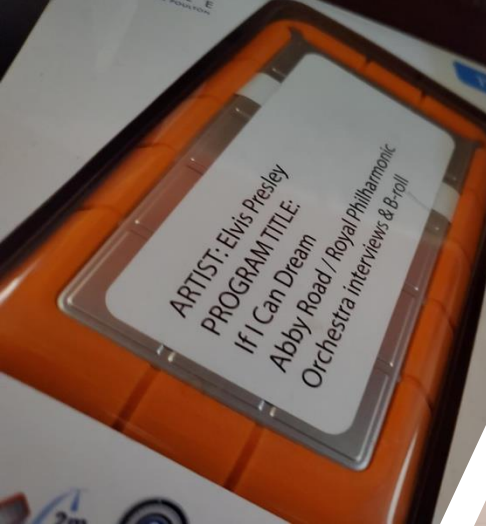
Mass Digitization: Media Hardware → Digital Archives



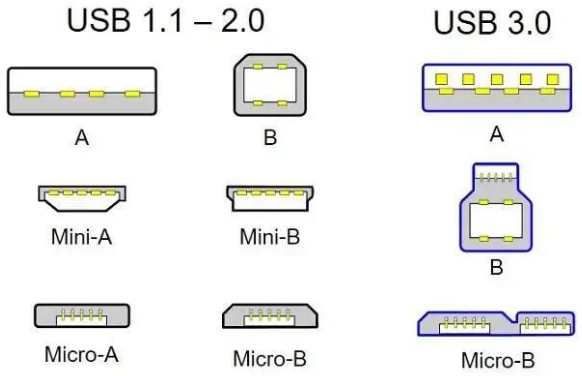
The company's US Archive's has thousands of external drives with content that requires multi-layered analysis

Additionally, there's an undetermined amount within the organization and 3rd party content creators

Each drive and disc requires leveraging video & audio production software, cables & adapters, and use of multiple platforms to access, inspect and archive data

Presently, it could take an Archivist a full workday to go through the end-to-end process of archiving the content on a signal drive

External Drives: Cables, Adapters & Hardware



USB C

External Drive Docking Station

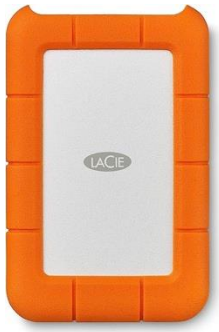
Hard Disk Drive (HDD)

Solid-State Drives (SSD)

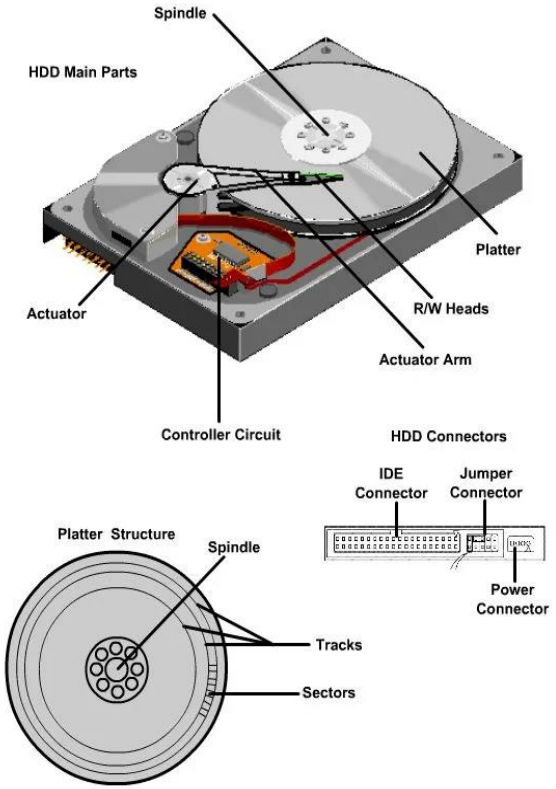
Thunderbolt

Firewire

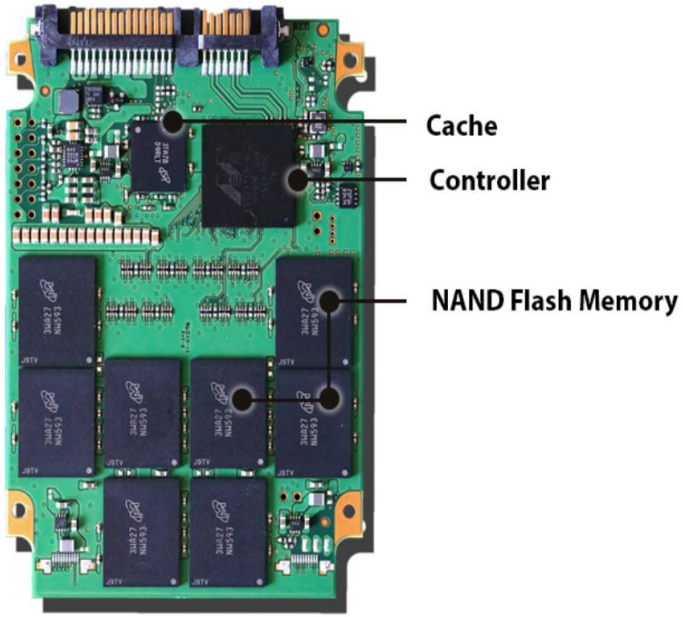
External Drives: Media Hardware Examples



Hard Disk Drive



Solid-State Drive

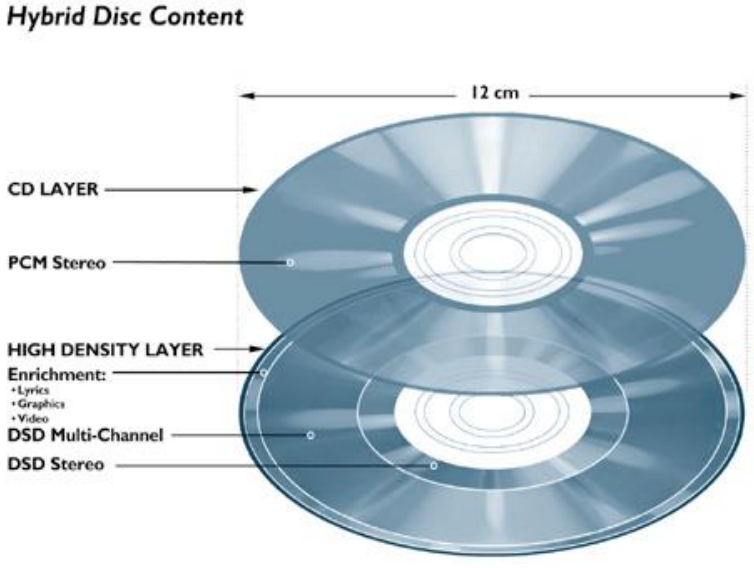


[External Drives Guide]

Additional Hardware Items…

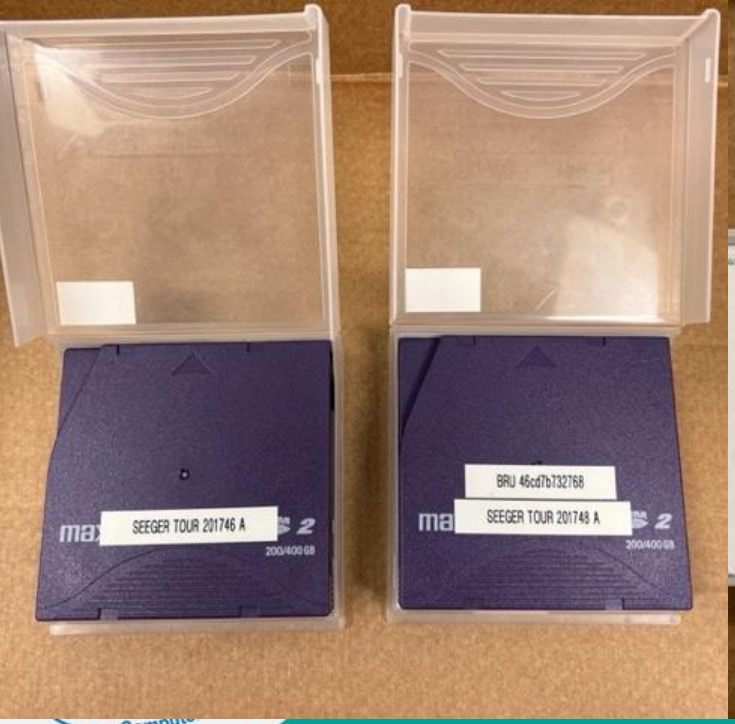
LTOs: Linear Tape Open (storage technology)

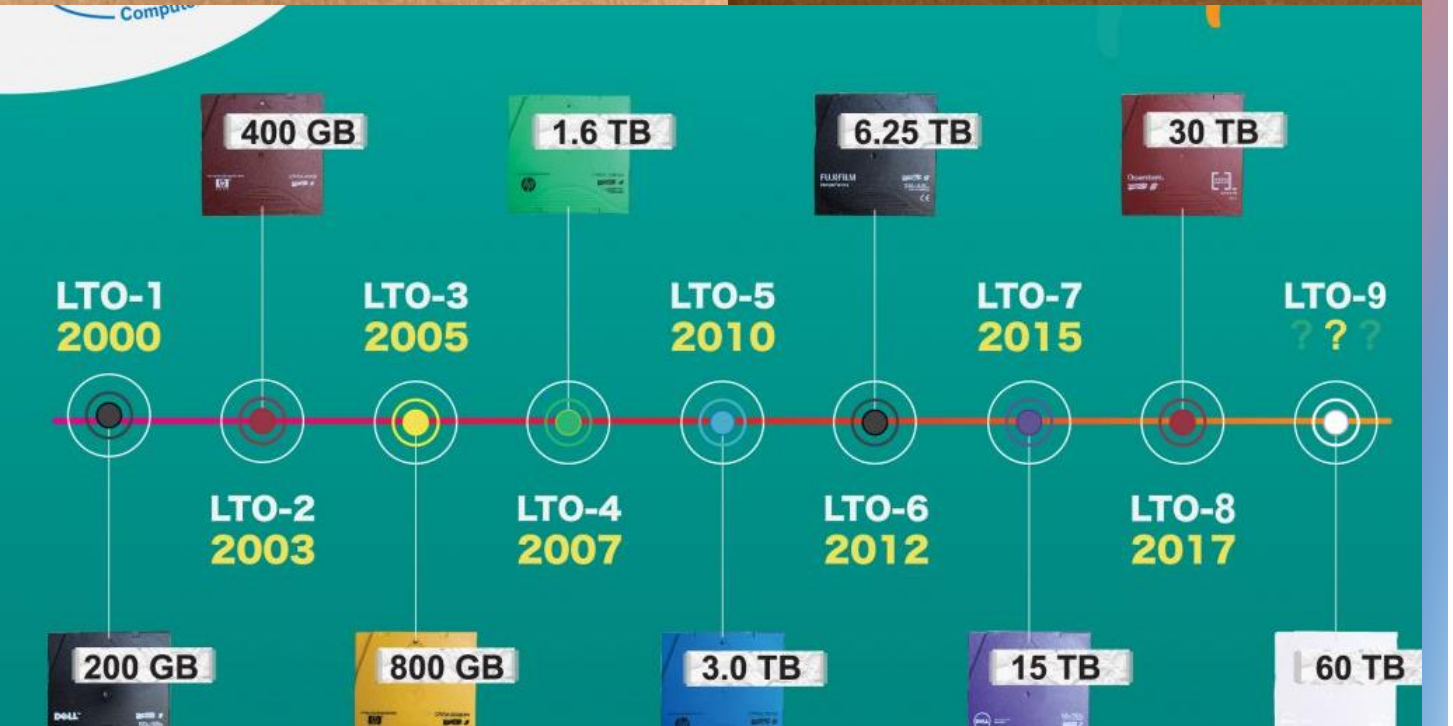
Compact Discs (CDs & DVDs)



[Disc Guide]

[LTO Guide]





LTOs: Linear Tape Open (storage technology)

- LTOs are linear media that only gets written and read in an end-to-end pass format
- It is a method meant for long term archival storage and backup, not 'in the moment' user collaboration
- Unlike hard drives, LTOs must be inserted into a machine where tape is separated from cartridge to read the contents
- Sony Music US Archive's has a substantial overlap in audio between backups to hard drives and backups to LTOs. They are generated as backups to hard drives
- LTOs represent verified and reviewed projects in their totality, often with associated metadata embedded
- We have on hand all types of LTOs, including LTFS formatted tapes (LTO-5 and above)

External Drives: Challenges & Realities

Time is Ticking:

External drives hardware, specifically Hard Disk Drives (HDDs), are prone to failure the older they are, making extraction of data challenging.

A Public Library:

Need a temporary library for content and metadata to be evaluated

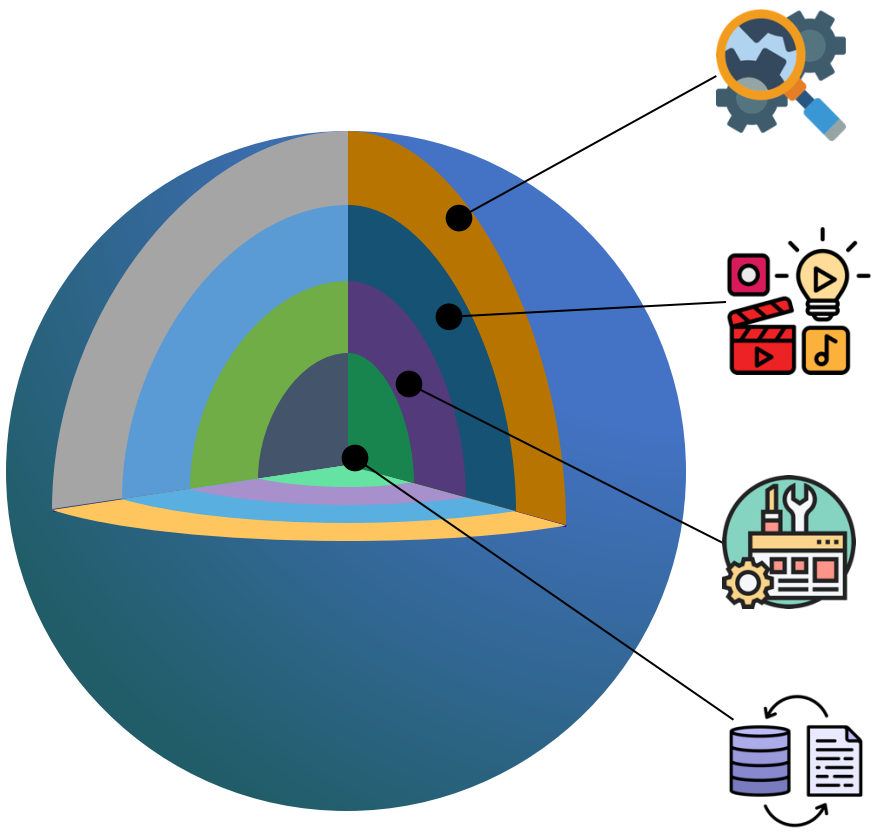
Lost in Space:

Drives are scattered in many places within Sony departments and with external partners

Input/Output System:

Require process mapping for collection of drives into a repository for centralized access

External Drives: Current as-is Workflow



Hardware Inspection

- Define the type and model of external drive; Hard Disk Drive (HDD) or Solid-State Drive (SSD)
- Line up appropriate cables and/or adapters to use for powering up external drive
- Run tests on the external drives to ensure there is access to the data
- If data is not accessible, work through different ways to retrieve it

Content Inspection

- Inspect database to disseminate what the content includes and that it is complete
- Ensure that there aren't any special characters in the asset file names
- Confirm that no unsupported files are present in the folders (Audio)
- Review any supplemental information that accompanied the external drives (Drive barcodes, summary sheets, etc.)
- Check to see if project data has been previously archived in the digital archive

Content Preparation

- Removal of characters/files that will be rejected by digital archive system
- Placement of content to another drive or cloud (if cannot access directly from external drive)
- Copying content into a zip file to upload into digital archive (if unable to re-format files)
- Completion of submission/input forms and entry of relevant metadata fields

Extraction → Ingestion

- Upload content into the digital archive
- If content fails to upload from the external drive, the Archivist will download the contents to a cloud server or local computer and try to upload to the digital archive
- Manual upload process is dependent on network bandwidth, toolset guidelines, etc.

External facing information for vendors

External Drives: Hardware Inspection

- Determine the drive type, HDD or SSD. This will decide which cables and supported adapters will be used
- There are three types of cables that would be needed depending on the model of the drives: FireWire, USB or Thunderbolt
- Due to the outdated technology of the drives, cable adapters will also need to be used for mounting the drive onto the user's computer
- A large percentage of the drives are formatted for Apple computers; therefore, a user must be hooked up to a Mac computer or have a Mac drive application
- User will check to ensure that the contents of the external drive can be accessed. If they cannot be accessed, a user will try different solutions to extract the data. In these instances, an SSD/HDD docking station could be used to extract the data from the drive
- The older the external drive, the likelihood increases that the drive may die, and the data wouldn’t be able to get accessed

External Drives: Content Inspection

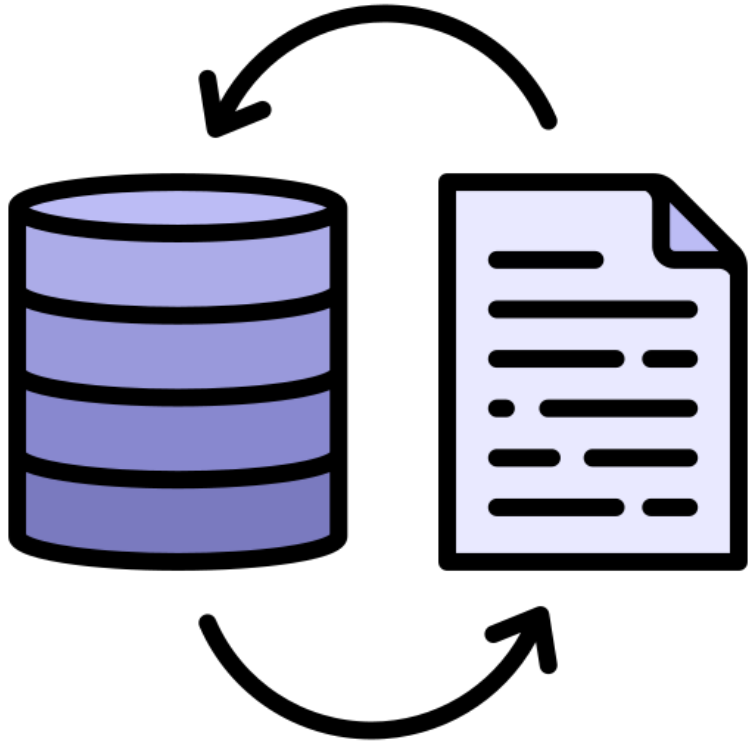
- Once the data is accessed, the Archivist performs quality control on the assets to digest and summarize what is present
- Review the file names to ensure there aren't any special characters (*Digital asset archiving system does not accept files with special characters)
- To keep the data from getting compromised, file names nor file formats can be changed without using the original software program used to create the files *(i.e., Pro Tools, Logic Pro, Premier Pro, Final Cut, etc.)* If the file name or file format is changed without it, then previewing the file in system will render a blank asset
- Remove any MD5 proxy files are present in the folders (Audio). The archiving system will not accept content that includes these proxy files, and they must be deleted in order to ingest *(Could take an hour to remove MD5 files from a 250GB drive)*
- Archivist will review any supplemental information that accompanied the external drives *(Drive barcodes, summary sheets, etc.)*
- Archivist will cross reference to confirm content is not already in the archiving system. This includes checking file names & folder sizes to see if there's a match.

External Drives: Content Preparation

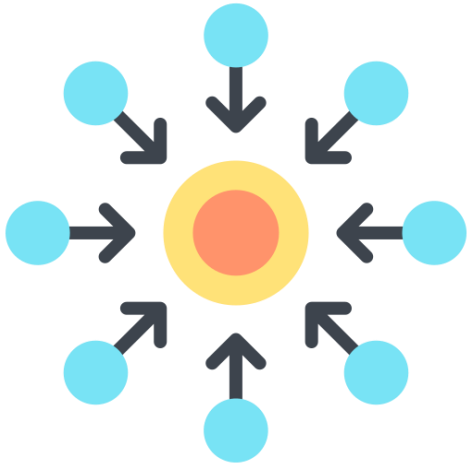
- If there is trouble accessing the data from the external drive, the Archivist will try and transfer the content to another drive or cloud-based service
- They will perform a search in a digital warehouse platform to see if the contents in the drive have connection to any products so that they can be assigned to them when uploaded into the digital asset archiving system
- If the Archivist does not have software tools available to make file changes, then they proceed to copy the content and create a zip file to upload into the system. Unfortunately, this means the end user must download the entire package in order to see all the contents. They cannot view individual files.
- The Archivist will clean up the asset characters/files that will be rejected by the archiving system's ingestion process
- Once content is ready, a system submission form is filled out with relevant metadata fields
- If a physical asset barcode exists (from physical inventory system), a form is filled out to input digital archive barcode linking the hardware and digitized asset

External Drives: Extraction & Ingestion

- If content fails to upload from the external drive after multiple attempts, the Archivist will download the contents to the media workflow tool or local computer and try to upload to digital asset archiving system.
- When uploading content into archiving system, there are many factors at play that determine how quickly the assets are ingested. These include age of the external drive, the network bandwidth, offsite internet connection, as well as how many multi-destination uploads are taking place from the drive.
- For example; uploading 100GB of audio/video content into archiving system from the HQ using IBM Aspera will take about 2.5 hours. This does not include the inspection, administrative, or registration work. When uploading 100GB from one of the offsite studios, it could take 3-5 hours.
- Video projects tend to work drives more than audio as they use more of the disc all the time and can sometimes make it harder to get the data from.
- IBM Aspera application is a high-speed transfer tool that assists in the acceleration of uploads/downloads into the various company systems.



External Drives: How can you help?



Centralize & Digitize

- Help with inventory management & media audit of external drives
- Assist in extracting content from hardware into a cloud platform

Content Management

- Stage content and sourced metadata for access in a consumable format
- Assist with data formatting to stay current with software technologies

Toolset Partnership

- Explore possible data integration points with the internal systems
- Help with automation opportunities