Digital Preservation in the Performance Library

Introduction

You may not realize as you go about your daily work that your digital files are valuable assets. They took time to create, and you don't want to do that work again. You stood at the scanner and then carefully edited those parts in Photoshop, or you spent days researching next season's repertoire. What should you do to make sure your files are safe and therefore accessible when you need them?

Unfortunately digital preservation is more challenging and complex than analog preservation, i.e. putting your printed music on the shelf. However, do not delay safeguarding your files. Doing something, even just one or two of these practices, is better than doing nothing.

Disclaimer: This document is primarily intended to introduce digital preservation concepts. Any specific utilities, tools, software, etc. included herein may be obsolete within a relatively short time after 2022.

Risk Management

We think about risks so we can plan to avert them. Prioritizing which situations are more likely to happen to you will help you know where to start mitigating those risks, and will give you a clearer path forward for preventing problems. Here are some examples:

- File format obsolescence. What files are you no longer able to open because you no longer have the software?
- Media degradation, e.g. CDs that have delaminated.
- Someone could accidentally delete your files.
- Big changes in your organization, including leadership changes, can put your digital assets at risk. For instance, a new administrator's vision for budgeting could negatively affect IT support.
- Technology refreshes - getting a new computer, or having new network hardware installed, could cause loss or damage.
- A single point of failure, such as the lone IT person on staff being the only one who understands your systems.
- Outsourcing storage: If there's a move to the cloud, be a part of those discussions to make sure your files will be safe, accessible, and secure.
- Natural disasters pose risks to digital files.

One of the best bits of advice is to keep things simple - do not create or rely on custom solutions that depend on a single person.

Know What You Have *(see Worksheet)*

To begin properly preserving your digital files, you need to know what you're preserving.

- How many files do you have, and how much room do they take up in terms of gigabytes or terabytes?
- You may also want to determine your average rate of growth, such as how many files per year and how many gigs per year your holdings are growing. You can determine this by checking in periodically on your numbers of files and their size. This will give you a sense of future storage needs, especially if you're managing your own files and backups. Or, you'll be able to more expertly advocate for increased resources.
- You should also know what types of files you have. Some of the more common types include PDFs, TIFFs, JPGs, .doc, and .xls.
- How are your files stored? Are they just local on your desktop computer? Or are they on a laptop, network storage, in the cloud in a personal or a corporate account, on an external hard drive, flash drives, diskettes, or other media?
- Who is responsible for managing your digital storage? Is it just you, the one IT person at work, or an outsourced IT service?
- Who has access to your files? Consider if that's just you, only certain staff, or way too many people.

Protect

A. Virus and malware protection

Make sure your virus and malware protection is up to date and working, especially if you are managing your own storage environment. Check it periodically, run the scans, and keep the software up to date. Practice good habits such as not plugging random thumb drives into your primary storage device and not clicking on links in emails from people you don't know. If you are not managing your own environment, it's a great chance to engage your IT folks in a conversation.

B. File structure

Being organized is a great way to protect your files. Good file structure provides necessary context. A flatter hierarchy (fewer levels) is easier to preserve than a deep one; e.g. backup utilities may not be able to preserve files or structural context beyond 5 levels deep.

When deciding on folder structure, there is no right answer since every set of files and their uses are different. But there are certain considerations that can help keep the digital files organized. Would it make sense to start with folders arranged alphabetically by composer and title? Or if by seasons and projects, perhaps a chronological arrangement is best. Organize everything so you'll be able to find it again quickly. Or, if someone else needed to find a particular file, consider how hey would most easily discover it.

C. File naming conventions

Correct file naming can help maintain the lifespan of digital files by ensuring they remain accessible across multiple operating systems. Though you may work on a Mac, your backup utility might be Linux-based. File names can and should also provide valuable metadata that helps everyone identify the files quickly and easily.

When it comes to naming your files, again consider how that impacts their organization. If you list individual scans by instrument name, they'll sort alphabetically. If you'd like them in score order, use a number system.

- Use only using letters, numbers, and underscores, and not spaces. Some operating systems may not recognize spaces.
- To separate words or dates use underscores, or "CamelCase," which is capitalizing the first letter of every new word.
- When organizing files chronologically, start with the date in year-month-day format. YYYY-MM-DD
- If you're imposing a file order, such a score order, use leading zeros before the instrument name. Score could be 001, flute 1 could be 002, etc.
- The only period in your file name should be the one your computer likely adds automatically right before the file extension.
- If you're working with multiple versions of files, including a version number at the beginning of the file names helps keep thing organized and readily recognizable. You may want to mark the final version "FINAL" at the beginning of the file name.

But whatever you do, BE CONSISTENT across all of your folders. Consistency helps with organization and quick recognition of files.

If you need to rename a lot of files, utilities such as Bulk Rename and Name Mangler can help automate the process.

D. Backups

LOCKSS is an easy-to-remember acronym for "Lots of copies keeps stuff safe." Making more than one copy, and utilizing more than one type of storage, mitigates a variety of digital preservation risks.

Keep at least one copy of your files easily accessible on your hard drive or in your organization' shared storage. You may want another copy in the cloud.

Automate your backup process with software or an online utility, such as BackBlaze. Automated processes help remove barriers to completing this important task by taking care of the job for you. Just remember to check that your backups do run and that they are capturing everything. This is where you find out if your file structure and naming conventions are either helping or hindering you. You may need to make adjustments such as flattening your hierarchy and renaming files so the backups can run smoothly and fully.

Make at least one additional copy on a less accessible storage medium. This may be the copy you keep in a different geographic location to protect against disaster. You may want to use an external hard drive for this. You could even partner with another orchestra where you trade backups, agreeing to keep them safe for each other.

If running backups are not up to you, ask your IT person how often/when is the network file storage backed up. Do they do overlapping backups? Once a day, once a week, once a month, and don't overwrite them immediately? This allows you to restore files from a backup weeks or months in the past, in case you only noticed recently that a file or folder was accidentally deleted.

E. Access/security

Moderating access to your files helps protect against accidental deletion, modification, and distribution, especially of copyrighted material. Can anyone and everyone in your organization access your files? If so, I recommend working with your IT to change that and set user permissions. For very sensitive information, you may consider locking files with passwords, or setting them up to be read-only for most people.

Copyright makes a distinction between ownership of the physical manifestation of a work, and the separate right to reproduce it (the right to copy). Digital material by nature does not align with this distinction. In the case of digital material, practices such as the preservation activity of making backups involve the deliberate or inadvertent creation of copies. The archival community is quite concerned about this issue.

However, when I worked at the US Marine Band, my biggest concern was not backup copies but rather the risk of illegal dissemination. I felt relatively comfortable having the files and copying them for backups, but we would never send the files to another organization unless authorized by the copyright holder to do so.

Thus copyright presents a situation where security of your files is paramount. While you would never illegally disseminate copyrighted digital music files, the marketing department intern who stumbles into your files might not know better.

F. Storage media obsolescence

Hardware and storage media obsolescence puts files at risk. Have you ever found a 5.25" disk or 3.5" diskette and had no way to view the files it contained? Or realized that your new laptop has no CD or DVD drive? Knowing what you have and how it is stored helps you prevent these data loss situations. You'll know to migrate files off of those digital carriers before you lose the ability to do so. Usually computer refreshes are announced before IT shows up to take your old machine away and install the new one. Theoretically you should have a window of time to migrate your files. If you have diskettes or optical media that you're no longer able to read, there are vendors such as EverPresent who can access the files and send them to you.

G. Refresh hardware

The typical lifecycle of computer hardware is 3-5 years. Estimate on the lower side for magnetic media and on the higher side for solid state media. Refresh hard drives, external drives, network drives, and so forth every 3-5 years. Do not wait for a drive to fail. Even if it does not fail, it will degrade over time and put your files at risk for corruption (see the section on "bit rot").

H. Preservation file formats

Some file formats last longer than others. Use formats recommended for preservation to best ensure the usability of your files over the long term, i.e. to prevent loss due to obsolescence. The Library of Congress is the leading resource for these recommendations:
https://www.loc.gov/preservation/resources/rfs/TOC.html

For preserving music, LOC most highly recommends paper. Next is music XML, then PDF/UA, and PDF/A. Native Finale, Sibelius, and Dorico files are not recommended for preservation. For image files, the recommended formats are .tiffs, then .jpg2000, then .png files.

I. Migration

If you run into file obsolescence, or want to try to prevent it, you would need to migrate your files to a newer or different formats that are still supported. There are digital tools for migrating files, which can be found online. Migration does carry a risk of damaging files.

If you have some old files that you can't open and don't recognize the file extension, you can use the Pronom database to identify them. Once you know what they are, you have a better chance of migrating them to a file format you can hopefully open. https://www.nationalarchives.gov.uk/PRONOM/

J. Resolution

The resolution you choose when scanning should be based on what you need. 600 dpi is usually appropriate for music going on the stand. Choosing greyscale, color, or black and white depends on your bandwidth, storage space, uses, and requirements. Higher resolutions mean crisper notes and staff lines for players, but also means larger files that take longer to move around, could be hard to share due email limits, and require more storage space.

Color and grayscale files are larger, so may need to be scanned at lower resolutions, without loss of clarity. Black and white scans can be done at higher regulations, since colors and gradations are not being captured. Generally speaking, scan at the highest setting you can manage, and scan to a format you can manipulate.

Monitor and Repair

Few of you may get into actively monitoring and repairing your files if something goes awry. While this is a bit more technical and typically done by digital preservation specialists and IT personnel, it's still good to understand what's going on.

A. Bit rot

Digital files deteriorate. This corruption is called "bit rot." Files are essentially made up of magnetic 1's and 0's. These are our fundamental bits, eight of which make up a byte. Those bits can flip, causing our files to become corrupted. Factors that can contribute to this include aging hardware, migration to new formats, and file transfers. *If* you are able to open corrupt files, either data, formatting, or both may be lost.

B. File fixity

Bit rot can be detected by checking file fixity. IT personnel should be checking this to make sure files are transferring, backing up, and being preserved in storage without corruption. If fixity and checksums are something you're interested in learning more about, there is a tool available for less than $50 per year called Fixity Pro that can help you check file fixity. If you do discover corrupt files, you'll need to be able to replace them from good copies, either from original files or backups run prior to the onset of bit rot.

Sustain

How do you sustain your files and the great digital preservation practices you've undertaken?

A. Document

Create documentation, if you haven't already. Filling out the worksheet is a great place to start. What sorts of things should you document? A lot of the points covered at the beginning, including where your files are located, security settings, backup protocols, access restrictions, file migrations, who is responsible for what, policies and procedures, the file structure, your file naming conventions, workflows, any digital utilities tools you use, and so forth.

a. Succession planning

When you leave, will your successor know where to find everything they need? They shouldn‘t have to waste time figuring out everything from scratch. It's always better to be able to hit the ground running.

b. Consistency across staff

Up-to-date documentation also very helpful for consistency. If there is more than one person in your library, documentation helps keep you all on the same page, helps establish procedures and policy, and can answer their questions when you're too busy relaxing on the beach and shouldn't be answering work emails anyway.

B. Review

Review your documentation at least annually. Put a tickler on your calendar to do this, and update your documentation every time anything changes.

Review your practices annually as well to make sure they're still working for you, your staff, your players, and your organization.

- Are your file formats, structure, and naming conventions still working?
- Is your backup strategy on track and serving you well?
- Are you achieving geographically distant backups?
- Are you running your backups on schedule, and if not, what barriers need removing to ensure they're occurring regularly?
- Have you monitored your data recently?
- Is it time to refresh hardware? Documenting when you purchased and installed new hardware and software is really helpful to know when equipment and tools are reaching end of life.
- Is there enough redundancy in place that you're avoiding single points of failure? Have single points of failure crept in because someone left, or did a piece of equipment fail and hasn't yet been replaced?

C. COOP

Digital music files should be considered in Continuity of Operations planning. This is where your offsite copies will become critical, as well as your documentation. Make sure your backups are geographically separate enough that they wouldn't be affected by the same disaster that befalls your organization. Put a reminder in your calendar to keep this backup refreshed as much as necessary. If your digital holdings are not represented in your organization's emergency preparedness plan, you'll want to fix that.

D. Advocacy

You and the contents of your library are valuable resources critical to your organization's success. The library's digital assets are an important part of that resource that should be prioritized, protected, and budgeted for as much as the physical music. Digital preservation should be integrated into your organization's culture.

Your data is growing, and the organization needs to budget for that. Storage may be cheap, but it, along with management, has associated costs. In digital preservation terms we're talking about advocacy for your library and its digital assets, and organizational policies that recognize and protect digital assets.

Advocacy includes making others in your organization aware of how the library's digital assets are beneficial. Did they save the rehearsal or the performance? Tell that story. Did they save someone time? Tell that as well. When appropriate, explain the investment necessary (time, money, personnel, other resources) to keep those amazing digital assets viable, safe, and accessible so they can continue to be so beneficial.

Advocacy is constant and ongoing, and part of that is making sure you are included in discussions about your files, such as IT meetings and budget meetings. Being the expert about your library's digital assets helps with your advocacy, and you become that expert by knowing your files, how they're stored, how they're managed, and how they've saved the day.

Further Resources

Reading:

● NEDCC leaflet:

https://www.nedcc.org/free-resources/preservation-leaflets/6.-reformatting/6.5-digital-preservation

● Library of Congress recommended formats: https://www.loc.gov/preservation/resources/rfs/

● File naming: http://www.controlledvocabulary.com/imagedatabases/filename_limits.html

● Digital Preservation Coalition: https://www.dpconline.org/

Video tutorials:

● Digital Preservation for Individuals and Small Groups:

http://www.ala.org/alcts/confevents/upcoming/webinar/043015

- Preserving Digital Collections: An Overview: http://www.ala.org/alcts/confevents/upcoming/webinar/031616

- Long-Term Storage in Digital Collections: http://www.ala.org/alcts/confevents/upcoming/webinar/pres/111412

- Field Guide to Digital Preservation: http://www.ala.org/alcts/confevents/upcoming/webinar/101619

- Audiovisual Digital Preservation: http://www.ala.org/alcts/confevents/upcoming/webinar/022719 and

- http://www.ala.org/alcts/confevents/upcoming/webinar/031319

- Digital Preservation for Small Repositories:

https://www.youtube.com/watch?v=uVb4IXQgTeI&feature=youtu.be

- Managing your Digital Collection: https://www.youtube.com/watch?v=PHD92p2cUjo&feature=youtu.be