AVPS

AudioVisual Preservation Solutions

Permanence Persistence Preservation

The Future of Content is in Your Hands

Preservation: The Shift from Format to Strategy

Presented By

Chris Lacinak

October 8th, 2006

121st Audio Engineering Society Convention

chris@avpreserve.com | www.avpreserve.com

Preservation: The shift from format to strategy

It’s simple

10 easy steps

1. Analog = Bad

- Boooo!!!!

2. Digital = Good

3. Buy Digital Storage Medium of Choice

The Cheaper and Bigger, the Better

Our Price: $349.99
600GB External USB 2.0 Hard Drive

Free up space on your desktop or notebook by stashing your photos, videos, music and other files on this massive 600GB hard drive.

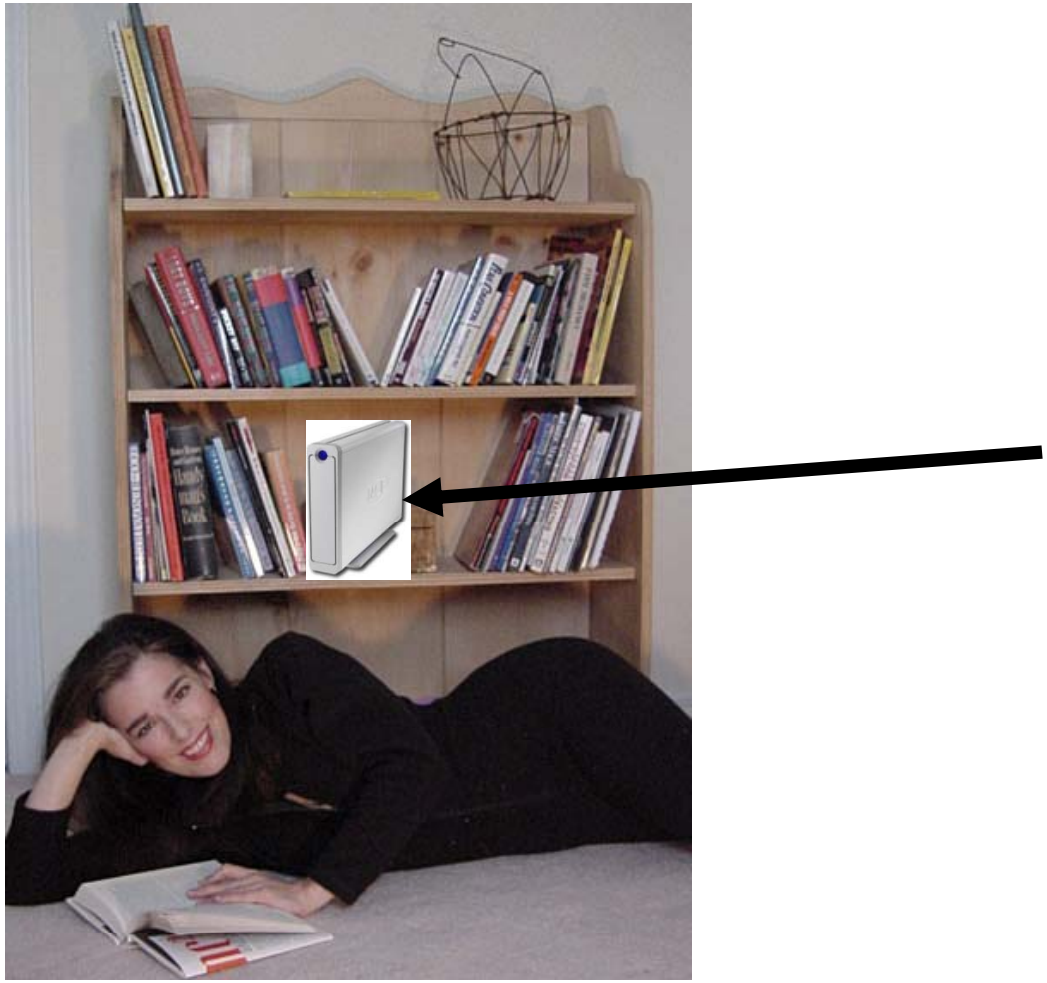
Do it online!

4. Transfer to Digital

Minimize your exposure to analog materials!

5. Put analog in it’s place

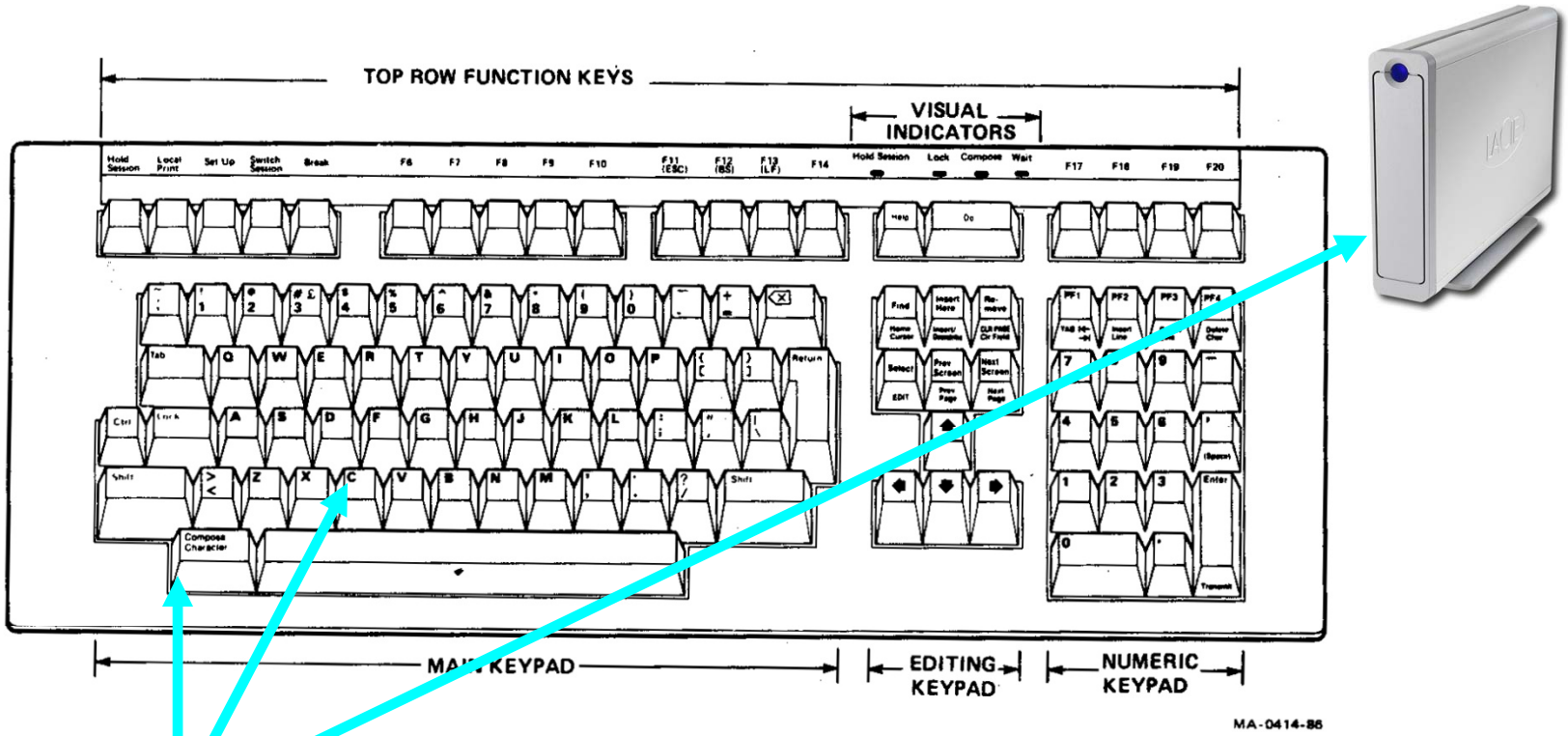
6. Place on a shelf



7. Go Have Some Fun and Relax…It’s Digital

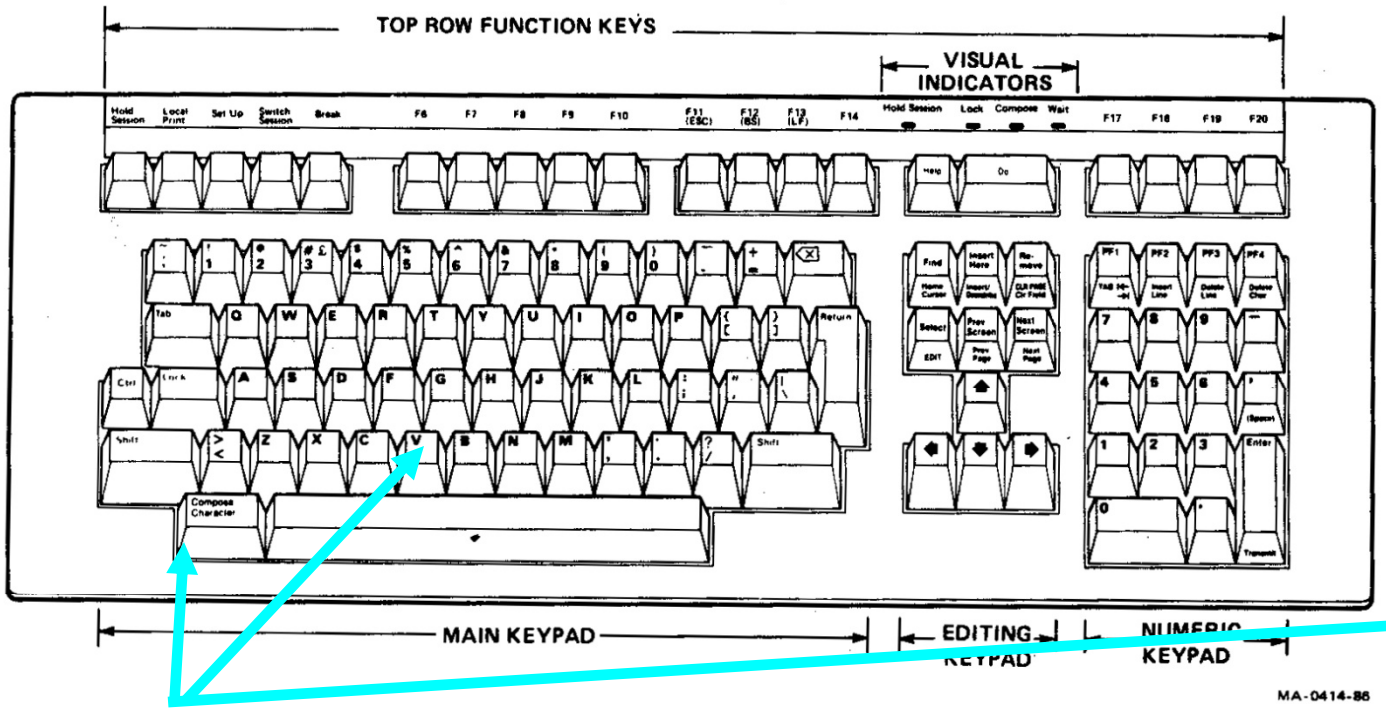
8. Come back 100 years later

9. Copy



CTRL C

10. Paste



CTRL V (Careful not to hit C again or you'll have to go back one step)

Newer/Better Version

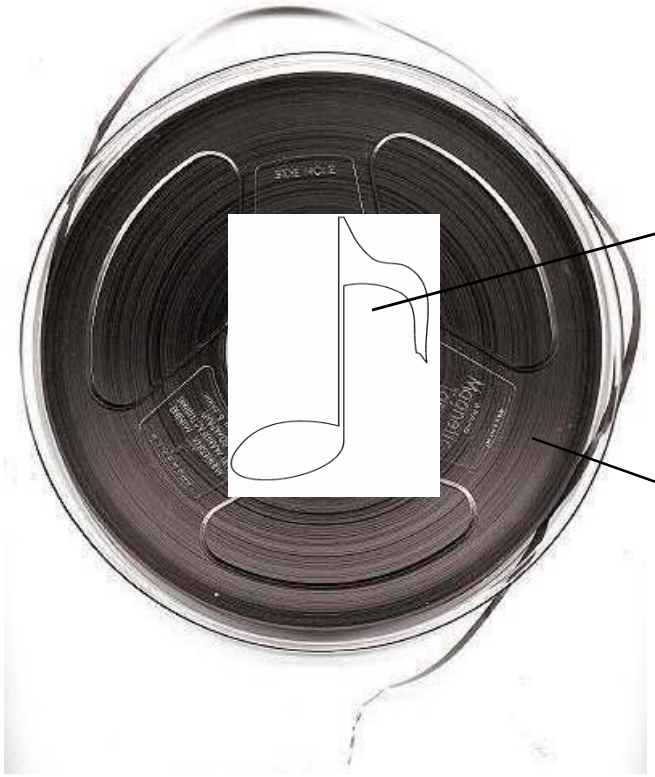
Now Go Preserve!

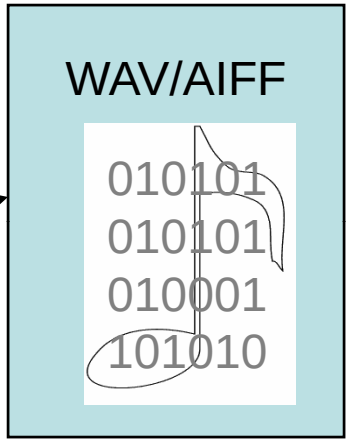
The End

Use of terminology throughout this presentation

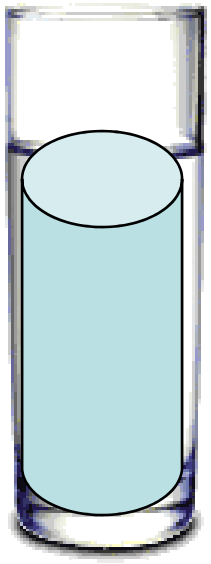
- Digital – An intangible object such as a WAVE file on a disc.
- Physical – A tangible object such as a reel of tape.
- Physical Digital – a tangible object containing digital information (CD, DAT,etc.)

Why the title
(and how the title is wrong)

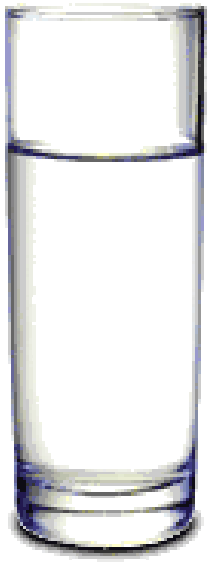


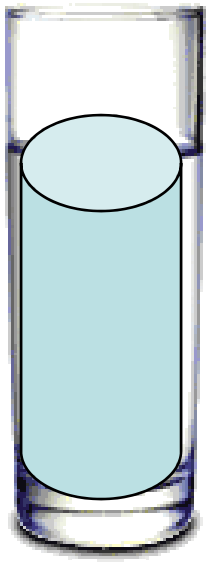


Object of Preservation

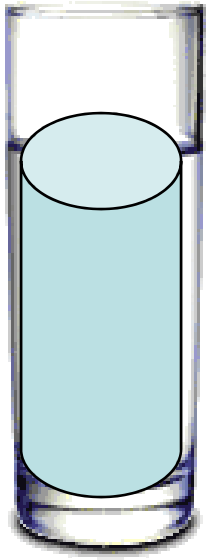


Object of Preservation

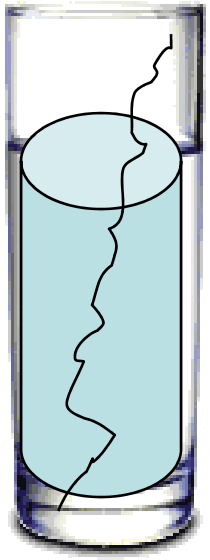




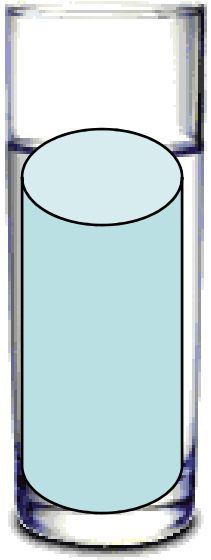
Object of Preservation

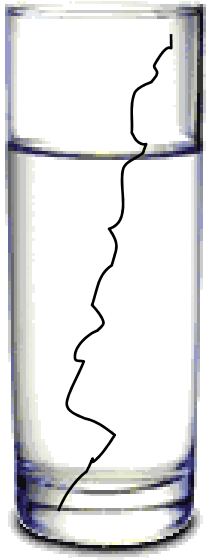


Object of Preservation

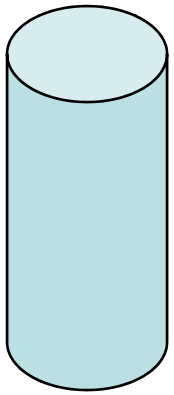


Object of Preservation

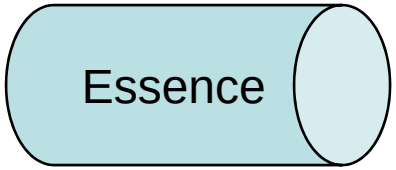




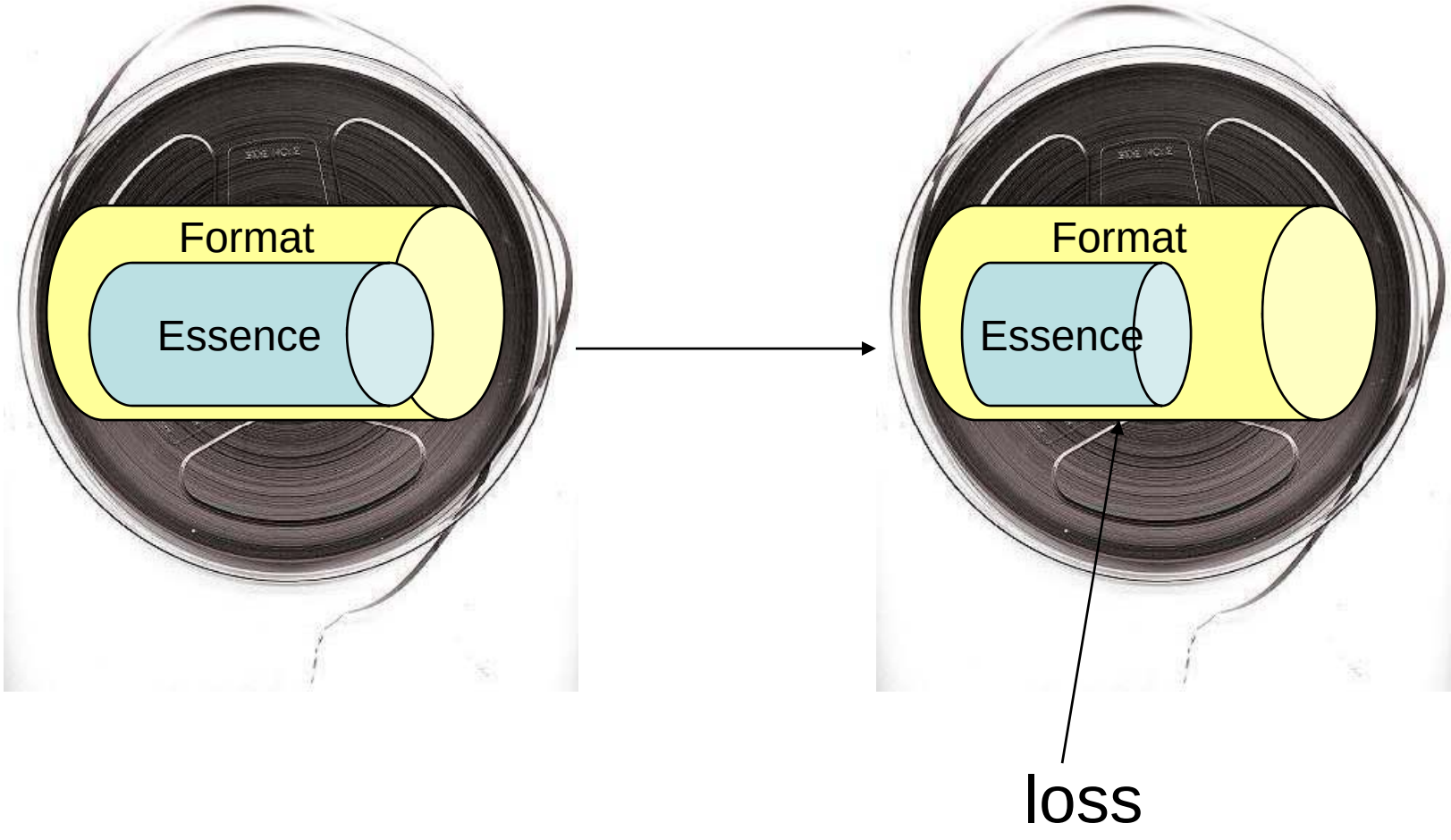
Object of Preservation



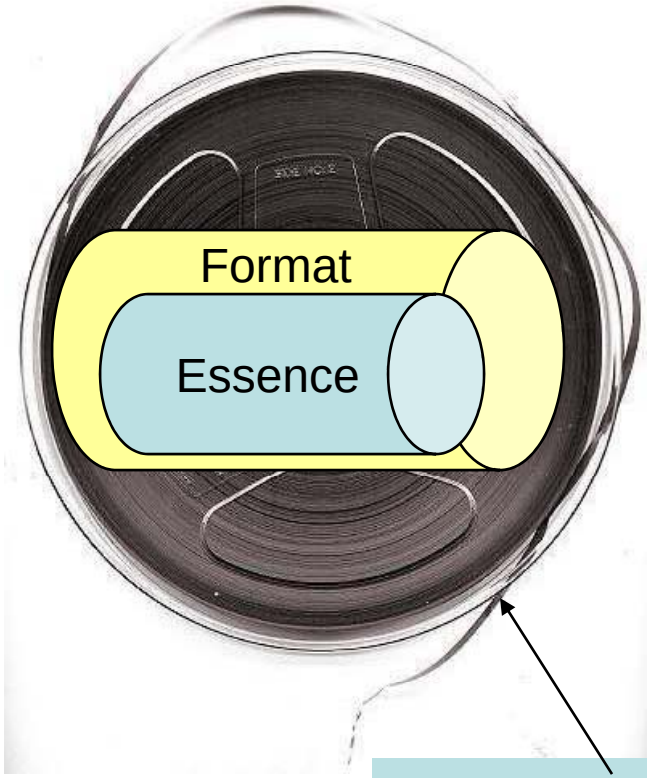
Object of Preservation



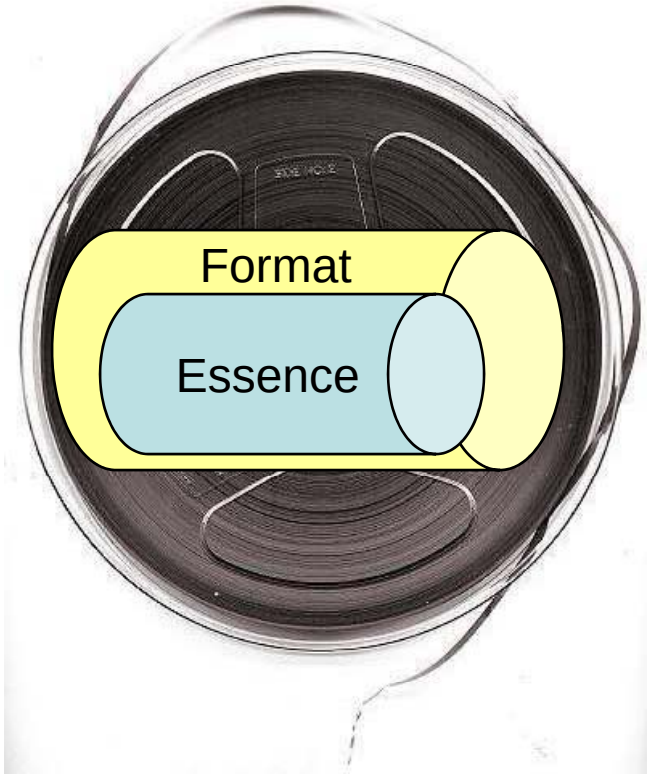
Physical Migration

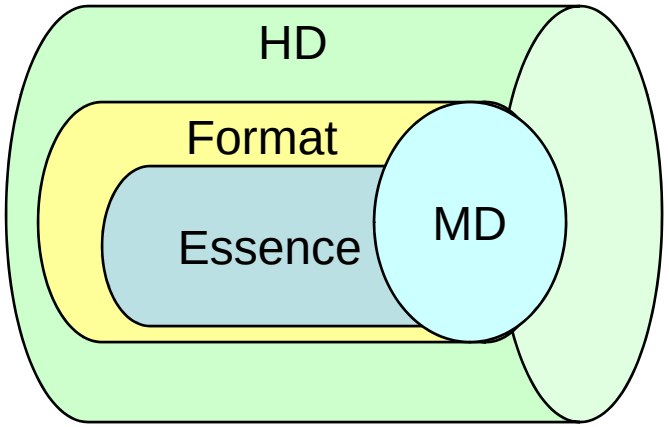


Physical object

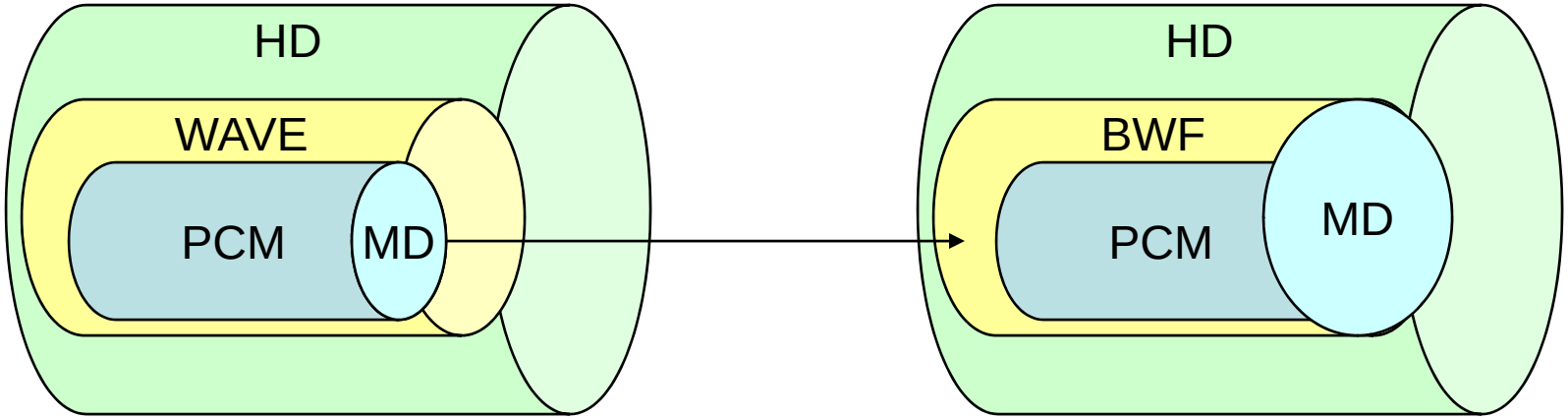


Digital Capture without Loss

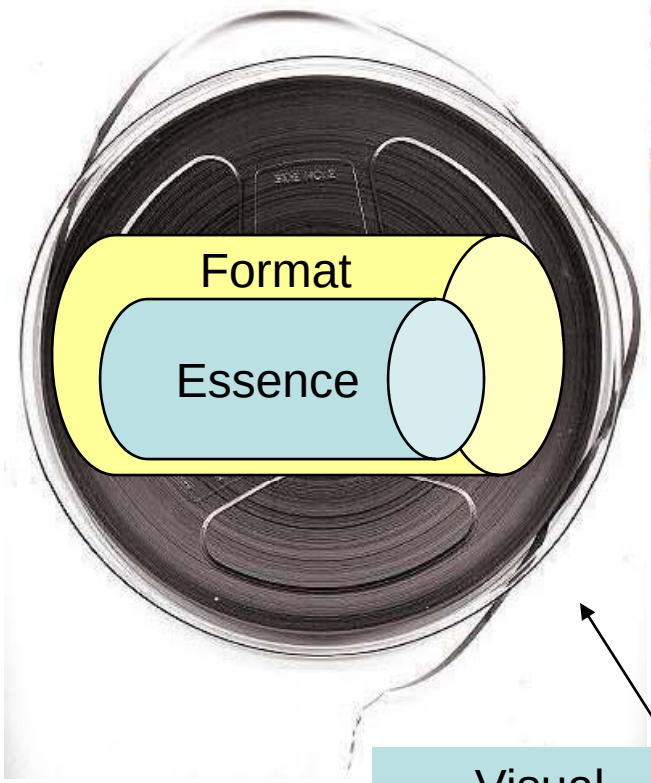


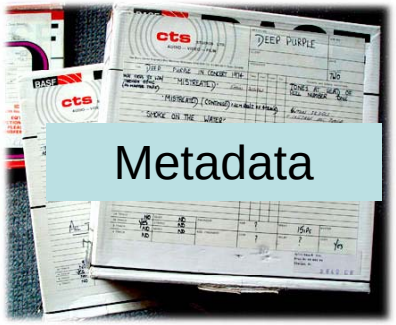


Lossless migration

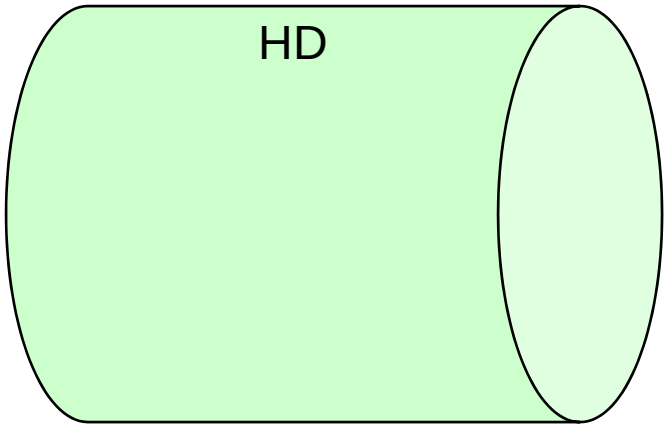


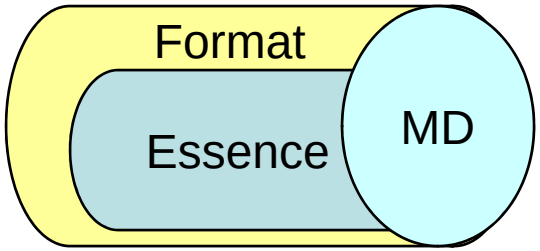
Independence of Media, Format, Essence and metadata



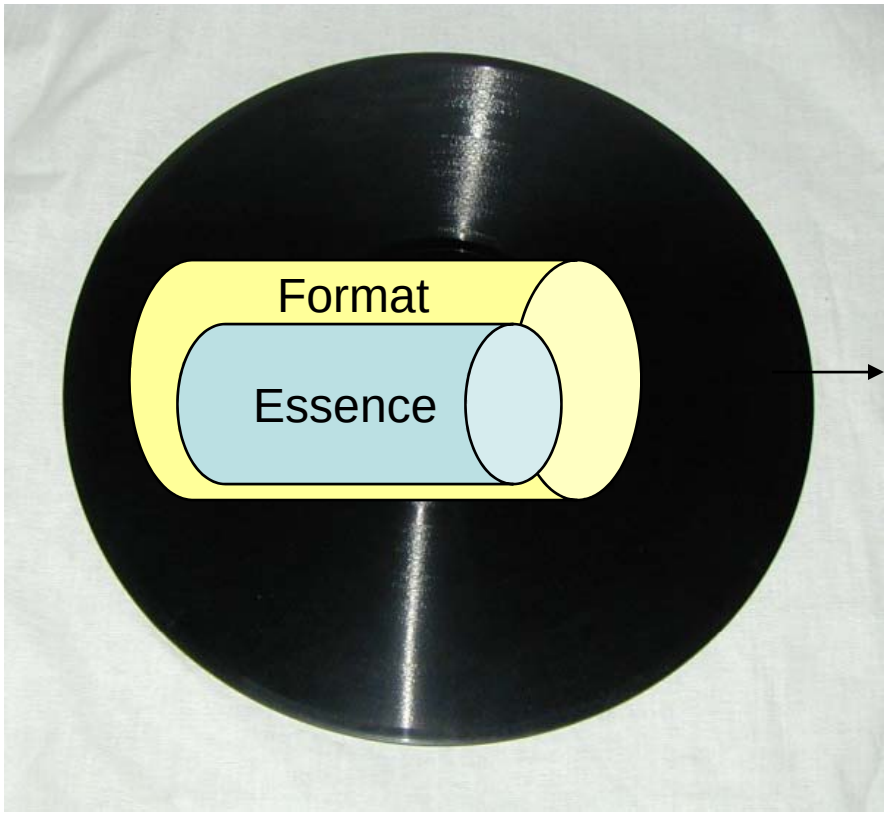


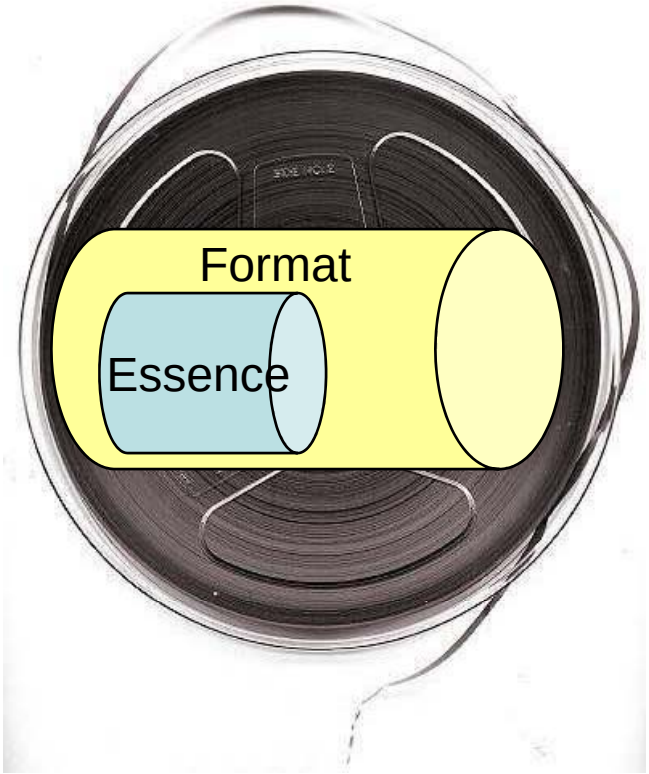
Visual Metadata



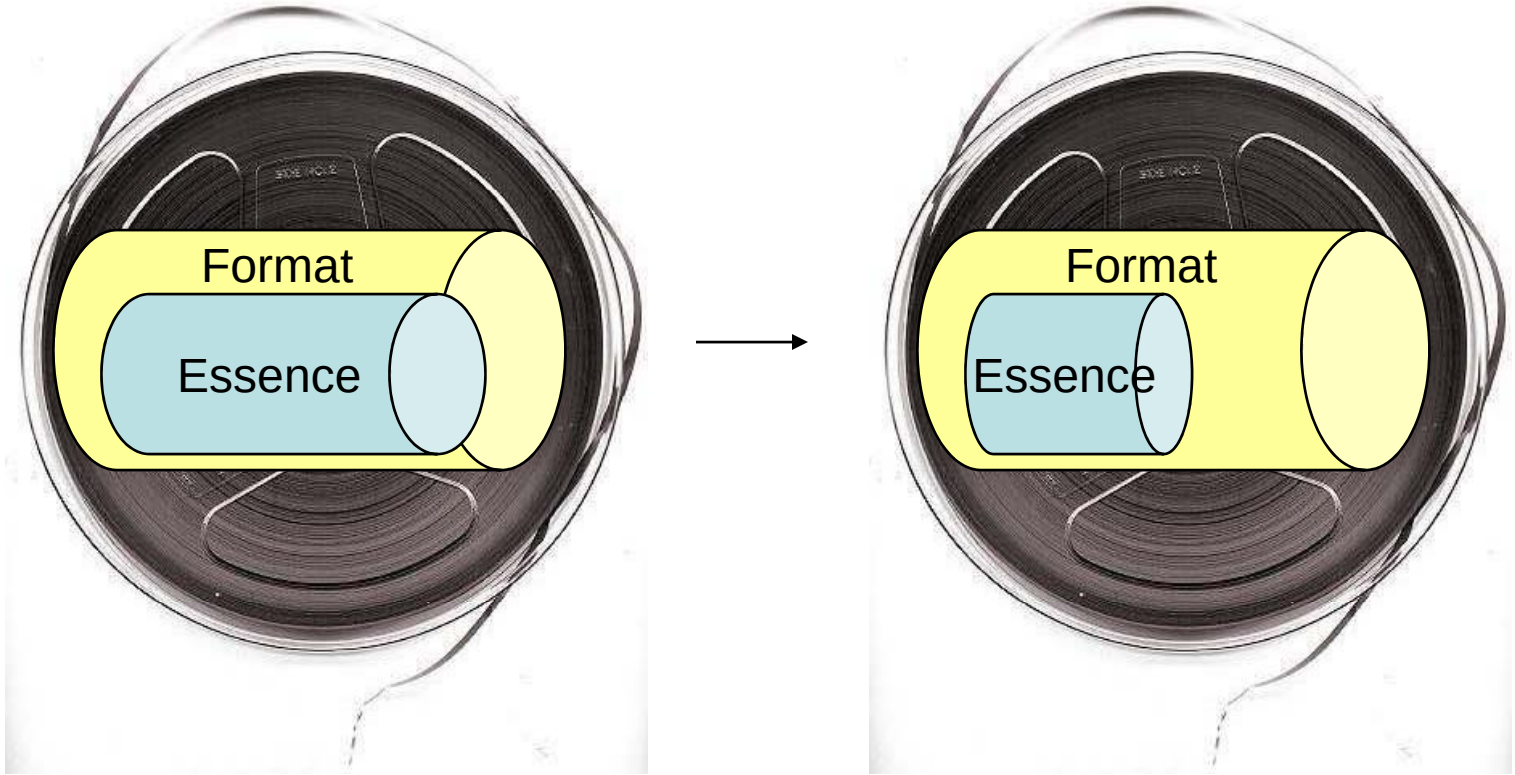


Separation and new terminology Migration

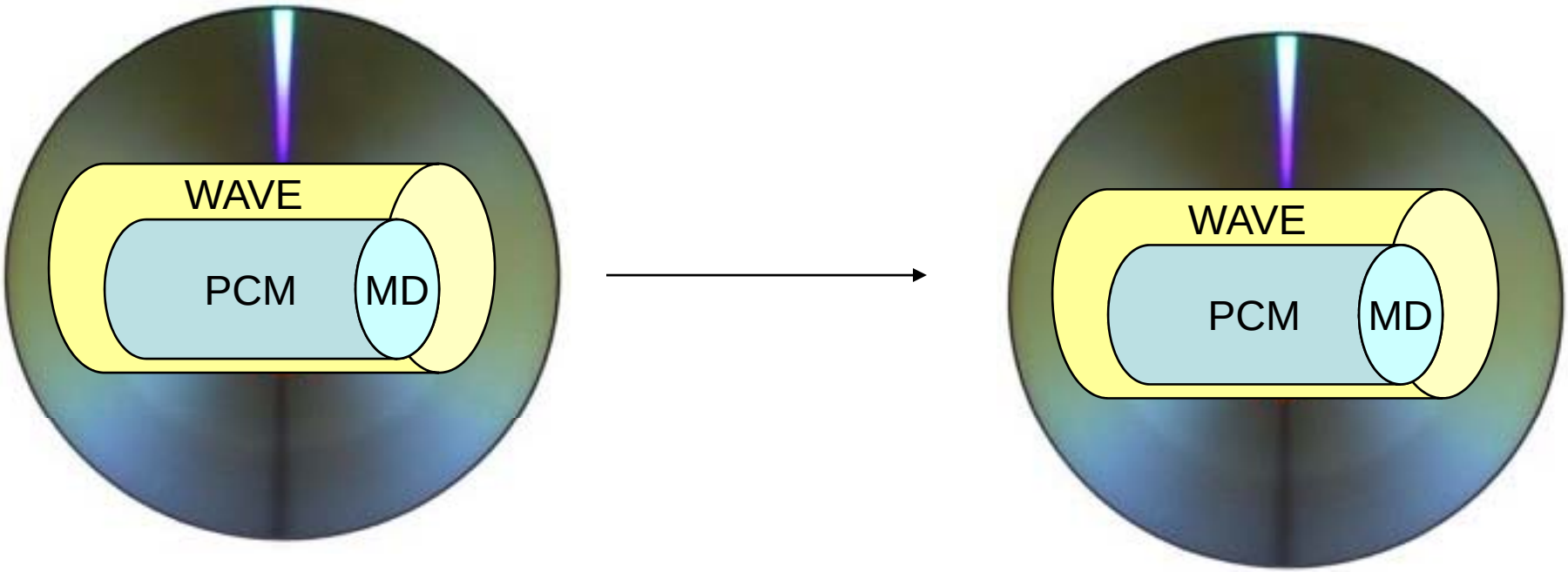




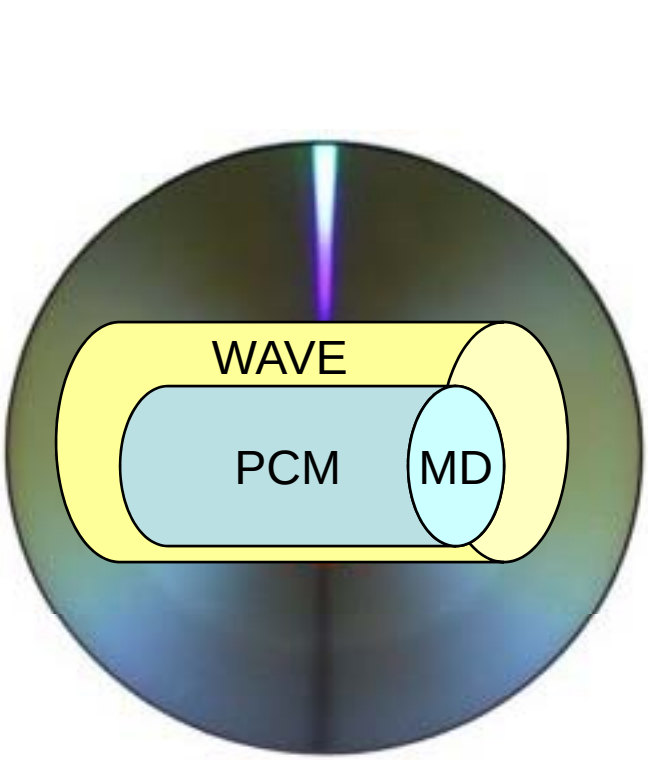
Separation and new terminology Migration

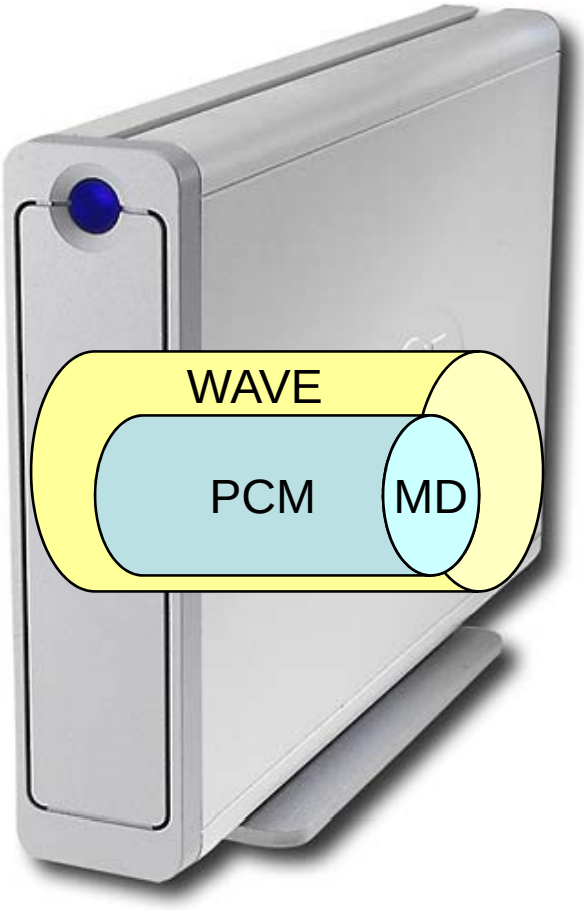


Separation and new terminology Refresh

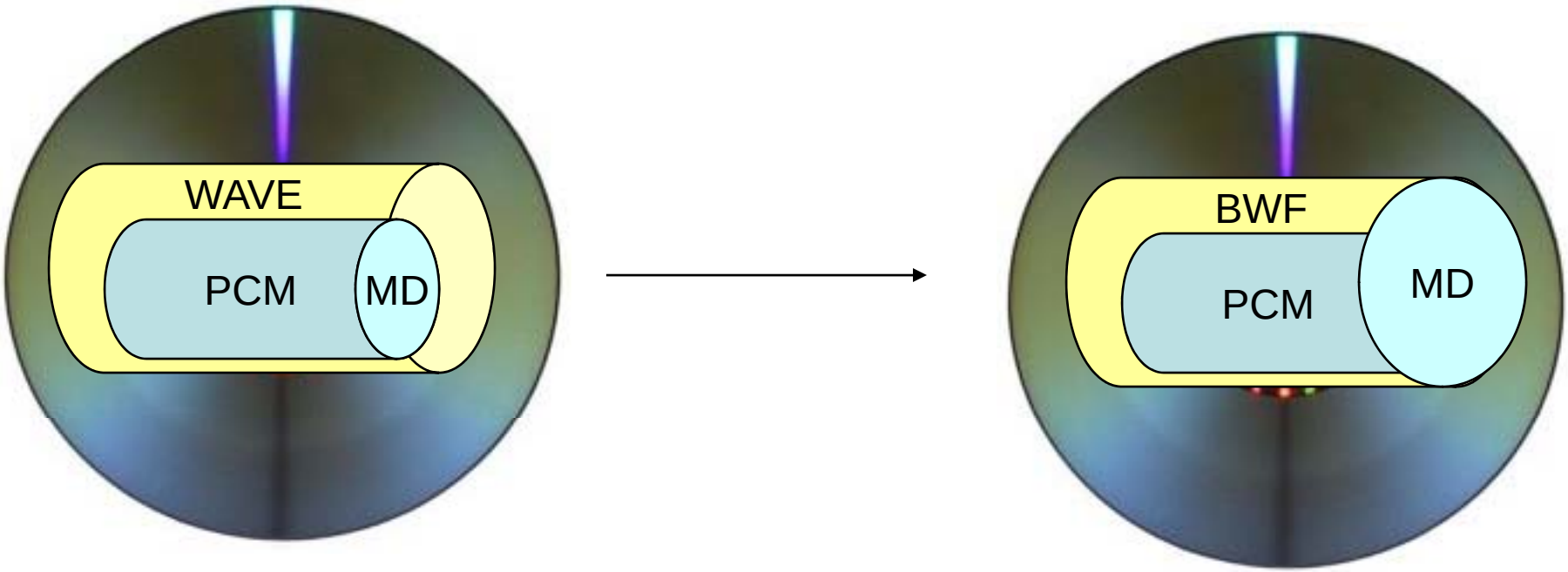


Separation and new terminology Refresh

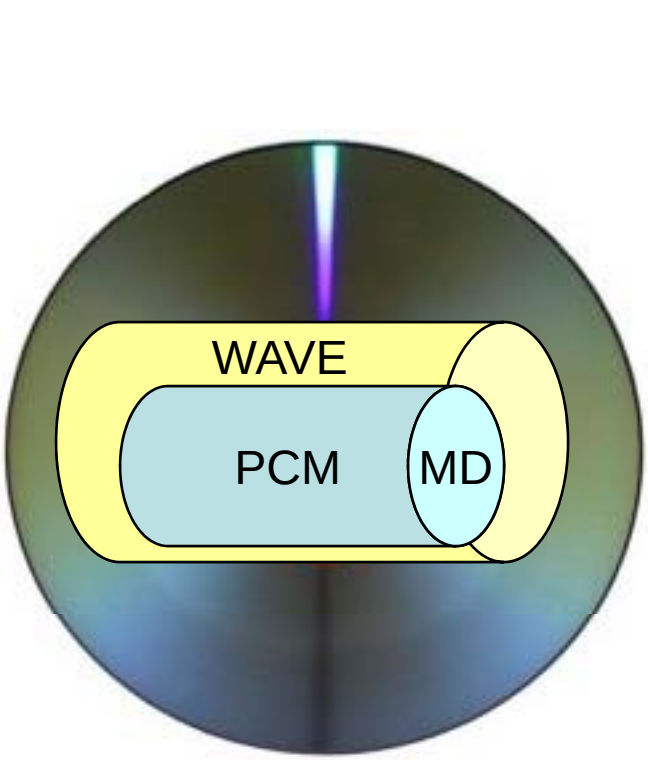


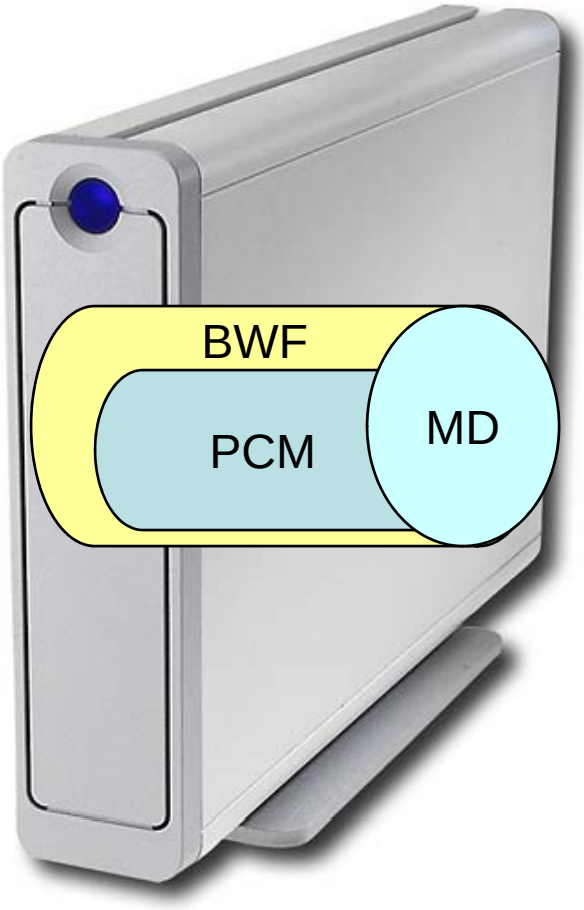


Separation and new terminology Migration



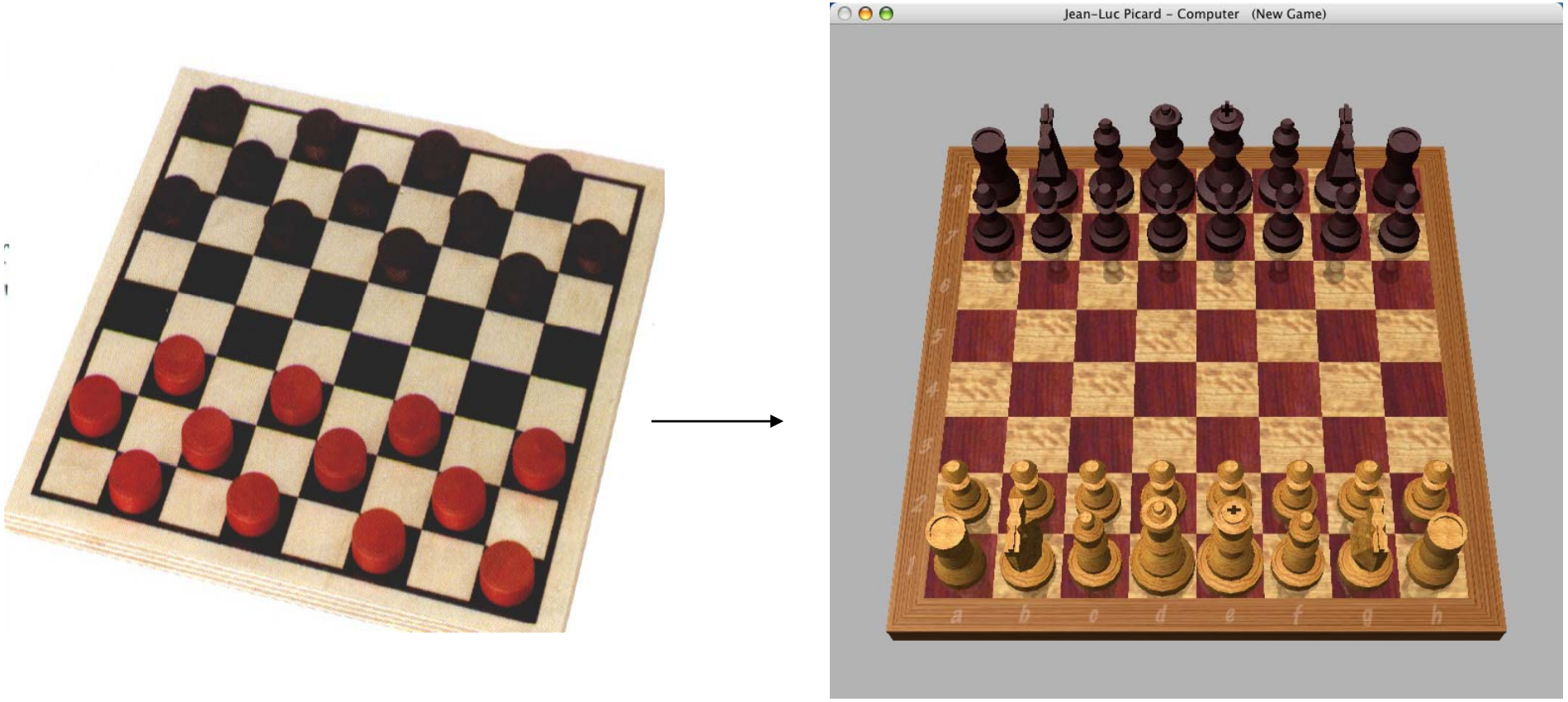
Separation and new terminology Migration





Our Game is Forced to Evolve

One move game plan to a complex strategy

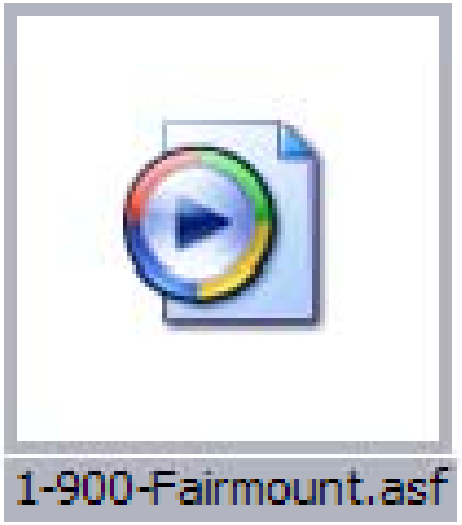


Call to Action

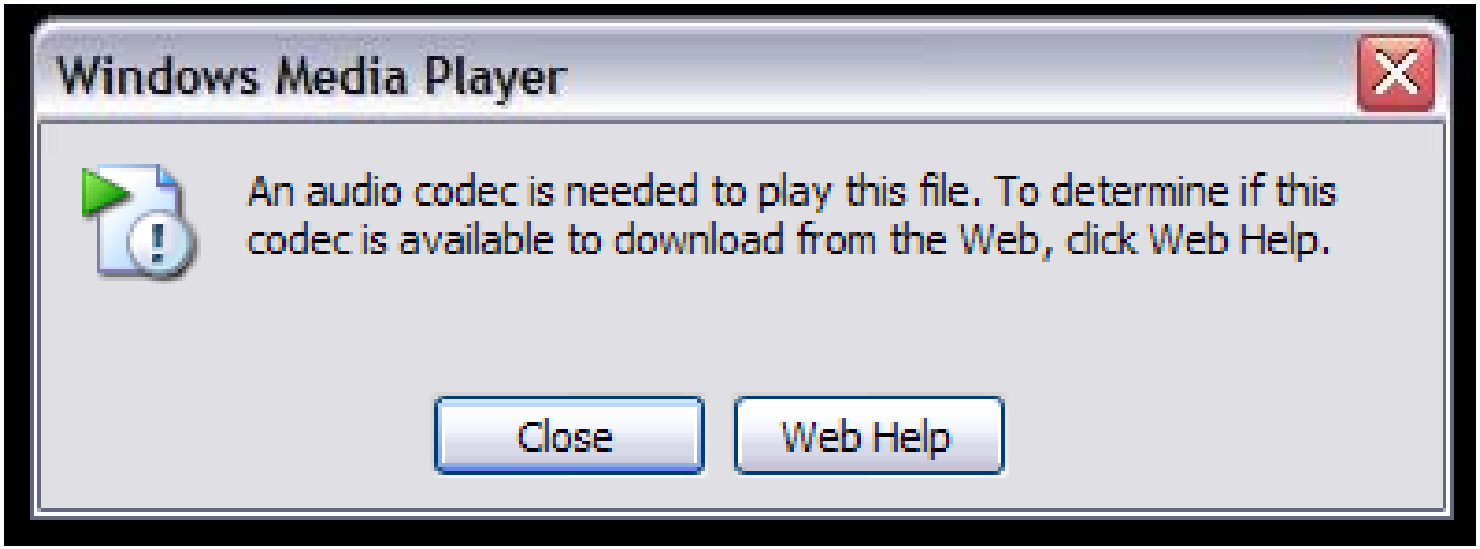
"In the analog world, previous formats persisted over time. Cuneiform tablets, papyrus, and books all exist until someone or something (fires, earthquakes) takes action to destroy them. But the default for digital information is not to survive unless someone takes conscious action to make them persist."

– Howard Besser of NYU out of the Handbook for Digital Projects

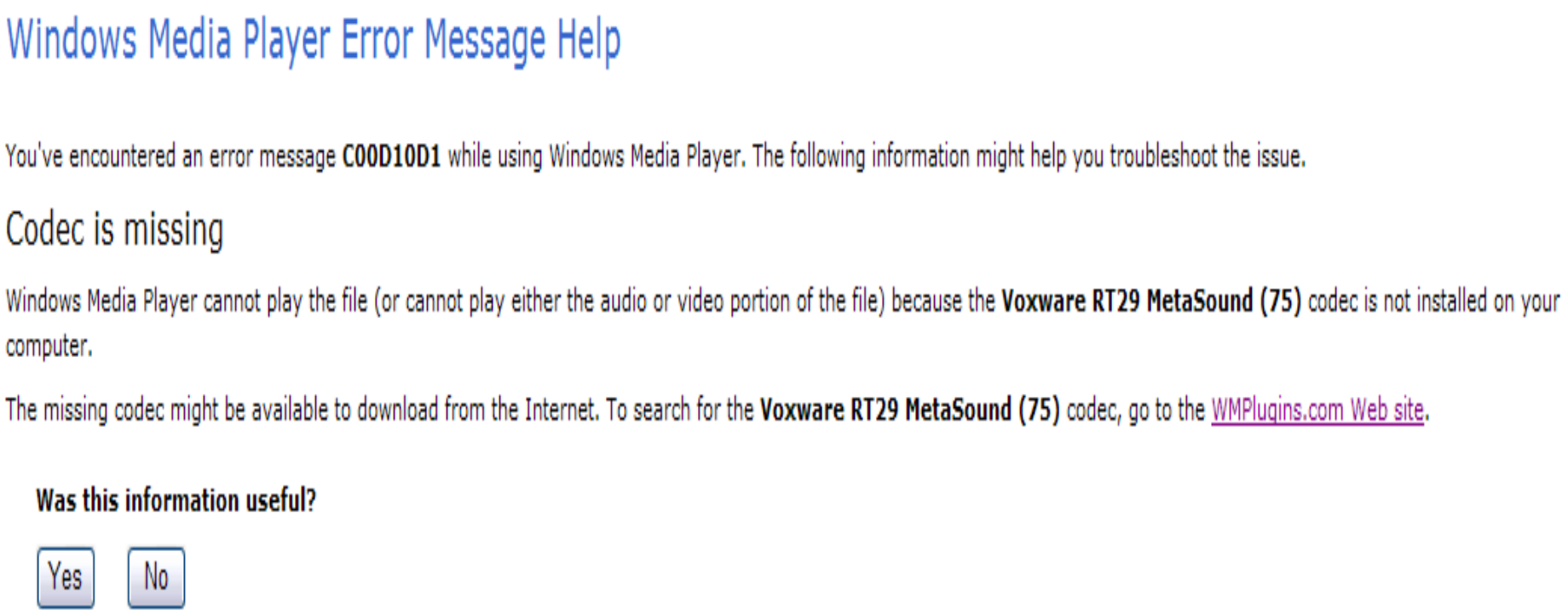
Call to Action



Call to Action



Call to Action



Call to Action

Place on a shelf

Joke…

but also realistic

Poor Follow Through

- Eager move to digital
- Lack of direction on digital preservation
- Default position of physical materials
- Storage on shelf again falling short of preservation, but this time with greater burdens to bear
- What does a preservation system consist of

Looking at a Preservation System

- Open Archival Information System
  - Ingest
  - Store
  - Manage
  - Maintain
  - Disseminate

By the way

- Physical collections call for the same requirements
- Never met due to excessive resource utilization

What is the act of preserving

"preservation is the totality of things necessary to ensure the permanent accessibility – forever – of an audiovisual document with the maximum integrity"

Ray Edmonson *Audiovisual Archiving: Philosophy and Principles*

Alignment

- Looking at preservation system requirements in the context of meeting the definition of preservation we run into some problems in the digital domain.
- These are problems that a preservation system needs to address

Digital System Problems

- Viewing
- Scrambling
- Inter-Relation
- Custodial
- Translation

Digital System Problems

The Viewing Problem

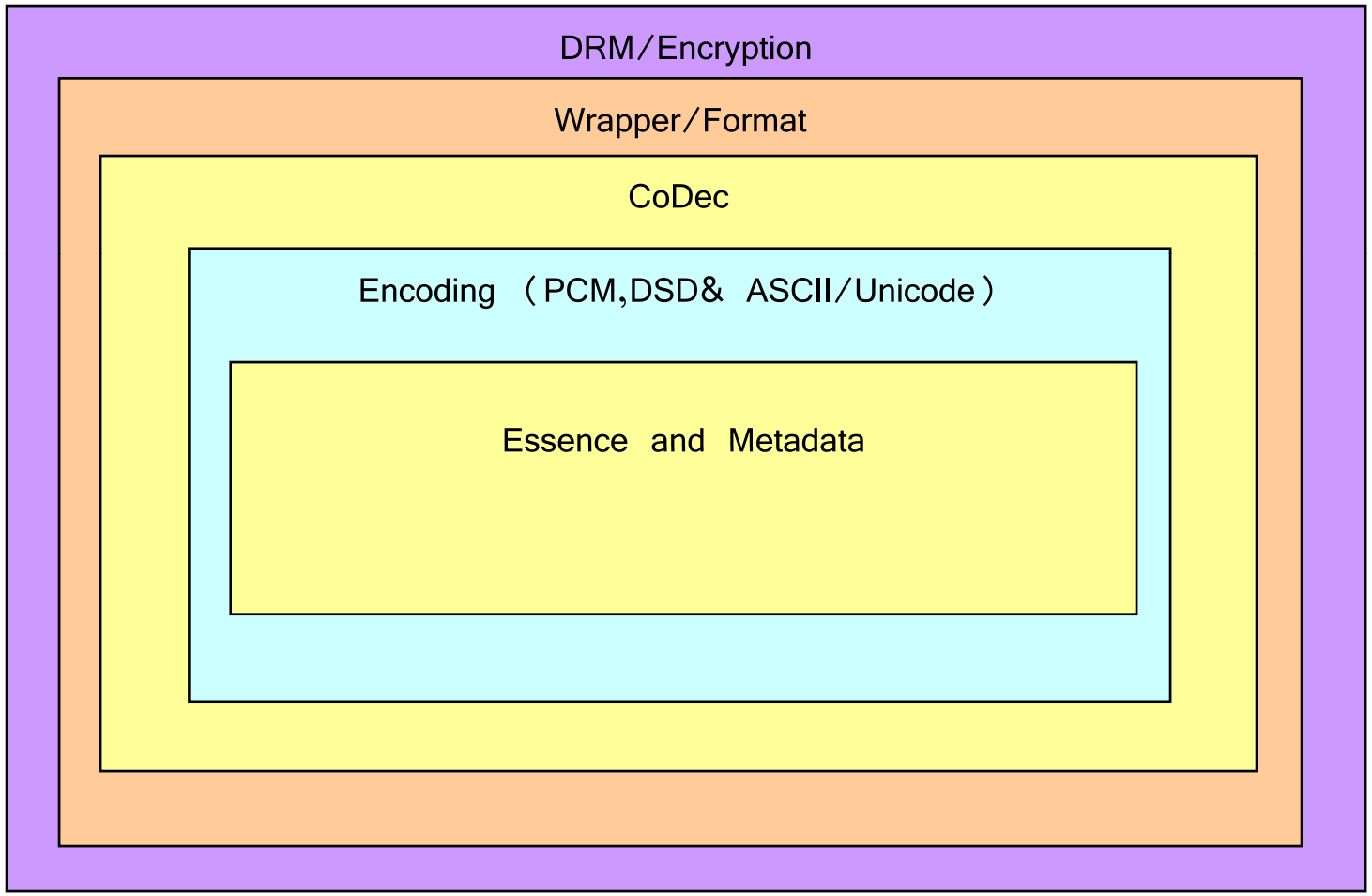
Maintaining the infrastructure . Same issue as with physical, but with an obsolescence rate that deters expertise and saturation. Also with independent components, there's more things to keep track of that are changing. (ie. codecs, file formats, wrappers, metadata, versions, hardware, etc…)

Digital System Problems

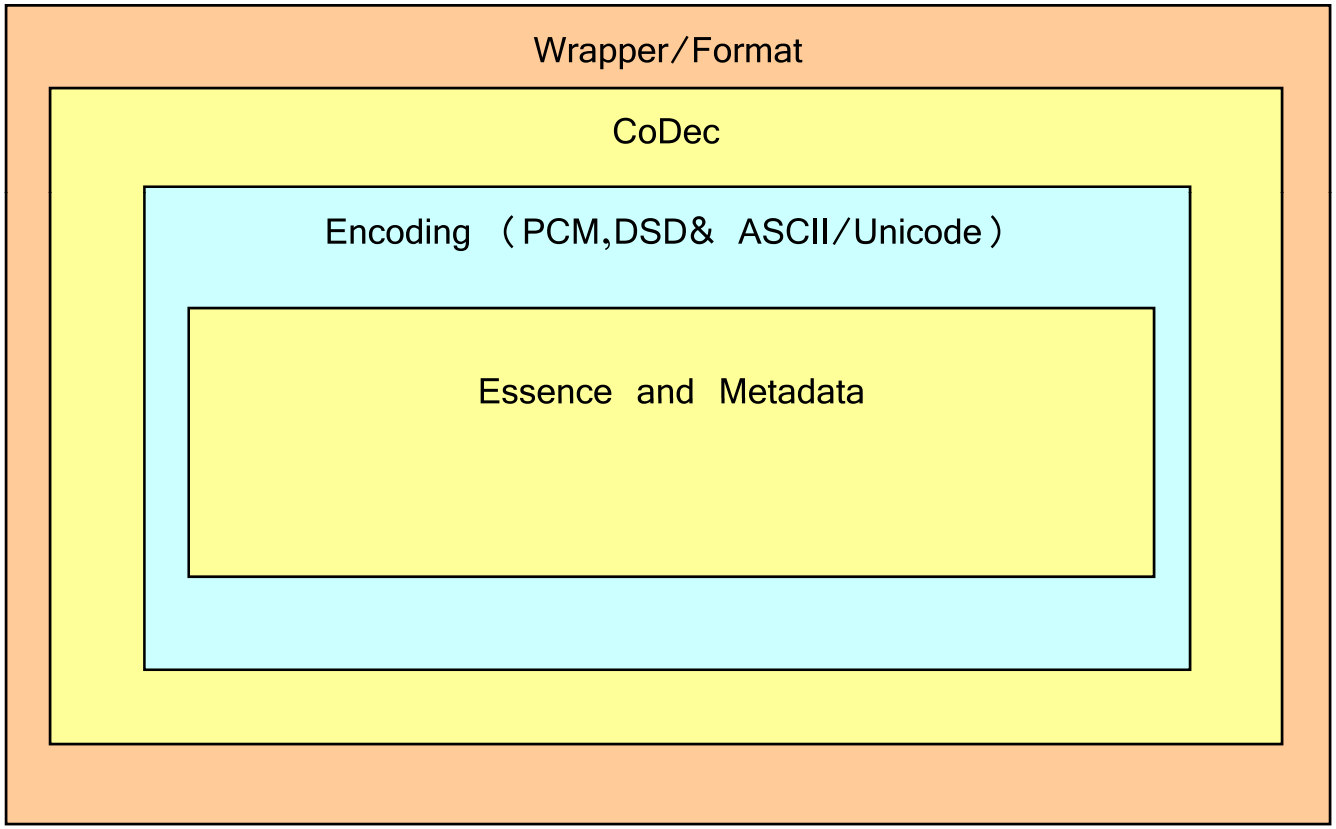
The Scrambling Problem –

Every Layer is another layer to decipher in the future. Each one adds complexity. These include Encoding, CoDecs, DRM, Encryption/Security, File Format Proprieties.

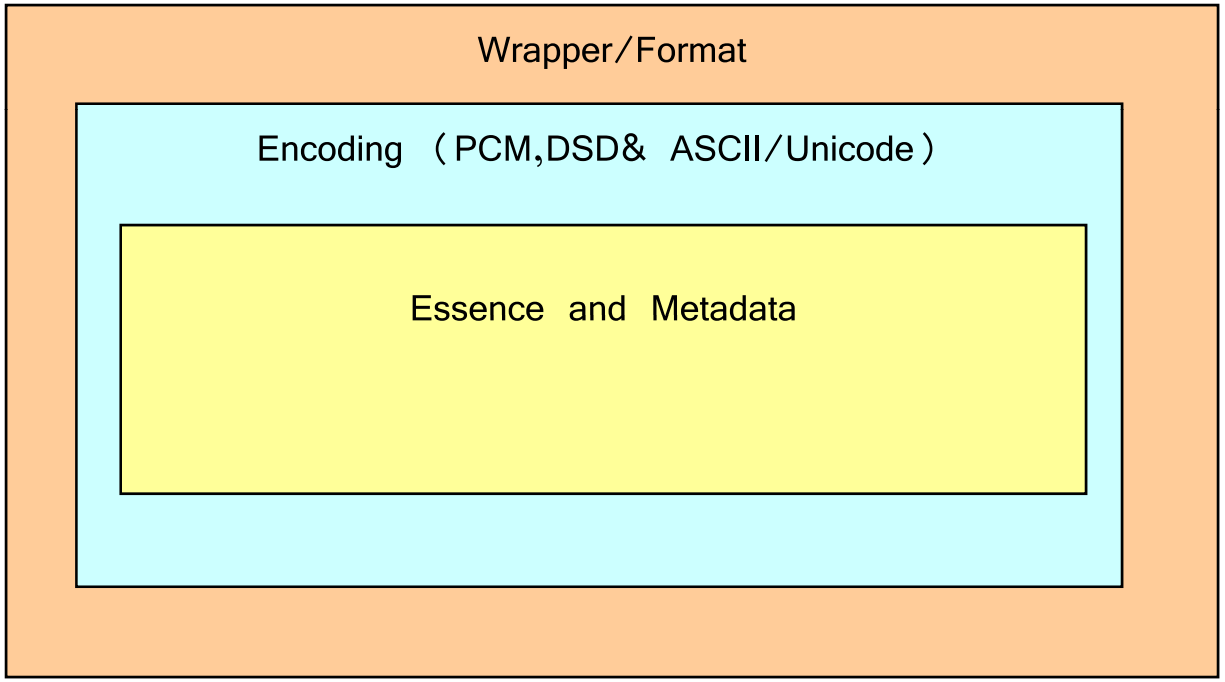
Barriers to Access



Barriers to Access



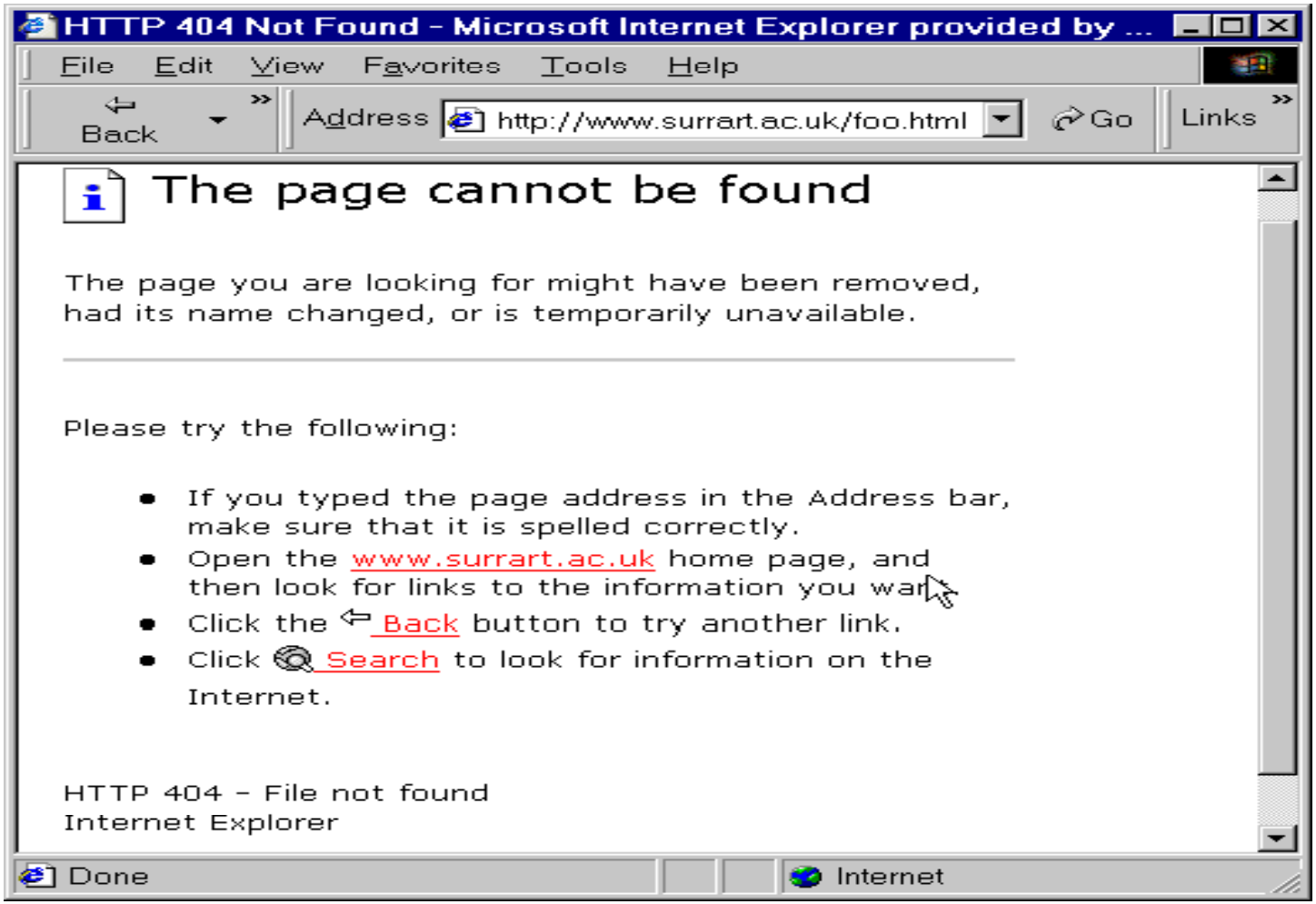
Barriers to Access



Digital System Problems

- **The Inter-relation Problem –** Proper preservation requires representing the original object to the greatest extent possible. Representing the relationships to the object maintains authenticity and proper representation of the object. From an access perspective as people continue to increasingly integrate digital audio and video objects into their work (Presentations, Classroom, etc.) persistence of that link becomes imperative to the preservation of the new work and an ongoing awareness of the referenced content.

Inter-Relation Problem



Digital System Problems

- **The Custodial Problem –** Lack of definition about whose responsibility it is for employing preservation oriented practices in the creation of the object and in the preservation of the digital object. IT or Archivist. Musician/Producer or Record Label. What does a digital object ready for deposit into a digital repository look like and who does it make sense for the assurance of conformance with that? What makes up a “Trusted” digital repository and who looks after and maintains that?

Digital System Problems

- **The Translation Problem –** This looks at interoperability factors. Formats such as MXF and AAF and standards such as AES31 look to solve this problem. This is the issue we face when moving from one system to another. We know that we will be forced to follow the evolution of systems and the formats that they support. So when migration is performed we have little doubt that we'll be able to migrate the bits that make up the essence, but what about the granularity and structure of metadata fields supported by future systems. Or how markers and playlists are represented.

Strategic Choices

Sustainability Factors of File Formats

- Disclosure
- Adoption
- Transparency
- Self Documentation
- External Dependencies
- Impact of Patents
- Technical Protection Mechanisms

Library of Congress “Sustainability of Digital Formats Planning for Library of Congress Collections” in ed. Fleischhauer, Carl and Caroline R. Arms
http://www.digitalpreservation.gov/formats/index.shtml

Sustainability Factors of File Formats

Disclosure

Disclosure refers to the degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. Preservation of content in a given digital format over the long term is not feasible without an understanding of how the information is represented (encoded) as bits and bytes in digital files.

Sustainability Factors of File Formats

Adoption

Adoption refers to the degree to which the format is already used by the primary creators, disseminators, or users of information resources. Saturation in the market and your community.

Sustainability Factors of File Formats

Transparency

Transparency refers to the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor. Digital formats in which the underlying information is represented simply and directly will be easier to migrate to new formats and more susceptible to digital archaeology.

Added layers such as compression, DRM, and encryption all add barriers to preservation.

Sustainability Factors of File Formats

Self-documentation

Digital objects that are self-documenting are likely to be easier to sustain over the long term and less vulnerable to catastrophe than data objects that are stored separately from all the metadata needed to render the data as usable information or understand its context.

Side Note about Metadata

- Metadata is the information expected to be required to manage and preserve an object.
- A digital object at a glance gives us a filename and that is it. Therefore the role of metadata in supporting preservation is of primary and absolute importance in preserving the essence from the perspective of both mid to long term preservation of enabling digital archaeology as well as enabling access and finding those objects

Sustainability Factors of File Formats

External dependencies

External dependencies refers to the degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments. The less dependencies the greater chance of sustainability.

Sustainability Factors of File Formats

Impact of patents

Patents related to a digital format may inhibit the ability of archival institutions to sustain content in that format. The existence of patents may slow the development of open source encoders and decoders and prices for commercial software for transcoding content in obsolescent formats may incorporate high license fees.

Patent law nuances dictate that patents may very well go well beyond the 20 year term defined. Mitigating such risks is the fact that patents require a level of disclosure that should facilitate the development of tools once the relevant patents have expired.

Sustainability Factors of File Formats

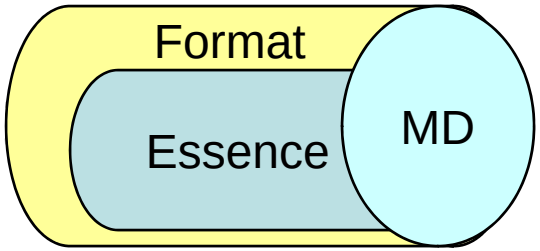
Technical protection mechanisms

custodians must be able to manage content in the face of changing technology. Content for which a trusted repository takes long-term responsibility must not be protected by technical mechanisms such as encryption, implemented in ways that prevent custodians from taking appropriate steps to preserve the digital content and make it accessible to future generations.

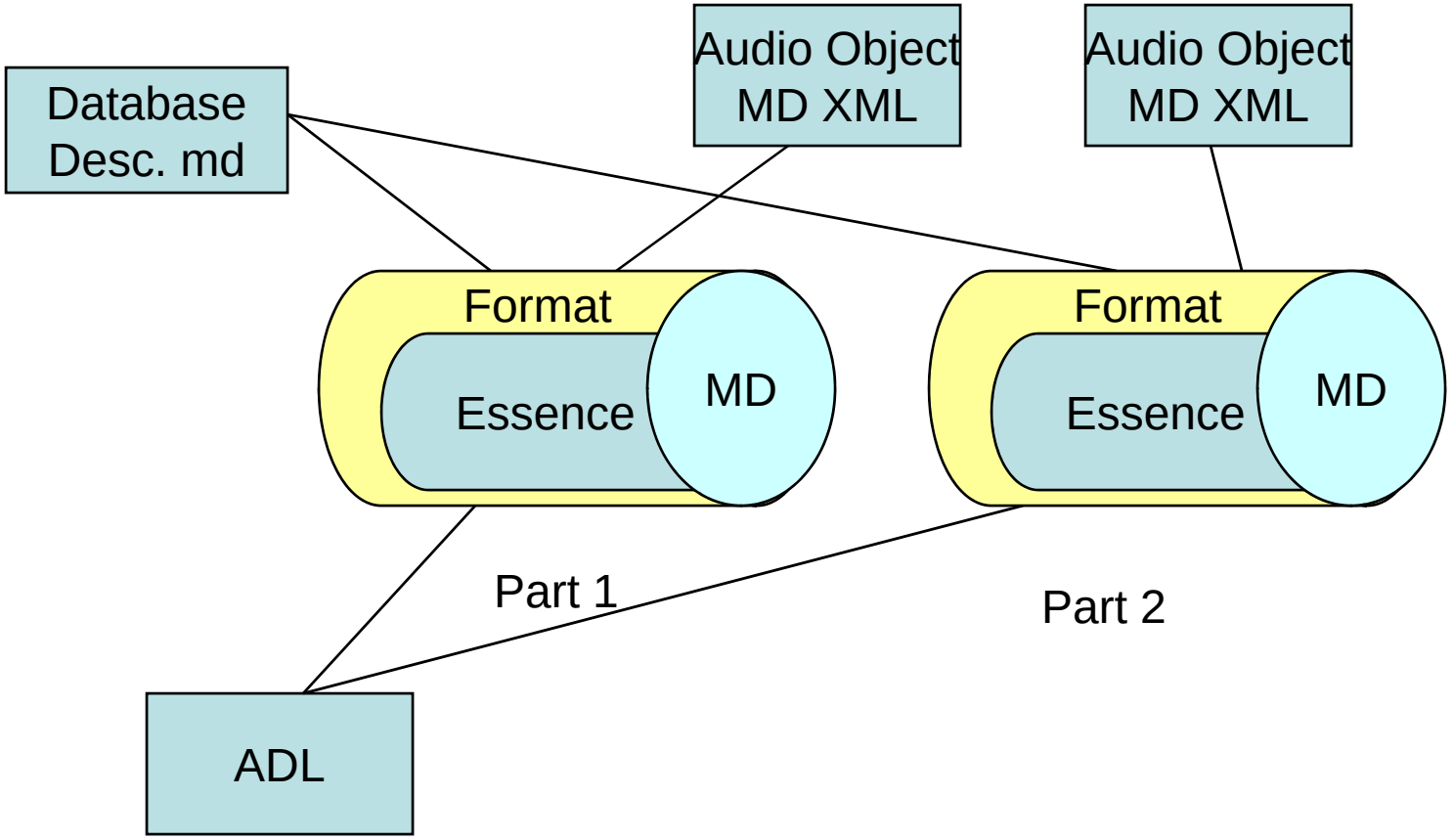
File Format to Package

- A collection of information
  - Interoperability (SIP and DIP in OAIS)
  - Stored and Managed Package (AIP in OAIS)

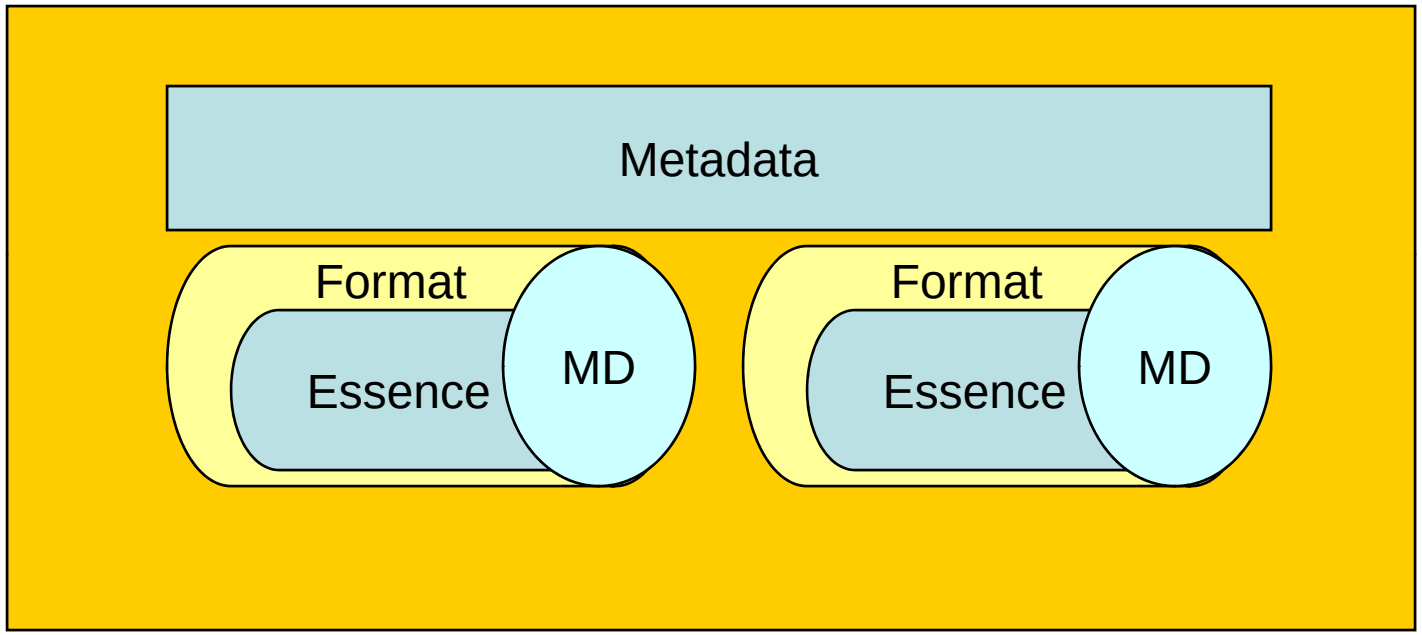
File Format to Package



File Format to Package



Package



Managing the Package

- Packages as part of a preservation system are engaged in the following activities
  - Ingest of essence and metadata
  - Metadata storage, access, updating (multiple databases, changes in ownership, new derivatives, file location change)
  - Access to essence
  - Data integrity checking
  - Refreshing
  - Migration
  - Creation of new derivatives (for distribution, research, or just updating access technologies)
  - Security/Authentication/Permissions (Not DRM or encryption. Just entry into the domain. Physical would be a secured physical location. Electronic would be who has access to the server.

Is this why digital preservation is so expensive?

- In the physical domain this collection of information exists but is an abstract package
- Performing Preservation system activities requires a great amount of labor
- Comparison of performing the same activities reveals analog much higher resource utilization

Physical Digital Objects CD/DVD/RDAT/DASH

- Worst of Both worlds
- Faced with media failure, degradation, reproduction issues and obsolescence in the physical domain. You will also be faced with having to actively address all of the problems previously mentioned in the digital domain. Not to mention that you are able to take advantage of very few, if any of the advantages of entering into the digital domain.

Properly managed digital preservation storage

- Enables much more opportunity for automated activates including
  - Data integrity measures
  - Documentation
  - Refreshing
  - Migration implementation
  - Creation of new derivatives
  - Security
  - Metadata dissemination/updating
- Greater Accuracy and efficiency

Comparison Chart

Higher # - Greater Resource Utilization

Rating factors were time, man/auto & opp. for error

	DAT	CD-A/ DVD-V	CD-R/ DVD-R	Data Tape	HD	DVD Jukebox	Tape Library	DMSS
Ingest (from)	12	12	9	5	4	6	6	3
Data integrity at regular intervals	14	11	11	8	8	6	5.5	3
Refreshing	15	11	9	7.5	5.5	6	4	3
Migration	14	14	11	9.5	7.5	6	5.5	3
Metadata storage, access, updating	15	15	15	11	11	10	9	3
Access	13	13	13	11	11	8	7	5
Creation of new derivatives	14	14	11	10	8	7.5	5	3
Security/Authentication/Permissions	13	13	13	13	6	8	8	3
TOTAL	110	103	92	75	61	57.5	50	26

Guidelines

1. Reformat from analog carrier with the highest integrity possible (you are forever imprinting the existence of the signal in the digital domain)
2. Create unaltered digital master representing a faithful reproduction of the original recording
3. Capture at the greatest resolution possible (primarily driven by limitations of the institution and archive size)
4. No data reduction employed (lossy very bad, lossless bad) (ie. MP3)
5. No encryption or DRM wrappers on preservation copies
6. Maximum ACCURATE documentation about the source and resulting audio objects
7. Maximum ACCURATE documentation about the path/processes from point to point in the life of the content.
8. Maximum ACCURATE documentation about the lifecycle of the object (physical condition of object documented at points of engagement, error rates over a period of time)

Guidelines cont'd

9. Routine data integrity measures are required
10. Automated data integrity measures are better than manual
11. Choose File Formats over digital streams (DAT, CD-Audio)
12. Error rates as low as possible, with refreshing well before uncorrectable errors occur
13. Migration must occur before obsolescence of the format. Leave plenty of time. Migration takes much planning and testing prior to implementation.
14. At least two copies made of preservation data and kept in separate physical locations
15. Access copies made to restrict access to master file
16. The more "open" the better (Less proprietary, More disclosure. Standardized. For instance BWAVE is a documented standard that can be validated for conformance and reverse engineered using the documentation down the road)
17. The greater the adoption in the market the better

AVPS

AudioVisual Preservation Solutions

Permanence Persistence Preservation

The Future of Content is in Your Hands

Preservation: The Shift from Format to Strategy

Chris Lacinak

AudioVisual Preservation Solutions

chris@avpreserve.com