



# Safe and Efficient Preservation of Images and Other (Biodiversity) Media

Mike Webster

Director, Macaulay Library, Cornell University

The **Cornell** Lab  of Ornithology  
**Macaulay Library**

# Biodiversity Media



Yellow-throated Vireo collected by John James Audubon

# Biodiversity Media



Yellow-throated Vireo collected by John James Audubon



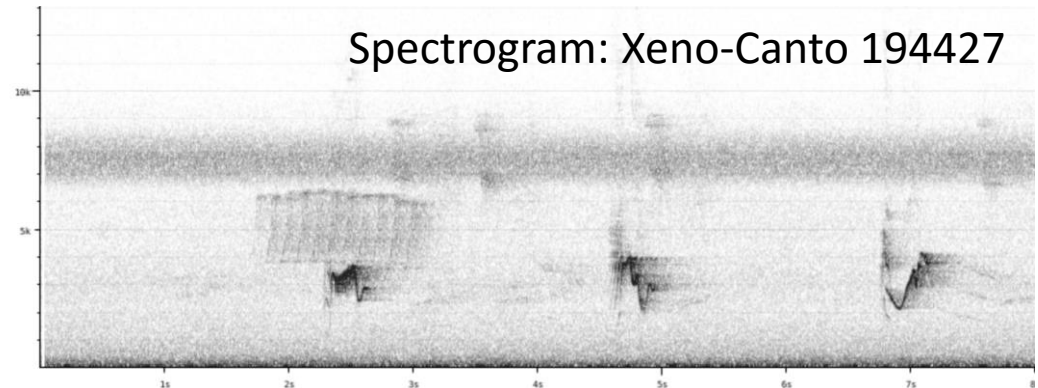
# Biodiversity Media



Yellow-throated Vireo collected by John James Audubon

Photo: nebirdsplus

Audio Recording: ML79436



# 5 Simple Principles

(it's about future-proofing the data)

1. Collect a lot of data (within reason)
2. Don't throw data away
3. Have a migration strategy
4. Spread your data around
5. Maintain data integrity

# 1. Collect a lot of data (within reason)



# 1. Collect a lot of data (within reason)



Audio: 96 kHz, 24-bit

Photos: Lots of pixels

Video: SD? HD? 4K? 16K?

← Good to use color standard and calibrated camera (?)

## 2. Don't throw data away



Audio: 96 kHz, 24-bit

Photos: Lots of pixels

Video: SD? HD? 4K? 16K?

### File Formats\*

Audio: .wav or BWF, maybe AIFF  
(not mp3)

Photos: Raw(?), DNG, TIFF, or  
JPEG2000 (not jpeg)

Video: ??? (MP4 or MPEG)

### \*Rules of thumb:

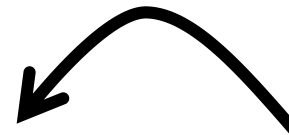
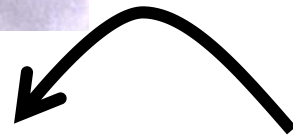
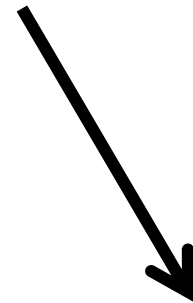
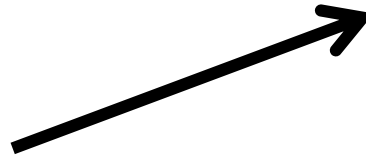
(1) Lossless is good

(2) Nonproprietary is great

(3) Widely accepted if best



# 3. Have a migration strategy



### 3. Have a migration strategy



Not even talking about this part...

General rule: Garbage in, garbage out

Think about: Media format & condition, playback hardware & calibration, A/D converter quality, etc, etc

Good to partner if possible...

# 3. Have a migration strategy

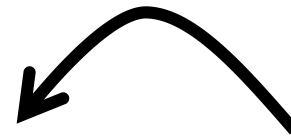
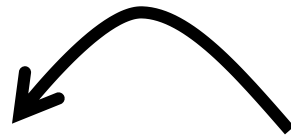
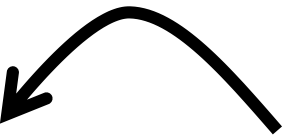
- Inevitable changes in format, carrier or other technologies
- Choose format that is widely-accepted industry standard
- Preferably not proprietary

## File Formats\*

Audio: .wav or **BWF** (not mp3), maybe AIFF

Photos: Raw(?), **DNG**, TIFF, or JPEG2000 (not jpeg)

Video: ???



## 4. Spread your data around

### ML Media Archive Storage

- Onsite HDD SAN utilizing RAID 5 redundancy architecture
- Offsite campus HDD mirror
- LTFS tape backup (2 copies)

# 5. Pay attention to data integrity

```
0001110100100100111110000111000101001010001100...
0001110100100100111110000111000101001010001100...
0001110100100100111110000111000101001010001100...
0001110100100100111110000110000101001010001100...
0001110100100100111110000110000101001010001100...
```

**Point:** Redundant copies and backup systems are without value unless one can confirm that the backup copies are true surrogates of the original.

Data integrity verification through use of hash algorithm (e.g., Digest algorithm 5 (MD5))

# Assumptions for the future

Support of analog technology will become increasingly challenging

Storage costs will continue to fall

New archival formats will evolve

# 5 Simple Principles

(it's about future-proofing the data)

1. Collect a lot of data (within reason)
2. Don't throw data away
3. Have a migration strategy
4. Spread your data around
5. Maintain data integrity

Some advice: Seek help