to maintain a near-replica of her file system between work and home:

P19: "Anything I worked on, I would just put on the flash drive and then I would take it home and load that to my laptop... I wasn't always good about doing this every day, but I would say that either once a week or every two weeks... I would transfer them from the laptop to the desktop... And the file folder arrangement really does look pretty identical [on] all my devices."

P17 described a complex process that involved several manual components regarding data storage and analysis in a research lab:

P17: "It's on you to make sure that it gets synched back up to the server and stored properly... we don't have any of the more automated... type of stuff... that's all manual. And then [for analysis], people move them on to the server for permanent storage."

When manual processes were relied on, participants described having to remember to sync files and discussed the consequences of forgetting (e.g., not being able to get work done at home, having to use a slower access method).

These contrasts between automated and manual processes illustrate a recurring theme among our findings – there is a tension between users' desire for control and understanding of what is happening when they sync versus wanting the ease and "set-it-and-forget-it" of automated systems. One way that users resolve this tension is by using multiple methods; another is by falling back on simple, well-understood methods such as manual copying.

Size and Number of Files

A majority of participants (12/16) indicated that the size of the data influences the sync strategy used.

P11: "For smaller things, I put them on a memory stick, a USB key. For real small things ... I'll email it to myself. For larger files... [I] copy the files to [a network drive] and then at home I can access the network and I can either just read them on there or copy them over."

Customized Solutions

Three participants from technical backgrounds described designing solutions tailored to meet their synchronization needs. For example, P16 developed a set of custom scripts to synchronize his work and home computer that were configured to run every 15 minutes – the amount of time that it took him to travel between home and work.

Cloud Synchronization Services

It was clear from the interviews that cloud-based storage and synchronization services were becoming increasingly important for many participants. Ten participants mentioned general use of cloud-based services (e.g., for maintaining access to files) and five described using them for synchronization purposes. Interestingly, some participants used Dropbox for maintaining access but did not fully take advantage of its automatic synchronization features. For example, P19 used Dropbox to facilitate manual synchronization and kept a "truncated version" of her file structure on Dropbox to support this process:

P19: "Sometimes I use the flash drive and then I've been trying to use Dropbox. So I would move the files to Dropbox and when I got home... I would just open Dropbox and... copy back those files... So there's a permanence to the file system that I've created for my laptop and my desktop, but for Dropbox, I really see [it] as a transfer system, so [its] file system is one that facilitates me being able to transfer things."

It was not clear from our interview with P19 if she was unaware of Dropbox's automatic synchronization features or if she consciously preferred do manual synchronization. Related to this, six participants described Dropbox as a place to store "working" files:

P27: "So if there are some temporary things that I just want to live on Dropbox for a while, like while I am working on a grant proposal or something... I'll use it for that. But I use the [departmental] file servers as more permanent storage."

Using cloud-based services to store working copies of files can also add to fragmentation and duplication of files:

P22: "I've got multiple versions of the document... one on Dropbox, one on my desktop, one in My Documents. And, even worse, one attached to an Outlook email. Then, even checking the latest date becomes... cumbersome."

Privacy and Security Concerns

Eleven participants cited concerns regarding data privacy, or security restrictions required by law or intuitional policies as factors that influence which tools or strategies they use to store and share information. While these types of restrictions are essential to maintain privacy and security, they also create complexity in users' synchronization practices. Systems need to help support users in creating and maintaining these types of boundaries.

P25: "I still have a little bit of... suspicion about [cloud technology]. Like is it really always going to be accessible or is it really secure? Are there things I should be concerned about keeping documents in that form?"

Complexity of Sync Technology

Synchronization technology (in its various forms) can offer users great power and control to manage their information across devices. We observed a natural tension between users' desire for understanding and control over how their information is stored, copied, and transferred versus their desires for simple, easy-to-use systems and processes. Amidst these tensions, the use of familiar, manual processes offered re-assurance:

P25: "I have a desire to not... have to think about where I'm transferring it. It would be wonderful to have it in one place. But my other concern is how much stuff would be in that one place, and what would be my structure and organization for getting it? Whereas, I can tangibly separate what's on USB drives."

Participants also described weighing the benefits of investing time learning new technologies versus using synchronization processes that were already in practice:

P11: "It's usually a limited subset of things that I want in different places. I suppose there are ways to set that up [automated sync], but I think it's more hassle than doing it [manually]."

On the other hand, we observed cases where participants had developed complex processes using combinations of technologies to meet their needs. For example, while at home, P22 used the following process to post a document from his work computer to a courseware system:

P22: "the document that I need is on my desktop. So I use remote desktop, get the document, put it on Dropbox through remote desktop, get out of remote desktop, go into the document from Dropbox, finalize it, log in to [the courseware system], and post it."

While there may have been other ways to achieve the same goal, P22 invested effort to develop a working solution using a combination of technologies.

In other cases, the complexity or design of the technology may obfuscate features. For example, many cloud-based synchronization services have both web-based and desktop clients. For the desktop clients, often the default installation is to create a special folder for things that should be synchronized. While this is a good default from a privacy and security standpoint, users often form mental models of software based on default configurations, and may not invest time to learn configuration options. For example, in the quote below, the default configuration of Dropbox seems to have encouraged an impression that it could not synchronize other folders:

P26: "The ideal, maybe this exists... it's probably a Dropbox feature that I just haven't been able to figure out yet... when you click save, you don't have to worry about putting it somewhere weird, or shipping it to another folder... I think where I have a problem using Dropbox and stuff like the cloud... you still have to manually put it there and I always forget. So if I have a USB drive... I just click save to it and it is done."

Participants also discussed gaps in their knowledge about the details of specific sync technologies. For example, some participants were not entirely sure about where files were physically stored when they used cloud-based synchronization tools such as Dropbox.

Summary

Participants used both automated and manual processes for synchronization and noted the impacts of needing to remember to sync. Size and number of files are often used to determine syncing strategy. Cloud-based services such as Dropbox were used but not always to their fullest extent. Privacy and security concerns and policies played a significant role in how people used technology to manage and synchronize files.

Boundaries and Work-Personal Separation

Work-personal separation and the management of files across this boundary was a commonly discussed topic during the interviews. All our participants were university faculty, and many conducted work at home and outside their office. A majority (12/16) reported that they had

separate work and personal computers. However, participants reported different approaches and philosophies about how they handled their *personal* PIM and their *professional* PIM. Some kept personal files separate from work files, while others were willing to mix the two to varying degrees. These results are consistent with and extend aspects of work/family border theory (Clark, 2008), illustrating how a person's file synchronization can play a role as an artifact in maintaining (or blurring) work-personal boundaries.

In the personal-to-work direction, two participants were concerned about keeping personal files off their work computer:

P16: "The home one has... our personal things, you know, tax documents and pictures, which I don't need to share."

Some participants were willing to keep less private types of personal files for use at work (e.g., music), but even these might be kept on a separate storage device:

P12: "So this [computer] has no personal information on it. Well, that's not entirely true. I have a separate drive that has all my music on it... but it's a separate drive."

In the work-to-personal direction, the use of work documents on a personal computer was discussed by six participants, but often as a temporary measure:

P15: "I park things on the desktop... especially if I think that eventually where I want them is at work. So I'll park them there to remind me to transfer them... A lot of times, if they're on the desktop, I know they're there temporarily."

When asked why work/personal separation was an important issue, two individuals cited security concerns and one of them also cited the importance of appropriate use of institutional resources:

P18: "My husband and I have had lots of discussions about security of things and just want to... keep that line separate. I don't want my personal family photos on my work laptop. I don't want my personal tax information on my work laptop... I just never want there to be a question of what I'm using [institutional] resources for"

These concerns highlight a key tension that our participants faced with synchronization and sharing technologies. As mobile workers, they had needs for flexible access from different locations to their currently active set of files. However, they also had needs to keep some files separated and secured, both on the work side and on the personal side. In response to these needs, participants used manual methods of *intentional fragmentation* (e.g., keeping work and personal separate) and *specialization of equipment* (e.g., using a separate computer only to review medical records, using a separate hard drive to store music).

Collaboration and Synchronization

In this section, we summarize methods and issues related to synchronization and sharing when collaborating.

Methods to Share Access

The most common methods of sharing access mentioned by participants included Dropbox (12), email (10), Google Docs/Drive (9), use of a shared server (6), and manual transfer such as handing over a physical USB drive (4). In addition to Dropbox, seven (7) participants mentioned other tools used for collaboration including: Sharepoint, Unison, EndNote, Yahoo listserv, FTP, Sakai, Zotero, and Wordpress. The variety of methods used illustrates both flexibility and challenges for collaborative PIM. On the one hand, users have many methods available for collaboration. However, team members must agree on tools and strategies for organization (Massey, Lenning, & Whittaker, 2014).

DEU for Control in Shared Collaborative Spaces

Two participants described uses of the download-edit-upload pattern noted earlier to support collaboration:

P18: "I would download it to my computer and make the change, and then.. change the file name and put it back on Dropbox so they would know that I had messed with it."

Providing "safe" personal spaces for collaborators to work individually can be an important aspect of collaboration (Marshall et al., 2012). When using shared storage space, the DEU pattern has advantages in that it provides a collaborator manual control, allowing them to move a file into their personal space for editing and/or experimentation before committing the changes back to the group. This flexibility also does not require coordination with the group – yet another way in which it supports users' desires for control of their data. However, a downside is that the practice could lead to conflicted file versions.

Issues and Strategies with Shared Access

Six (6) participants discussed simultaneous editing of collaborative project documents. The ability to simultaneously edit documents in tools such as Google Drive is a feature that several participants liked, and three participants specifically mentioned using Google Drive to facilitate collaboration.

P16: [Describing a remote collaboration using Google Drive] "We spent an hour and a half together yesterday morning working and editing on the file... He was really happy that he could not only see what I was doing, but he could also change the other entries... We collaboratively worked on it, and now we have something really good."

Other tools that are focused primarily on file storage often do not include built-in support for simultaneous editing. However, many of these detect and flag when multiple, "conflicted" versions of documents have been created. One participant described being cautious about this:

P27: "I haven't run into too many issues where two of us were trying to edit a file at the same time. So, I think we have just been able to be careful about it."

Another participant indicated that the number of collaborators distinguished the projects where multiple versions were less of a concern:

P17: "The only time where I've run into questions – on the rare occasion when I know there's been a problem – is where we've had many people, like four or five people, editing something. That gets to be a little tricky. But if it's just something I'm working on, or just two of us... I'm very confident that I always have the most up to date."

Eleven participants described having issues with multiple versions of the same file, seven of which specifically mentioned issues when collaborating (rather than managing their own personal files). Participants described using the following methods to address issues of multiple versions:

- File renaming (11 participants): "Typically we'll do draft whatever number, our initials and the date, so that we know that it's the latest and who did it." [P25]
- Out-of-band token passing (10): "There's four of us, so we were very clear on who was doing edits when. So, [Person X] did edits for the one week when I was away and then she passed it off to me, and then I just passed it off to the next person." [P18]
- Track changes (4): "Then one of the team members wanted stuff emailed, like let's email documents and we'll have Track Changes on." [P23]
- Levels of access (3): "We have differential access for the faculty and students who are accessing it." [P25]

Management of Information by Groups

Participants also discussed how the behaviors of an individual in a shared storage space can impact the other group members: running out of space, collaborators not deleting files, multiple versions, not saving things, and the effects of one person "cleaning" or moving items in the shared space.

P17: "We're still trying to come up with a [lab] policy that works... [Person X] never deletes anything. He has hundreds of gigabytes of files, and he can't find anything ever. [Person Y] almost never saves anything... everyone's supposed to not copy, but move the files. But that doesn't always happen."

Non-Tech Teammates

Six participants specifically mentioned how the use of technology for collaboration was impeded by non-tech-savvy teammates. In particular, they noted how teammates who were not as adept at using the technology (or who simply refused to use it all together) often ended up not being equal partners in the collaboration process:

P23: "The main issue was that two of the team members wouldn't participate. That was a big problem I thought, because it meant that they were not drawing on the

common body of materials that we were collecting. And anything they happened to know about or collect, if I were writing something, it was not factored in."

This finding is complementary with prior work that has investigated sources of co-organizational failures and problems groups encounter when using shared repositories (Rader, 2009; Massey, Lenning, & Whittaker, 2014).

Preferences of Collaborators

In determining which tool or strategy collaborators would use on a project, four participants indicated that their choices were influenced by the preferences of their collaborators.

P13: "I have [both Dropbox and Google Drive] on my computer and have used them for specific projects. I'm collaborating with a group and they have a Dropbox site, or I'm collaborating with people that have Google Docs, but I don't use it personally for any of my own files."

Rader (2009) reported a similar result that collaborators' preferences had a strong influence in a group's tool choice.

The size of the files can also influence the choice of tool:

P17: "In the beginning, it's easy to email them around. Toward the end, after you've worked on them for a while, and you start to stuff figures into them, they get too big. And so that's when we just share across the server."

Emergent Themes

In this section, we present several themes from our data that cut across dimensions previously discussed.

New Method Transition/Development

Nine participants discussed aspects of how they had transitioned from old methods to new ones. Often these transitions were a result of technology changes (e.g., floppy to USB) or new software capabilities (e.g. Dropbox, new access policies at work, etc.).

Four participants mentioned transitioning from physically carrying USB drives to accessing files remotely. For example, one participant gave a detailed description of his transition from an older system of carrying a USB stick, to using remote desktop from home to access a network drive at work, to finally directly mounting a network drive from home. This last improvement was enabled by a software upgrade at work that allowed direct (and secure) mounting of drives from home. This evolution through multiple methods shows the adaptability of users and also underscores how technology and policy changes are continually impacting PIM and synchronization practices. Similar notions were echoed by participants describing their adoption of cloud-based storage services such as Dropbox:

P23: "Dropbox has basically taken over what used to be 'I'm putting it on my flash drive, I've got it on my neck all the time and I'm carrying it around.'"

Participants also described method transitions in how they created back-ups and archives:

P15: "I synchronize between the two computers and they function as backups for each other. I used to backup on the zip drives and, then I was backing up on CDs and I had a tape drive at one point. And this was just a mess."

Improvements in external hard drives have also had impacts on automated backup strategies:

P21: "Before you used to save everything on DVD, and now they have a two or three terabyte USB drive, USB 3.0 connectors which is very, very important."

Participants mentioned collaboration, workgroups, and co-workers as sources for learning about new PIM methods:

P18: [Describing a backup strategy] "I think I learned it back when I was a resident. My residency director told me that she did that... I always made it a habit after that."

"Should Do" Items

Despite reporting many sophisticated synchronization and archiving systems, there were many times that our participants expressed feeling that they were not well organized and that there were things that they "should do" to improve their organization and synchronization. These comments echo findings of previous studies of PIM (Jones, Bruce, & Dumais, 2001; Jones & Teevan, 2007), suggesting that although improvements have been made in software and technologies, users still feel critical of their own PIM behaviors.

DISCUSSION

Accessing files – Our participants' descriptions of how they access files paint a picture of how technological developments have improved workflow (e.g., remote desktop connections, cloud services) but illustrate that people continue to use manual methods for synchronization and access (e.g., carrying notebook computers, USB drives, emailing to self). Our findings suggest that while methods for access continue to evolve with technology changes, users still make extensive use of familiar – and reliable – patterns of access. Users are adapting and using new cloud technologies both as new ways to achieve existing patterns of access (e.g., manually copying files between devices via Dropbox) and to support new workflows and collaboration.

We observed use of the download-edit-upload (DEU) pattern in several contexts for maintaining temporary, working access to files. This pattern offers users many advantages (e.g., simplicity, control, temporary backup) and is benefitted by newer technologies such as remote desktop and cloud-services that allow it to be used in additional situations. While DEU involves manual synchronization, working on files locally was important for many users.

These results suggest that when new systems are developed, they can benefit users by providing flexibility in how they

are used and allowing users to leverage familiar, manual methods for synchronization and sharing.

Synchronization – Participants used both automated and manual processes to synchronize files among devices and described a number of factors that influenced their choice of method: file size, number of files, speed of Internet connections, the costs of a forgotten or failed sync, and privacy and security concerns. These are consistent with other studies that have reported similar factors (Marshall and Tang, 2012; Dearman and Pierce, 2008) and illustrate the types of decisions that users make regarding synchronization and sharing in daily work. Improvements to technology will continue to help minimize the effects of some of these factors (e.g. file size, number of files). Systems could help users deal with factors such as privacy and security by making it easier to understand and control how data will be synchronized and to which devices.

Boundary management – One of the biggest challenges our participants faced was maintaining and dealing with boundaries imposed by their preferences for work-personal separation and by institutional requirements and policies. These factors played a major role in how many participants organized their file storage, implemented workflows, and architected their synchronization. While some participants were willing to let work spill over into personal spaces, there were concerns about putting personal information on work devices and also about maintaining the security and privacy of protected work information. Prior studies have also found that users have concerns and problems dealing with these synchronization in the context of these boundary issues (Voida, Olson, & Olson, 2013; Marshall & Tang, 2012; Dearman & Pierce, 2008).

There is inherent tension between users' desires for simple, "anywhere" access to their information and personal/institutional needs to segment and secure files. In order to implement and maintain boundaries, complexity is introduced in the form of fragmentation, access control, and configuration, with which users must then grapple when using and managing information. Similar to the issues with synchronization, systems and interfaces are needed that recognize and support boundaries and use that knowledge to create access models that help users in creating appropriate synchronization configurations. Like many other PIM activities, synchronization is often a support activity rather than the user's primary work focus. Users develop differing levels of mental models about how particular synchronization technologies work (Marshall & Tang, 2012). As such, systems also need to help convey appropriate mental models of synchronization and, as Dearman and Pierce (2008) have called for, to show users what will be affected before actions are taken, not after. Systems could allow users to delineate and maintain information boundaries while also providing support for temporary boundary-crossing when necessary.

Collaboration – When collaborating, our participants noted many issues related to synchronization and sharing including: file naming and renaming, multiple versions, tracking changes, and collaborators of varying levels of technical skill. These findings echo those of recent studies that have focused on how group use cloud services for collaboration (Massey, Lenning, & Whittaker, 2014; Volda, Oslon, & Olson, 2013). We found evidence that the preferences and existing expertise of teammates and the size of the group could influence the choice of collaborative tool, mirroring a similar finding by Rader (2009). Tools that support users' synchronization and sharing needs for collaboration are still evolving and gaining adoption. Our findings suggest that users need tools that give them confidence that they will be aware of changes (with options to view things the way they were before changes), that they can undo actions or retrieve data, and that will respect the out-of-system communication that is likely to take place.

CONCLUSION

Through qualitative analysis of the 16 interviews of faculty at a large university, we identified problems and considerations that influence users' file synchronization and sharing practices. Our findings suggest that while methods for access continue to evolve with technology changes, users still make extensive use of familiar patterns of access, including automated and manual methods. We observed a common pattern of download-edit-upload that was used to transport small sets of working files for temporary use on a different device, and then return the edited version to its original location. We reported factors that influence users' choices of methods for synchronization, including privacy and security concerns, data size, institutional policies, and knowledge of systems. We also described insights about how boundaries and data separation affect synchronization practices and about synchronization issues users faced when collaborating. Our findings extend the larger body of research about PIM practices with synchronization and sharing technologies and we provide insights that can be used to develop synchronization and sharing systems that better fit users' needs.

ACKNOWLEDGMENTS

We thank Justin Brinegar for his assistance with this research. This research was funded by a UNC SILS Frederick and Eleanor Kilgour Research Grant.

REFERENCES

Anderson, J.Q. and Rainie, L., 2010. The future of cloud computing. Pew Internet & American Life Project.

- Brinegar, J. and Capra, R. 2011. Managing Music Across Multiple Devices and Computers. In Proc. iConference 2011, ACM, 489-495.
- Brustoloni, J.C., Villamarin-Salomon, R., Djalaliev, P., and Kyle, D., 2008. Evaluating the usability of usage controls in electronic collaboration. In Proc. SOUPS 2008, ACM, 85-92.
- Capra, R., Khanova, J., Ramdeen, S. 2013. Work and personal e-mail use by university employees: PIM practices across domain boundaries. JASIST 64(5): 1029-1004.
- Clark, S.C. 2000. Work/family border theory: A new theory of work/family balance. Human Relations 53(6): 747-770.
- Dearman, D. and Pierce, J.S., 2008. It's on my other computer!: computing with multiple devices. In Proc. CHI 2008, ACM, 767-776.
- Erickson, T. 2006. From PIM to GIM: personal information management in group contexts. Communications of the ACM 49:1, 74-75.
- Jones, W. 2007. Keeping Found Things Found: The Study and Practice of Personal Information Management. San Francisco: Morgan Kaufmann.
- Marshall, C. and Tang, J.C., 2012. That syncing feeling: early user experiences with the cloud. In Proc. DIS 2012, ACM, 544-553.
- Marshall, C., Wobber, T., Ramasubramanian, V., and Terry, D.B., 2012. Supporting research collaboration through bi-level file synchronization. In Proc. GROUP 2012, ACM, 165-174.
- Massey, C., Lenning, T., Whittaker, S. (2014). Cloudy Forecast: An Exploration of the Factors Underlying Shared Repository Use. Proc. CHI 2014.
- Rader, E. 2009. Yours, mine, and (not) ours: social influences on group information repositories. In Proc. CHI 2009, ACM, 2095-2098.
- Tang, J., Brubaker, J., and Marshall, C. 2013. What do you see in the cloud? Understanding the cloud-based user experience through practices. In Proc. Interact 2013, LNCS 8118, Springer, 678-695.
- Volda, A., Olson, J.S., and Olson, G.M., 2013. Turbulence in the clouds: challenges of cloud-based information work. In Proc. CHI 2013, ACM, 2273-2282.
- Volda, S., Edwards, W. K., Newman, M. W., Grinter, R. E., & Ducheneaut, N. (2006). Share and share alike: exploring the user interface affordances of file sharing. In Proc. CHI 2006, ACM, 221-230.