Understanding Researchers' Shared File Organization Practices in Cloud Storage for Collaborative Projects

Kyong Eun Oh (Simmons University, USA)

Introduction

Cloud storage platforms such as Google Drive or Dropbox, which provide ubiquitous and simultaneous access to files, are widely used among researchers for their collaborative projects. However, despite the obvious benefits of cloud storage, there are challenges as well. One is managing shared files that are organized by others with different practices (Bergman, Whittaker, and Frishman 2019; Massey, Lennig, and Whittaker 2014). File management can be supported by systems, especially when their design is informed by an understanding of user behavior and its determinant factors (Dinneen and Julien 2020, E1). However, while the challenges of managing shared files have been discussed, we still don't know much about how shared files are currently organized in cloud storage, when researchers organize them, whether researchers have rules or norms for organizing shared files, how satisfied researchers are with their shared file organization practices, and what variables are related to researchers' satisfaction. This study will address this gap in knowledge by investigating researchers' shared file organization practices in cloud storage for collaborative projects.

Theoretical framework

Personal information management (PIM) is "the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use, and distribute the information needed to complete tasks and fulfill various roles and responsibilities" (Jones 2007, 453). In PIM, there are two models that identified major PIM activities: Jones (2007) who suggested 1) finding/re-finding, 2) keeping, and 3) meta-level activities (464); and Whittaker (2011) who suggested 1) keeping, 2) management, and 3) exploitation (10). In both models, organizing is a crucial part of meta-level activities and management. PIM behaviors have been often studied while focusing on these main activities, and a number of studies have investigated how people organize personal information, especially in digital forms (Henderson and Srinivasan 2011; Jacques et al. 2021; Oh 2017, 2021; Oh and Belkin 2014). Fewer studies examined *shared* file

management in cloud storage; one by Bergman and his colleagues' studies focused on the re-finding of shared files by conducting experiments (Bergman et al. 2015; Bergman, Israeli, and Whittaker 2020; Bergman, Whittaker, and Falk 2014). However, thus far the field lacked a study that specifically investigated researchers' shared file organization practices in cloud storage.

Research questions

RQ1: What are the researchers' primary ways of organizing shared files in cloud storage?

RQ2: When do researchers organize shared files in cloud storage?

RQ3: To what extent does the research project team have rules or norms for organizing shared files in cloud storage?

RQ4: How satisfied are researchers with their shared file organization practices and what variables are related to researchers' satisfaction?

Methodology

Data were collected by distributing an online survey to 360 universities randomly selected from the list of doctoral universities in the United States (Carnegie Classification of Institutions 2018). At each university, administrative representatives were asked to distribute the email with the link to the survey to their department's listservs, inviting any researchers who have an ongoing collaborative research project which uses shared cloud storage to volunteer to take the survey. In total, 535 responses were analyzed by conducting statistical analyses. Table 1 presents participants' information.

Table 1. Demographics of Participants (*N*=535)

		Frequency	Percentage
Gender			
	Male	180	35.2
	Female	313	61.3
	Non-Binary	18	3.5
Age			
	20s	214	41.9
	30s	152	29.7
	40s	77	15.1
	50s	47	9.2
	60+	21	4.1
Ethnicity			
	Asian	84	16.6
	Black/African American	15	3.0

	Latino/Hispanic	33	6.5
	White/Caucasian	351	69.4
	Other	23	4.6
Discipline			
	Arts & Humanities	110	21.5
	Sciences	185	36.1
	Social Sciences	214	41.8
Job			
	Graduate Student	296	57.9
	Postdoc and Other	57	11.2
	Professor	158	30.9

Results

The primary way of organizing shared files

When asked how the files of the project are primarily organized in their shared cloud storage, the top answers were "in folders based on tasks" (23.5%, n=123), "as a list based on file name" (21.4%, n=112), "in folders based on research phases" (17.4%, n=91), "use two or three of the methods in the answer choices almost equally" (13.2%, n=69), and "as a list based on the last updated date" (9.2%, n=48).

When to organize shared files

When asked when they are most likely to organize files in the shared cloud storage, the top answers were organizing files "when the storage looks cluttered/messy" (22.1%, n=115), "when there is a new task" (21.7%, n=113), "when creating a new file" (21.7%, n=113), "when you are done with a certain task "(11.7%, n=61), "not organized in a specific way" (9.4%, n=49), and "when you cannot find the files you need" (8.3%, n=43).

Having explicit rules or implicit norms for organizing shared files

When asked if their research team has agreed-upon rules or norms for organizing files or folder structures for the files in the shared cloud storage, 45.5% said they have implicit norms (n=237), 40.5% said no rules or norms (n=211), and 14.0% said they have explicit rules (n=73). In the case of naming conventions of files, 44.0% said no (n=235), 41.8% said they have implicit norms (n=223), and 12.0% said they have explicit rules (n=64). In the case of controlling different versions of files, 50.7% said no (n=271), 34.3% said they have implicit norms (n=183), and 11.8% said they have explicit rules (n=63).

Satisfaction and variables related to researchers' satisfaction with shared file organization practices

When asked to rate how satisfied they are with their file organization practices in the shared cloud storage on a 7-point Likert scale (1=extremely dissatisfied, 7=extremely satisfied), the mean was 5.37, indicating that participants were slightly or moderately satisfied.

To identify which organization practices are associated with participants' satisfaction, a series of one-way ANOVA were conducted. The results showed that satisfaction was significantly higher for participants who are most likely to organize files when creating a new file (M=5.61, SD=1.54), when they are done with a task (M=5.61, SD=1.45), or when there is a new task (M=5.50, SD=1.62) than those who organized files when they cannot find the files they need (M=4.30, SD=2.00), F(6, 490)=4.24, $p < .01$. In addition, satisfaction was significantly higher for participants whose research team had rules (M=5.76, SD=1.43) or norms (M=5.54, SD=1.56) for organizing files or folder structures than for those who don't (M=5.04, SD=1.77), F(2, 493)=7.45, $p < .01$. In a similar vein, participants' satisfaction was significantly higher for those whose research team had rules (M=5.77, SD=1.40) or norms (M=5.55, SD=1.50) for controlling different versions of files than for those who don't (M=5.15, SD=1.78), F(2, 489)=5.34, $p < .01$. Variables not significantly related to researchers' satisfaction were researchers' primary way of organizing shared files and whether the researchers' project team had rules or norms for naming conventions of files.

Discussion

The findings showed that both folders and lists were heavily used for organizing shared files, unlike previous PIM studies that reported people's strong preference for folders (Jones et al. 2005; Krtalic, Marcetic, and Micunovic 2015). This indicates that for shared file organization, folders are not as prevalent as for personal file organization. In terms of when researchers are most likely to organize files, the most frequent answer was "when the storage looks cluttered/messy" showing that shared files are often organized based on ad hoc needs rather than by following rules or plans. This observation is reinforced by the responses regarding having rules or norms. In all three questions, more than 40% of participants responded that they don't have explicit rules or implicit norms for organizing files, naming conventions, or controlling different versions of files. However, the analysis showed that those whose research team had rules or norms for organizing files/folder structures or controlling different versions of files had significantly higher satisfaction than those who don't, highlighting the importance of having rules or norms. This is further supported by the finding that those who were most likely to organize files when they cannot find needed files had significantly lower satisfaction than those who were most likely to organize files when creating a new file, when they're done with a task, or when there is a new task. Interestingly, having rules or norms for naming conventions of files, which was often recommended as a best practice for

personal file management (Jones et al. 2015; Kearns et al 2014), was not associated with satisfaction.

Conclusion

This study investigated researchers' shared file organization practices in cloud storage for their collaborative projects. The findings examined different practices and revealed possible ways of making shared file organization practice more satisfactory by identifying variables associated with satisfaction. By conducting a large-scale study with researchers in different disciplines, this study deepens our understanding of shared file organization practices for collaborative research projects. It also informs the design of systems and tools that better support researchers' shared file management in cloud storage, which is directly related to researchers' productivity. It would be helpful to further investigate researchers' shared file organization practices, challenges, and strategies via in-depth interviews.

REFERENCES

Bergman, Ofer, Elyada, Oded, Dvir, Noga, Vaitzman, Yana, and Ben Ami, Adir. 2015. "Spotting the latest version of a file with Old'nGray." *Interacting with Computers 27*(6): 630-639.

Bergman, Ofer, Israeli, Tamar, and Whittaker, Steve. 2020. "Factors hindering shared files retrieval." *Aslib Journal of Information Management* 72(1): 130-147.

Bergman, Ofer, Whittaker, Steve, and Falk, Noa. 2014. "Shared files: The retrieval perspective." *Journal of the Association for Information Science and Technology* 65(10): 1949-1963.

Bergman, Ofer, Whittaker, Steve, and Frishman, Yaron. 2019. "Let's get personal: The little nudge that improves documents retrieval in the Cloud." *Journal of Documentation* 75(2): 379-396.

Carnegie Classification of Institutions. 2018. "Standard Listings." https://carnegieclassifications.iu.edu/lookup/standard.php#standard_basic2005_list

Dinneen, Jesse, and Julien, Charles-Antoine. 2020. "The ubiquitous digital file: A review of file management research." *Journal of the Association for Information Science and Technology* 71(1): E1-E32.

Henderson, Sarah, and Srinivasan, Ananth. 2011. "Filing, Piling & Structuring: Strategies for Personal Document Management." In *Proceedings of the 44th Hawaii International Conference on System Sciences*, *San Diego, CA, 2011*, 1-10. IEEE.

Jacques, Jerry, Mas, Sabine, Maurel, Dominique, and Dorey, Jonathan. 2021. "Organizing personal digital information: An analysis of faculty member activities." *Journal of Documentation* 77(2): 401-419.

Jones, William. 2007. "Personal information management." *Annual Review of Information Science and Technology* 41: 453-504.

Jones, William, Capra, Robert, Diekema, Anne, Teevan, Jaime, Perez-Quinonoes, Manuel, Dinneen, Jesse, and Hemminger, Bradley. 2015. "'For Telling' the Present: Using the Delphi Methods to Understand Personal Information Management Practices." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, *Seoul, Republic of Korea, 2015*, 3513-3522. New York, NY: Association for Computing Machinery

Jones, William, Phuwanartnurak, Ammy Jiranida, Gill, Rajdeep, and Bruce, Harry. 2005. "Don't Take My Folders Away! Organizing Personal Information to Get Things Done*."* In *Proceedings of the 2005 Conference on Human Factors in Computing Systems, Portland, OR, 2005,* 1505-1508. New York, NY: ACM.

Kearns, Lorna, Frey, Barbara, Tomer, Chrisinger, and Alman, Susan. 2014. "A study of personal information management strategies for online faculty." *Journal of Asynchronous Learning Networks* 18(1): 19-35.

Krtalić, Maja, Marčetić, Hana, and Mičunović, Milijana. 2016. "Personal digital information archiving among students of social sciences and humanities." *Information Research* 21(2): Paper 716.

Massey, Charlotte, Lennig, Thomas, and Whittaker, Steve. 2014. "Cloudy Forecast: An Exploration of the Factors Underlying Shared Repository Use." In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, *Toronto, Canada, 2014,* 2461-2470. New York, NY: Association for Computing Machinery.

Oh, Kyong Eun. 2017. "Types of personal information categorization: Rigid, fuzzy, and flexible." *Journal of the Association for Information Science and Technology* 68(6): 1491-1504.

Oh, Kyong Eun. 2021. "Social aspects of personal information organization." *Journal of Documentation* 77(2): 558-575.

Oh, Kyong Eun, and Belkin, Nicholas. 2014. *"*Understanding What Personal Information Items Make Categorization Difficult. In *Proceedings of the 77th Association for Information Science and Technology (ASIS&T) Annual Meeting, Seattle, WA, 2014,* 1-3. Association for Information Science & Technology.

Whittaker, Steve. 2011. "Personal information management: From information consumption to curation." *Annual Review of Information Science and Technology* 45: 3-62.