





Journal of eScience Librarianship

putting the pieces together: theory and practice

Why Can't I just use Dropbox? A comparison of cloud file storage platforms used for research

Tobin Magle, Northwestern University, Evanston, IL, USA, tobin.magle@northwestern.edu 
Deb McCafferey, Arizona State University, Tempe, AZ, USA 

Abstract

Objective: Many researchers use cloud file storage platforms such as Box and Google Drive as the sole data management platform for all of their research data throughout the course of their projects. Researchers have lost access to their preferred platforms due to changes in licensing agreements and cost to their institutions, leaving researchers to figure out a new system on their own. This paper describes differences between these platforms that affect research data workflows.

Methods: We selected four commonly used cloud file storage vendors (Box, Dropbox, Google Drive, and Microsoft SharePoint/OneDrive) to assess. The authors read public user documentation for the platforms and identified several differences, such as maximum file size, that could affect research data workflows. For each difference, we recorded the specifics of each platform then narrowed the scope to features that vary across the platforms.

Results: We identified three areas where cloud platforms differed in ways that affect data management: data stewardship, storage capacity, and file organization. Data stewardship is affected by variations in the platforms' approaches to individual vs. group ownership of files and how user roles are defined and assigned. The platforms also differ in limits on total capacity, individual file size, total number of files, and the amount of data that can be moved in and out of the platform. Some of these limits vary by access method (e.g. API vs web interfaces). Finally many differences affect how data can be organized, such as number of files per folder, synchronization, and file sharing restrictions.

Conclusions: Some research data workflows may be negatively affected by moving between cloud file storage platforms. Researchers will need assistance identifying differences between the platforms and modifying disrupted data workflows.

Received: July 18, 2023 **Accepted:** November 20, 2023 **Published:** December 20, 2023

Data Availability: All data collected for this study is contained within manuscript tables.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

See <https://creativecommons.org/licenses/by/4.0>.

♾️ OPEN ACCESS

Introduction

Researchers at academic institutions collect, analyze, share, publish, and archive data throughout the research data lifecycle, often within the same storage platform. Many institutions once offered unlimited cloud file storage with all the features necessary for the entire lifecycle at no cost to researchers. However, restricted storage quotas (Weers 2021) and increased prices (Krazit, n.d.) caused institutions to seek more cost-effective solutions and migrate university data including research data to other storage platforms, despite the enormous amount of work required by such migrations. While these platforms might seem interchangeable on the surface, subtle differences can cause some researchers to have to change their data workflows. This paper will discuss differences among these platforms that can affect research data management.

Methods and Results

We selected four common cloud file storage vendors to evaluate: Box (Box, n.d.), Google Drive (Google, n.d.), Dropbox (Dropbox, n.d.) and Microsoft SharePoint/OneDrive (Microsoft 2023b). We chose several criteria (see Supplemental Material) and recorded information by reviewing public user documentation provided by the vendors. We then identified factors that (1) cannot be negotiated with the vendor and (2) vary between platforms. We divided the types of limitations into three groups based on their impact on data management: data stewardship, capacity, and file organization. Each group is discussed in turn below.

Data stewardship

Research data is governed by policies that vary by funding agreements (United States: National Archives and Records Administration: Office of the Federal Register 2013; Blum 2012), copyright law (OpenAIRE, n.d.; Kuschel and Dolling 2022), and institutional policies (Government of Canada 2021). While the details vary, the goal of these policies is to keep the data secure and usable into the future. The way cloud file storage platforms handle file ownership and access affects the risk of data loss and exposure.

In assessing the platforms, we divided them into two categories: individual and group storage. Files within individual storage are owned by an individual user and can be shared with others. However, when that user's account is deactivated, the files are removed from the platform, even when shared with other research group members. All platforms we evaluated have individual storage associated with each user account (Table 1). Using individual storage introduces the risk of data loss at worst and creates another task when offboarding employees at best. In contrast, files within group storage aren't tied to an individual user: they are associated with the enterprise license. Individual users can be granted access to files in group storage. Notably, the Box platform does not have group storage, though some universities have used institutional logins that are not tied to an individual to simulate this feature.

Table 1: Cloud file storage platforms by types of storage

Vendor	Platforms	
	Individual storage	Group storage
Box	Box	None
Dropbox	Dropbox	Dropbox Team
Google	Google Drive	Shared Drive
Microsoft	OneDrive	SharePoint

Access to files in either type of storage can be extended to other users. The roles available vary greatly by name and function (Table 2). All four individual storage platforms have versions of owner, editor, and viewer roles (Box 2020e; Dropbox, n.d.; Google, n.d.; n.d.; Microsoft 2021; n.d.). Owners have full rights to the content: they can create, add, view, delete, copy, move, and share files and folders they own. Editors can generally edit, share, move, and delete files. Viewers can view the files on all platforms and can also make comments on files in Dropbox and Box. Google Drive and OneDrive have separate roles for commenting (commenter or “can review”). Group storage has roles similar to editors (member, contributor, etc.) and restricts these roles from sharing. Separate roles—like admin, owner or manager—control access. Additionally, Google Shared Drive contributors cannot delete content from the drive, but “content manager” can delete files and may be able to share content at the discretion of the manager. The variation in the names of and permissions associated with these roles could cause confusion when moving between platforms and could result in researchers sharing data in an unintended way.

Capacity

These platforms also differ in the volume of data they can accommodate. The total capacity, how capacity is distributed, and what counts against a user’s quota vary among the platforms (Table 3). Box and Dropbox continue to offer unlimited storage at increased prices that vary by storage capacity (Box, n.d.; n.d., Dropbox, n.d.). Google offers a set amount of storage for an entire institution that can be distributed by storage administrators through Google Workspace (Google, n.d.). For example, Google Workspace for Education is offered to qualifying institutions at no cost and includes 100 TB of pooled storage for all users with the option to purchase more storage for the pool (Google, n.d.). Microsoft distributes storage to users uniformly (1 TB for individual and max 25 TB for group storage) (Microsoft, n.d.; 2023b). Storage administrators can limit how many SharePoint sites are created and what their quotas are (Microsoft 2023a; 2023c). Files owned by a user in individual storage count against the quota, including previous versions and items pending deletion depending on the platform. Additionally, platforms that are integrated with email and messaging applications (Microsoft and Google products) include file attachments against quotas for account’s individual

Table 2: User roles and permissions by each platform

Vendor	Individual storage					Group storage				
	Role	Share	Edit, move, delete	Comment	View	Role	Share	Edit, move, delete	Comment	View
Box	owner	x	x	x	x					
	co-owner	x	x [§]	x	x					
	editor	x	x [§]	x	x					
	viewer			x	x					
Dropbox						admin	x			
	owner	x	x	x	x	owner		x	x	x
	editor	x	x	x	x	editor		x [†]	x	x
	viewer			x	x	viewer			x	x
Google	owner	x	x	x	x	manager	x	x	x	x
						content manager	x [†]	x	x	x
	editor	x	x	x	x	contributor		x ⁼	x	x
	commenter			x	x	commenter viewer			x	x
Microsoft	owner	x	x	x	x	owner	x	x	x	x
	can edit	x	x	x	x	member		x	x	x
	can review			x	x					
	can view				x	visitor				x

* Box also includes unique roles, uploader and previewer, and combination roles "viewer uploader" and "previewer uploader".

§ cannot delete top-level folders

† only admins can move folders out of a team folder

‡ optional, managers can restrict this option

= contributors cannot move or delete files

storage in OneDrive or Google Drive (Box 2022; Dropbox, n.d.; Google, n.d.; Microsoft, n.d.). Notably, Dropbox also counts files that users outside a user's enterprise subscription share with them against their quota, counting the same file against multiple users' quotas (Dropbox, n.d.). Thus, the amount of storage a given dataset will take up could vary between the platforms.

These platforms also limit the maximum file size, number of files, and data movement in and out of the platform (Table 3). Individual file size limits range from 150 GB to 5 TB (Box 2020d; Dropbox, n.d.; Google, n.d.; Microsoft, n.d.). Further limits are applied to different access methods on some platforms. For example, Box limits files uploaded through their mobile app and sync client to 15 GB and by email to 50 MB, and Dropbox limits files uploaded through the API to 350 GB and web interface to 50 GB (Box, n.d., Dropbox n.d). Limits on the total number of files allowed differ by orders of magnitude, with Dropbox having no stated limit (Table 3) (Box 2020b; Microsoft 2023b). All platforms limit data movement in and out of the platform, either by number of files or amount of data uploaded or downloaded in a defined time period (Box 2020c; Dropbox, n.d.; Google, n.d.; Microsoft, n.d.). The implications of these limits will be covered in the discussion section.

Table 3: Capacity limits by vendor

Vendor	Total	File size	File number	Data movement
Box	Unlimited	150 GB	1,000,000 per user	1 TB/user/month download
Dropbox	As much space as needed	2 TB	None	4 TB/ day download
Google	100 TB per institution	5 TB	400,000 per shared drive	750 GB/day upload
Microsoft	5 TB per user, 25 TB per SharePoint library	250 GB	30,000,000 total per user	250 GB or 10,000 files per day download

File Organization

Limitations that affect file organization vary widely among the platforms (Table 4). Box and Google Drive limit the number of files users can have in a folder (Box 2020b; Google, n.d.). Google also limits the number of files in Shared Drives (Google, n.d.). Furthermore, platforms limit the number of files or folders that can be shared and synced, some by number of items, other by the depth of the folder hierarchy, the total number of shared folders or the number of users or groups they are shared with (Microsoft, n.d., Dropbox, n.d., Google, n.d.). These limitations force a choice between reorganizing files or effort sharing many smaller folders individually. Finally, these platforms also experience performance issues when syncing many files

using desktop applications. Box reports service degradation when syncing >100,000 files, while Dropbox and Microsoft suggest syncing less than 300,000 files (Box 2020a; Dropbox, n.d.; Microsoft, n.d.).

Table 4: Limitations that affect file organization

Platform	Files per folder	Sharing	Sync
Box	15,000	Unlimited	<100,000 files
Dropbox	Unlimited	30,000 folders 1,500 subfolders	<300,000 files
Google Drive	500,000	100 groups 600 members	Unlimited
Microsoft OneDrive/ Sharepoint	Unlimited	50,000 files	<300,000 files

Discussion and Conclusion

While cloud file storage platforms seem interchangeable on the surface, this article identifies several differences that affect the risk of data loss, sharing and collaboration, the size of the files that can be stored, and how files can be organized. Some limitations, like total capacity, are flexible because institutions can buy more storage, as cost prohibitive as it may be. Though the increased cost of these platforms has been disruptive, this change makes sense: storage with more capacity and features tends to cost more to operate than storage without these features. This tradeoff is similar to high speed data storage attached to high performance computing clusters being more expensive than storage on other research data storage clusters. However, researchers are largely still insulated from the economics of data storage because their institutions offer cloud file storage at no cost, but charge for dedicated research storage. Changing the relative cost of storage platforms offered to reflect the actual cost to the university could nudge researchers to use cloud file storage for their intended purpose (sharing and document collaboration) and find more cost-effective solutions for storing and archiving large datasets.

Other differences are integral to the platforms and cannot be negotiated. These hard limitations affect particular research projects differently:

- Individual file size limits can prevent researchers from storing large files like videos or 3D images.
- File number limitations affect projects that generate many small files like outputs from computer models that vary multiple parameters.

- Sync limits also affect researchers who work with many small files, exacerbated by unstable internet connections.
- API upload and data movement limits restrict automation of high-throughput data workflows.

When encountering these workflow disruptions, researchers need help identifying the problem and reworking their data workflows to accommodate.

Finally, other differences require researchers to shift their mindset about data storage and organization. Many researchers are accustomed to organizing their data in individual storage based on a system that works for them and sharing specific files and folders with collaborators. Using group storage space mitigates the chance of data loss. However, because group storage grants access to a group of people by default, researchers need to think about an organization strategy that works for everyone and only store things that everyone should have access to in that space. For example, creating group storage for an entire research group seems like a good strategy, but won't work for projects that contain data subject to data use agreements that require only certain named individuals have access. A project based organization approach will work better in this case. Whether using group or individual storage, differences in limits to what a user can share also affects data organization. For example, a migration from Box to Dropbox means a shift from unlimited sharing to a limit of 30,000 shared folders and a limit of 1,500 subfolders in a folder. For a lab that runs cryo-EM experiments generating many small files and folders (Baldwin et al. 2018), this could mean that data has to be reorganized before migration.

In short, these platforms have differences that affect research data management. These differences became apparent when universities moved between cloud file storage platforms. As research data and computing support professionals, we can help researchers adapt to this change. Additionally, as these platforms become less useful as data management platforms, our profession can help researchers and our institutions think more holistically about researchers' data storage needs and how to incentivize good data management practices.

Data Availability

All data collected for this study is contained within manuscript tables.

Acknowledgements

We would like to thank Brian Cors from the University of Michigan for conversations and input on enterprise licenses for these platforms from an enterprise IT perspective.

The content of this article is based on the poster entitled "Why can't I just use Dropbox?" originally presented at [RDAP Summit 2023](#), available from Open Science Framework: <https://osf.io/yxmuj>.

Competing Interests

The authors declare that they have no competing interests.

References

- Baldwin, Philip R., Yong Zi Tan, Edward T. Eng, William J. Rice, Alex J. Noble, Carl J. Negro, Michael A. Cianfrocco, Clinton S. Potter, and Bridget Carragher. 2018. "Big Data in cryoEM: Automated Collection, Processing and Accessibility of EM Data." *Current Opinion in Microbiology* 43(June): 1–8. <https://doi.org/10.1016/j.mib.2017.10.005>.
- Blum, Carol. 2012. "Access to, Sharing and Retention of Research Data: Rights & Responsibilities." Council On Governmental Relations. https://www.cogr.edu/sites/default/files/access_to_sharing_and_retention_of_research_data_-_rights_%26_responsibilities.pdf.
- Box. 2020a. "Maximizing Box Sync Performance." Box Support. February 26, 2020. <https://support.box.com/hc/en-us/articles/360044194753-Maximizing-Box-Sync-Performance>.
- . 2020b. "Preparing Your Content for Migration." Box Support. February 26, 2020. <https://support.box.com/hc/en-us/articles/360043693654-Preparing-your-content-for-migration>.
- . 2020c. "Understand How Box Measures Bandwidth Usage." Box Support. February 26, 2020. <https://support.box.com/hc/en-us/articles/360044194313-Understand-How-Box-Measures-Bandwidth-Usage>.
- . 2020d. "Understand the Maximum File Size You Can Upload to Box." Box Support. February 26, 2020. <https://support.box.com/hc/en-us/articles/360043697314-Understand-the-Maximum-File-Size-You-Can-Upload-to-Box>.
- . 2020e. "Understanding Collaborator Permission Levels." Box Support. February 26, 2020. <https://support.box.com/hc/en-us/articles/360044196413-Understanding-Collaborator-Permission-Levels>.
- . 2022. "Storage Space Consumed by Shared Folder." Box Support. June 7, 2022. <http://support.box.com/hc/en-us/community/posts/6786566582291-storage-space-consumed-by-shared-folder>.
- . n.d. "Box — Secure Cloud Content Management, Workflow, and Collaboration." Accessed July 14, 2023a. <https://www.box.com/home>.
- . n.d. "Choose the Best Plan for Your Business: Business Plans." Accessed July 17, 2023b. <https://www.box.com/pricing/business>.
- . n.d. "Choose the Best Plan for Your Business: Individuals and Teams." Accessed July 17, 2023c. <https://www.box.com/pricing/individual>.
- Dropbox. n.d. "Compare All Dropbox Plans - Dropbox." Dropbox. Accessed July 17, 2023a. <https://www.dropbox.com/plans>.
- . n.d. "File Transfer." Dropbox. Accessed July 17, 2023b. <https://www.dropbox.com/features/share/send-large-files>.
- . n.d. "How to Manage Your Dropbox Sharing Permissions." Accessed July 17, 2023c. <https://help.dropbox.com/share/set-file-folder-permissions>.

———. n.d. "Restrictions and Limitations for Team Deployments of Dropbox." Accessed July 17, 2023d. <https://help.dropbox.com/plans/large-deployments>.

———. n.d. "Secure Team Collaboration - Dropbox Business." Dropbox. Accessed July 14, 2023e. https://www.dropbox.com/business/?_tk=paid_sem_goog_biz_b&_camp=142947702&_kw=dropbox%7Ce&_ad=649958535546%7C%7Cc&gclid=CjwKCAjw5MOIBhBTEiwAAJ8e1jdhFrN3jKgiCazBSOTRhS_UGvLAUfLYHEScFSACcKExM7ej7HRUxoCnzcQAvD_BwE&gclid=aw.ds.

———. n.d. "What Is the Dropbox File Size Limit?" Accessed July 17, 2023f. <https://help.dropbox.com/sync/upload-limitations>.

———. n.d. "Why Can't I Create a Shared Folder?" Accessed July 17, 2023g. <https://help.dropbox.com/share/shared-folder-faq#cantcreate>.

———. n.d. "Will Joining a Shared Folder Count against My Storage Space?" Accessed July 17, 2023h. <https://help.dropbox.com/storage-space/shared-folder-count-against-storage>.

Google. n.d. "How File Access Works in Shared Drives - Google Workspace Learning Center." Accessed July 17, 2023a. <https://support.google.com/a/users/answer/12380484>.

———. n.d. "How Many Items Can I Have Directly in a Folder?" Accessed July 17, 2023b. <https://support.google.com/a/answer/2490100?hl=en#zippy=%2Cchow-many-items-can-i-have-directly-in-a-folder>.

———. n.d. "Overview of Google Workspace for Education Storage - Google Workspace Admin Help." Accessed July 17, 2023c. <https://support.google.com/a/answer/10403871?hl=en#zippy=%2Cwhat-does-pooled-storage-mean-what-counts-toward-pooled-storage>.

———. n.d. "Personal Cloud Storage & File Sharing Platform - Google." Accessed July 14, 2023d. <https://www.google.com/drive>.

———. n.d. "Share Files from Google Drive - Computer - Google Drive Help." Accessed July 17, 2023e. <https://support.google.com/drive/answer/2494822?hl=en&co=GENIE.Platform%3DDesktop#zippy=%2Cshare-with-specific-people>.

———. n.d. "Shared Drive Limits in Google Drive - Google Workspace Learning Center." Accessed July 17, 2023f. https://support.google.com/a/users/answer/7338880?hl=en#shared_drives_file_folder_limits.

———. n.d. "Storage and Upload Limits for Google Workspace - Google Workspace Admin Help." Accessed July 17, 2023g. <https://support.google.com/a/answer/172541>.

Government of Canada, Innovation. 2021. "Tri-Agency Statement of Principles on Digital Data Management." Innovation, Science and Economic Development Canada. January 20, 2021. <https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-statement-principles-digital-data-management>.

Krazit, Tom. n.d. "Would You Pay 1,000% More for Box? - Protocol." Accessed April 21, 2023. <https://www.protocol.com/newsletters/protocol-enterprise/would-you-pay-1000-more-for-box>.

Kuschel, Linda, and Jasmin Dolling. 2022. "Access to Research Data and EU Copyright Law." *JIPITEC* 13(3). <https://www.jipitec.eu/issues/jipitec-13-3-2022/5558>.

- Microsoft. 2021. "Managing External Guests in SharePoint vs Teams." December 16, 2021. <https://learn.microsoft.com/en-us/microsoft-365/community/managing-external-guest-in-sharepoint-vs-teams>.
- . 2023a. "Manage Site Creation in SharePoint - SharePoint in Microsoft 365." April 6, 2023. <https://learn.microsoft.com/en-us/sharepoint/manage-site-creation>.
- . 2023b. "SharePoint Limits - Service Descriptions." April 25, 2023. <https://learn.microsoft.com/en-us/office365/servicedescriptions/sharepoint-online-service-description/sharepoint-online-limits>.
- . 2023c. "Manage Site Storage Limits in SharePoint in Microsoft 365 - SharePoint in Microsoft 365." June 21, 2023. <https://learn.microsoft.com/en-us/sharepoint/manage-site-collection-storage-limits>.
- . 2023d. "Microsoft 365 Education - Service Descriptions." June 28, 2023. <https://learn.microsoft.com/en-us/office365/servicedescriptions/office-365-platform-service-description/microsoft-365-education>.
- . n.d. "Changes to Microsoft 365 Email Features and Storage - Microsoft Support." Accessed July 17, 2023a. <https://support.microsoft.com/en-us/office/changes-to-microsoft-365-email-features-and-storage-e888d746-61e5-49e3-9bd1-94b88e9be988>.
- . n.d. "Cloud Storage Pricing and Plans." Accessed July 17, 2023b. <https://www.microsoft.com/en-us/microsoft-365/onedrive/compare-onedrive-plans>.
- . n.d. "Restrictions and Limitations in OneDrive and SharePoint - Microsoft Support." Accessed July 17, 2023c. <https://support.microsoft.com/en-us/office/restrictions-and-limitations-in-onedrive-and-sharepoint-64883a5d-228e-48f5-b3d2-eb39e07630fa>.
- . n.d. "Understand Groups and Permissions on a SharePoint Site - Microsoft Support." Accessed July 17, 2023d. <https://support.microsoft.com/en-us/office/understand-groups-and-permissions-on-a-sharepoint-site-258e5f33-1b5a-4766-a503-d86655cf950d>.
- OpenAIRE. n.d. "How Do I Know If My Research Data Is Protected?" OpenAIRE. Accessed September 25, 2023. <https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>.
- United States: National Archives and Records Administration: Office of the Federal Register. 2013. "Uniform Administrative Requirements For Grants And Agreements With Institutions Of Higher Education, Hospitals, And Other Non-Profit Organizations (OMB Circular A-110)." In *Grants and Agreements. Title 2*. Office of the Federal Register, National Archives and Records Administration. <https://www.govinfo.gov/app/details/CFR-2013-title2-vol1/CFR-2013-title2-vol1-part215>.
- Weers, Catherine. 2021. "Google Workspace Storage Limits Are Coming." *Amplified IT* (blog). February 17, 2021. <https://www.amplifiedit.com/workspace-storage-limits>.