SU LAIR

Stanford University LIBRARIES &
ACADEMIC INFORMATION RESOURCES

QuickTime Uncompressed 10-bit 4:2:2 for video archiving

Hannah Frost
Manager, Stanford Media Preservation Lab

AMIA - IASA
Philadelphia • November 5, 2010

SULAIR

Deciding on QuickTime

- Adopted at Stanford Libraries in 2007
- Reaffirmed in 2010

Why QuickTime Uncompressed 10-bit 4:2:2?
Why not something else?
What else?

SULAIR began using QuickTime opportunistically around 2007, and we recently reaffirmed that decision when building out an in-house lab for video digitization. In this presentation, I will go over our decision making process –
why QT, why not something else, what else are we doing in our preservation strategy? Decisions are never made in a vacuum; context is critical, so contextual factors that influenced Stanford's decision are framed in the discussion.

SUNLAIR

Year 2007: Background and Context

Monterey Jazz Festival Collection Project

- Funded by GRAMMY Foundation, NHPRC, and Save America's Treasures
- Digitally preserve 800 audio recordings
- Surplus funds enabled preservation of 250 hours of video

Stanford wanted to cease the practice of remastering analog originals to Betacam SP for preservation. The Monterey Jazz Festival collection project – which benefitted from a surplus of grant funding – presented an unusual opportunity to make the switch to file-based preservation reformatting.

In planning for the project specification, the first question we asked was: What do we want in an archival file format for video?

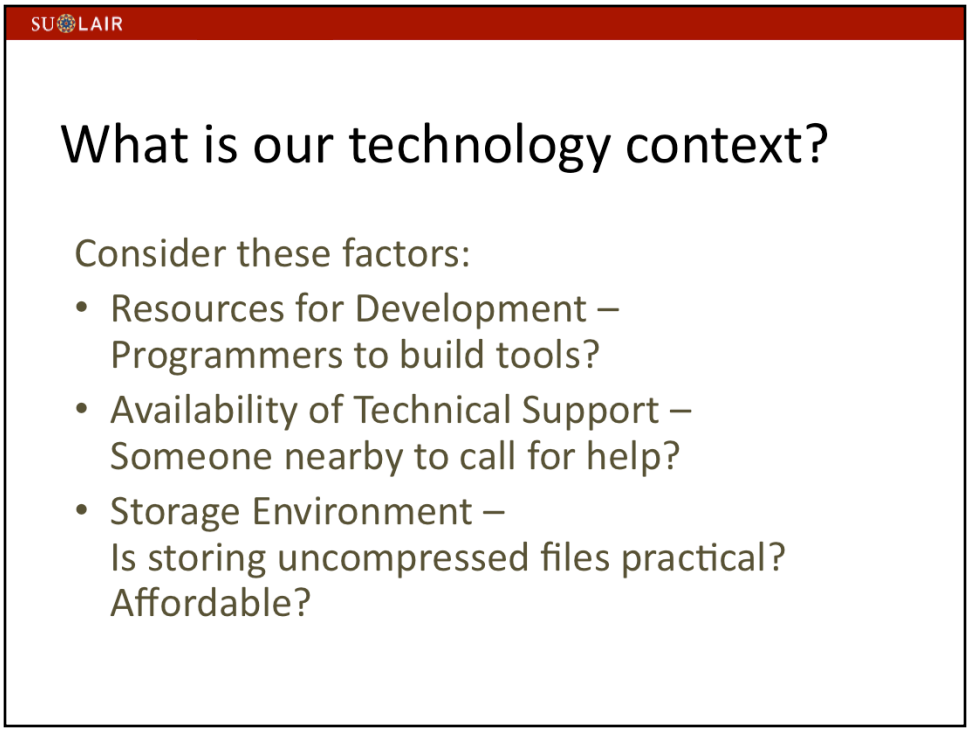
SUNLAIR

What does the user want?

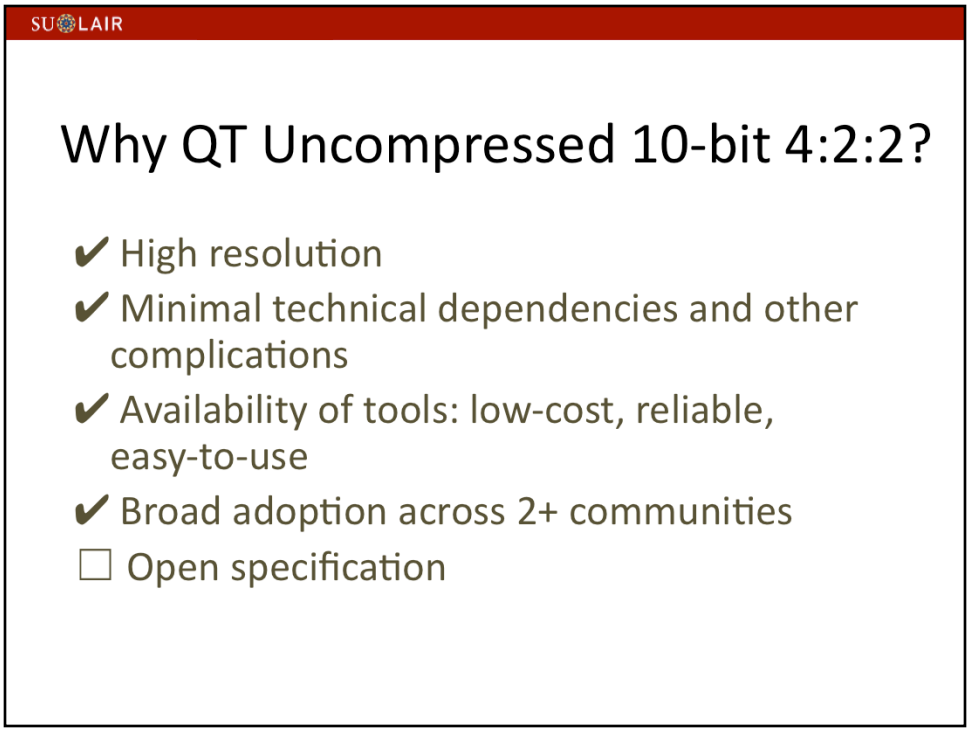
Hollywood director interested in using Monterey Jazz Festival material

- Primary requirements are:
  - High resolution
  - Good interoperability

Secondly: What format do users want?



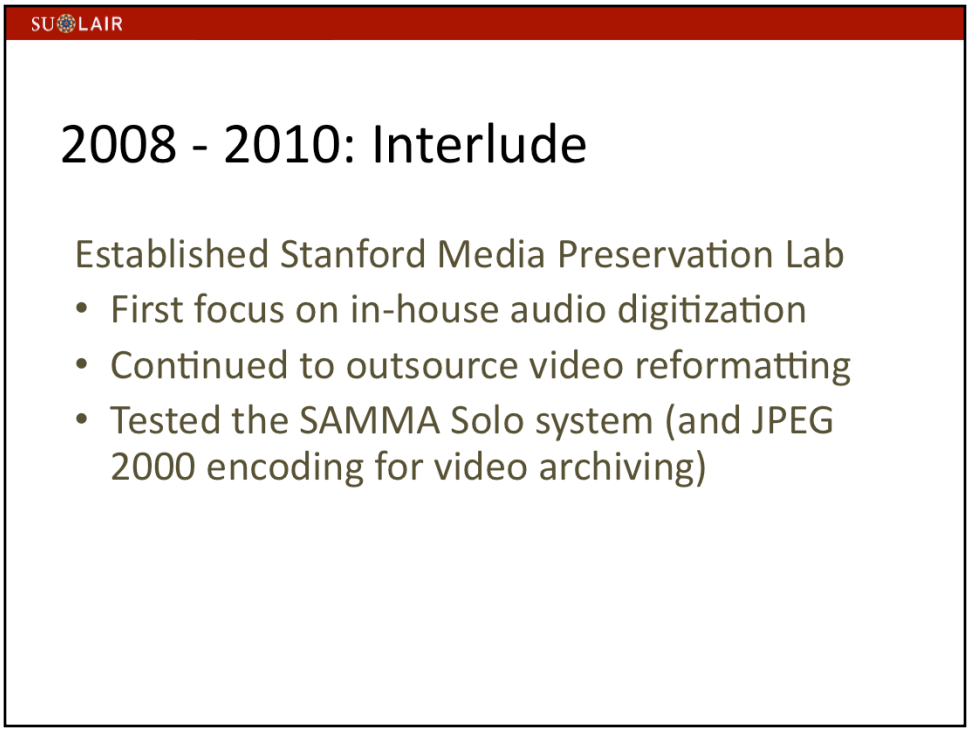
And finally, what is our technology context? Without ample development support for building and testing tools, it is not possible to be experimental. Stanford Libraries, once a PC-dominated environment, offered increasing technical support for Mac-based systems and tools. Would it be affordable to store uncompressed video files? Yes: As a partner in the Google Book Search Project, Stanford was already looking at storing millions of scanned books; availability of storage for a set of large video files was possible. And there are other contextual factors, most importantly broad organizational / philosophical support for digital video archiving.



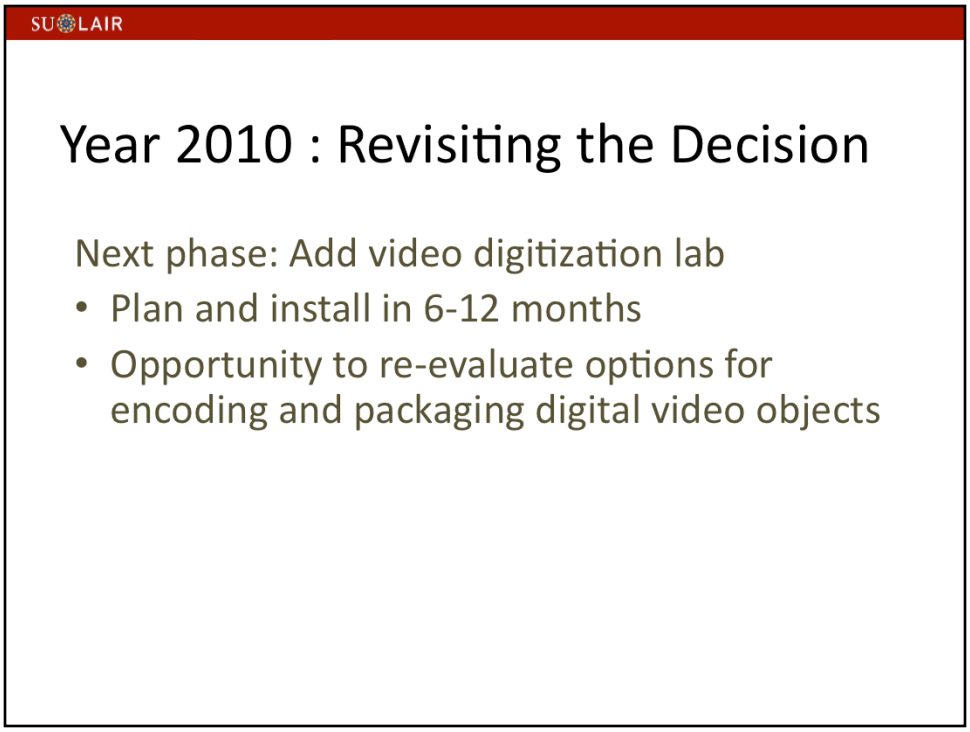
For the project, we chose QuickTime uncompressed 10-bit 4:2:2, as this particular codec/wrapper combination suited our requirements:

- 10-bit 4:2:2 is entirely adequate for capturing analog standard definition material
- Ample storage means we can capture highest resolution (no compression)
- Mac-support growing strong in the libraries
- Final Cut Pro, as a tool for working with QT files, is becoming ubiquitous;
- A strong group on the Stanford campus of video experts offered a rich local resource for working with video in the Mac environment.

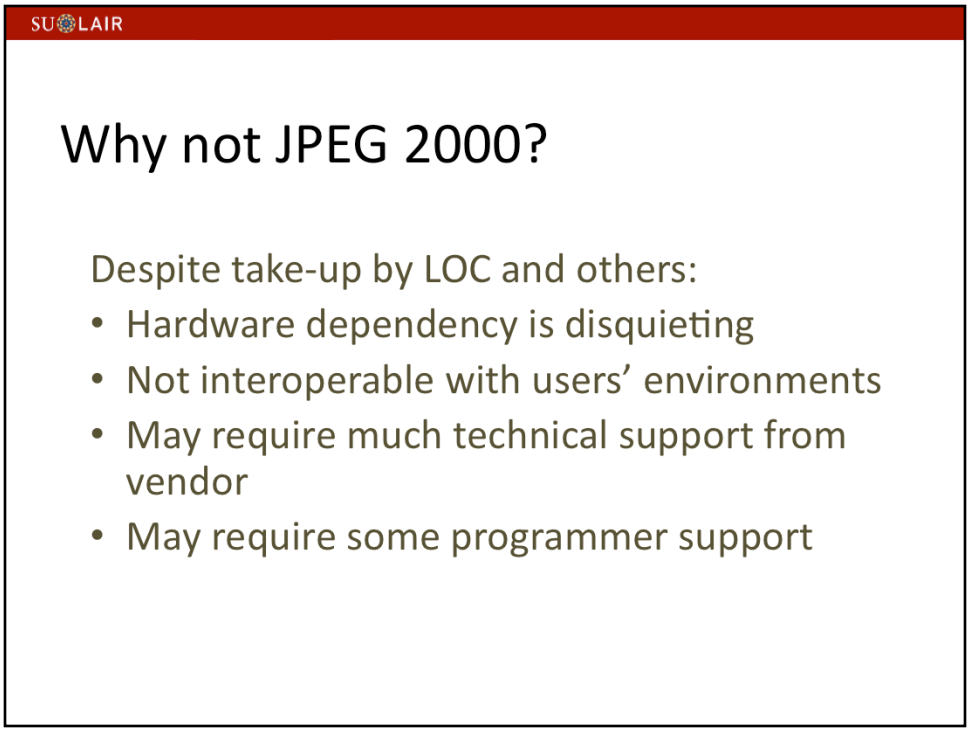
The project involved outsourcing the digitization to a vendor, and resulted in the production of about 25 TB of video files for the Monterey Jazz Festival collection.



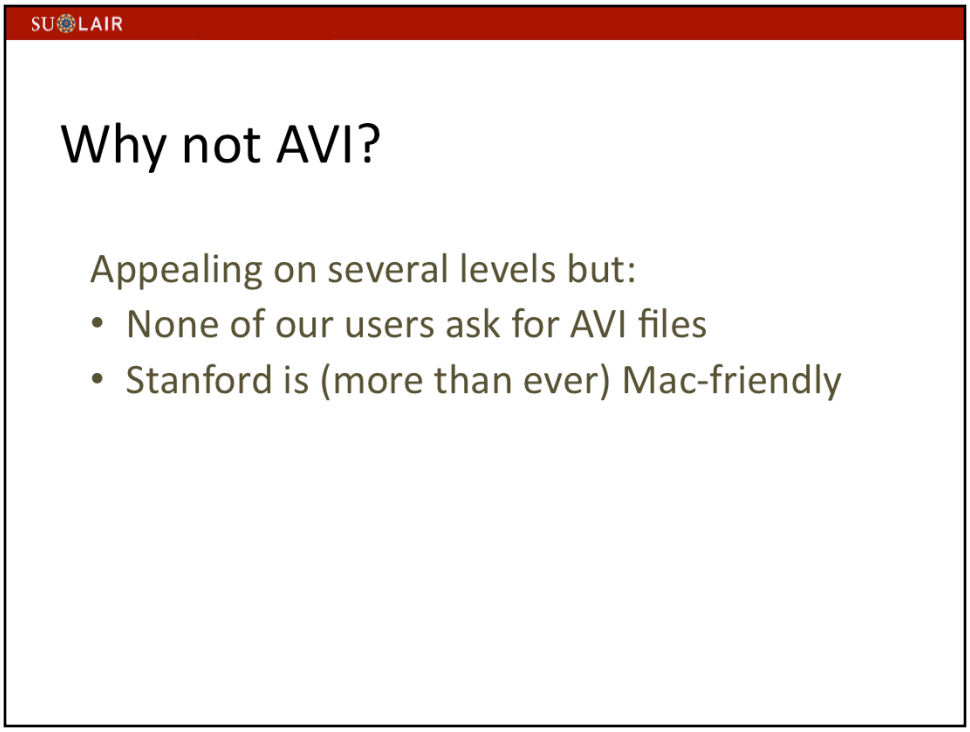
In 2008-2010, we spent two years establishing the Stanford Media Preservation Lab for preserving sound and moving image materials. The first phase of the lab development concentrated on audio digitization capabilities; during this time, we continued to outsource video reformatting to vendors using the same QuickTime spec. We had no issues with the files received from vendors as specified, and these files served to work well for the copy requests from researchers that we received. Also during this time, Stanford participated in a beta test of the SAMMA Solo system, so we had a chance to explore the use of JPEG 2000 encodings wrapped in MXF.



In late 2009, planning began for the Stanford Media Preservation Lab's expansion to handle video. It was time to reconsider the options for encoding video for preservation. We had had good luck so far with QT, and good reason to stick with specs that work. Maintaining consistency in file encoding specifications used will simplify long–term management of the content, so we were predisposed to continue use of QT. Nonetheless, we did step back and reconsider the options.

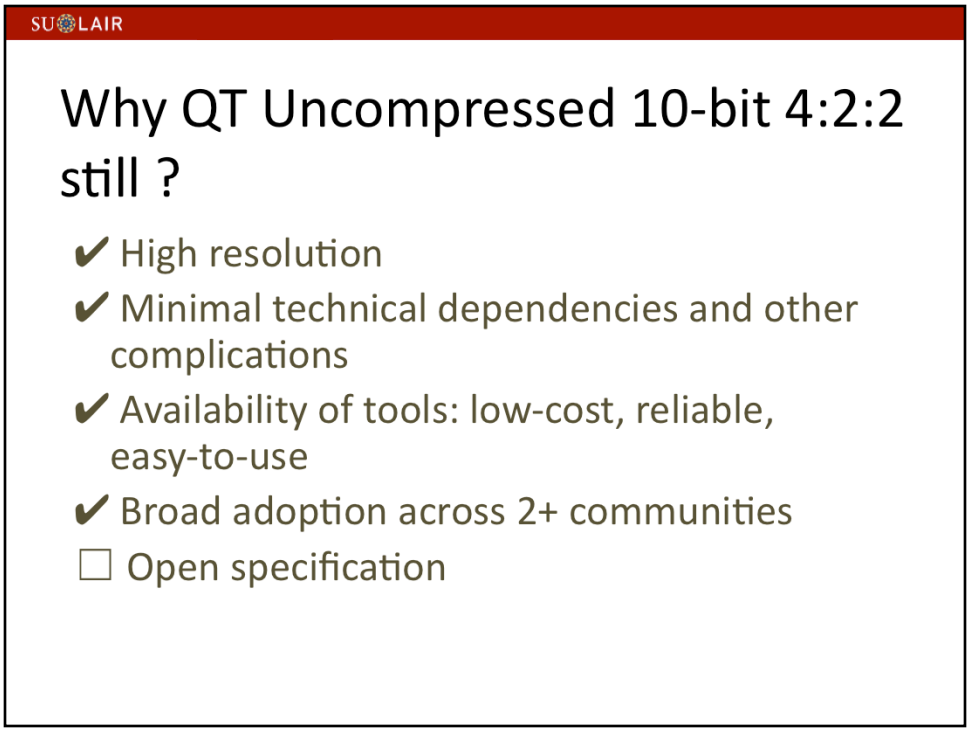


An obvious choice to consider was JPEG 2000 (as implemented in the SAMMA Systems products). On the pro side of the argument, it has gained some traction in the cultural heritage community through its adoption by the Library of Congress and others in the community, which certainly strengthens it as an option. Also, because it employs compression, less storage is required. Stanford is not completely averse to lossless compression in preservation masters: in fact, we are starting to archive JPEG 2000 image files for the books scanned by Google. On the other hand, the hardware dependency for playing back the SAMMA-produced files is disquieting, and the fact that the master JPEG 2000 files would not by interoperable with systems employed by users of the collections is also a concern. Nevertheless, we are still weighing the risk factors of employing compression for long-term management of original (unpublished) content, especially generated from sources that may not be able to be reformatted ever again in the future. Also, we anticipate that use of JPEG 2000 for video could require a fair amount of technical support from the vendor and ongoing local programmer support. We are concerned about down-time and productivity.



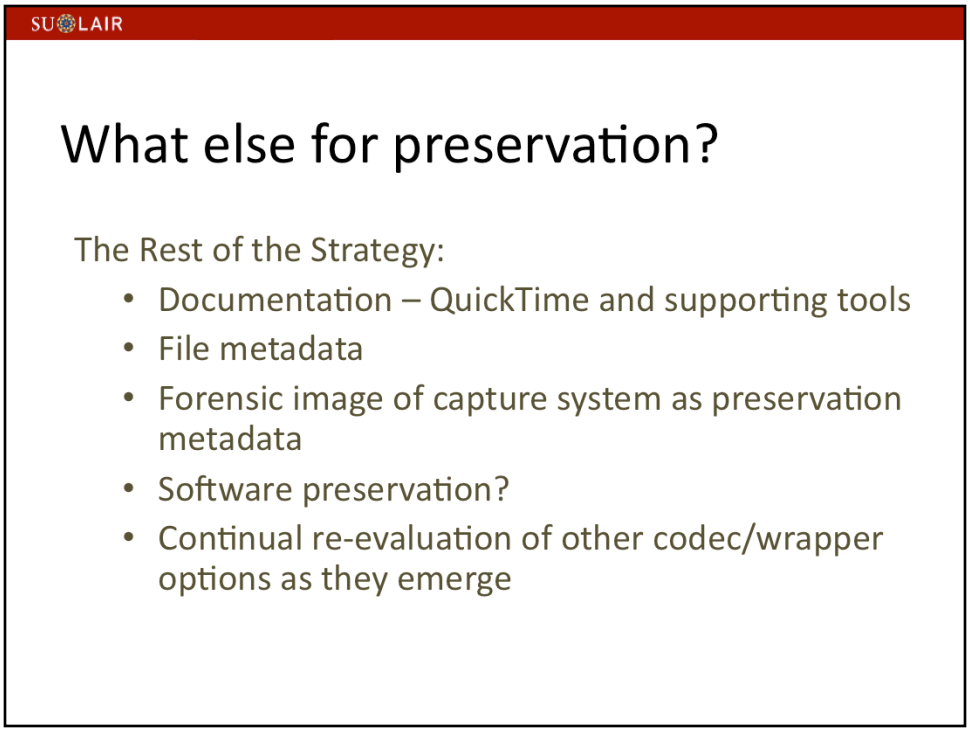
AVI is interesting, and has many things going for it, including the fact that it is based on the RIFF structure, like WAVE which we use for audio preservation. On the other hand, no end user has requested AVI (though presumably one could easily transcode from a QuickTime file. AVI dominates in the PC market (though it is supported in Mac systems and tools); at Stanford Libraries, Windows support remains available, but to an increasingly lesser degree.

What is our technical context now? Since 2007, the Digital Library Systems & Services division (where the Media Lab is located organizationally at Stanford Libraries) had grown and matured: there are greater developer resources, and the digital library infrastructure and its component systems are beginning to take shape, with a current emphasis on building digitization workflow tools for the range of digitization labs at Stanford. The greater Stanford community is greatly invested in Apple tools (though archiving efforts in the non-library production environments are not the same as in the research library environment). The anticipated abundance of storage is real. And the Media Lab is staffed, but it is noteworthy that it is staffed for production; there is not a lot of time for R&D and experimentation.

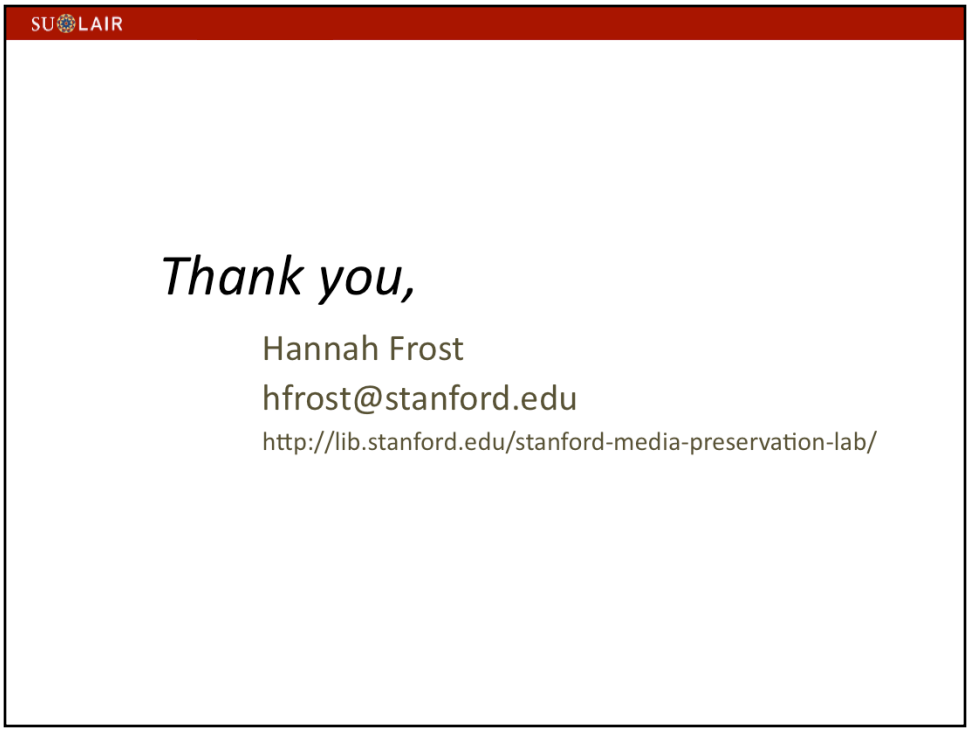


So we concluded that QuickTime still meets our requirements, but one: open specification. The concern for the proprietary nature of QuickTime is a matter of defending against a change in Apple's management of and support for QuickTime into the future. Apple has changed their technology fundamentals before -- remember resource forks? And remember when Microsoft dominated the technology market? Will future versions of tools drop compatibility for this codec? Unlikely, but possible. One piece of hope to offset the proprietary nature of QuickTime: it is the basis for MPEG-4 (ISO standard published circa 2001). This fact doesn't make the spec open, but it could strengthen support for it.

The decision to continue with QuickTime is never final; we can still change our mind after we observe how the market for open source formats/codecs/wrappers and supporting tools evolves. They will emerge, it is inevitable, and in particular we are hopeful about the work going on amongst the Federal Agencies these days with respect to MXF.



Additional aspects to our strategy include documentation of tools and formats we are using, technical metadata about the files and their creation environments. We may explore using forensic imaging of the capture systems as a means of documenting the tools and libraries present at the time of analog-to-digital conversion. And we will continue to re-evaluate this decision as other technology options emerge.



Please contact Hannah with any comments, suggestions, questions, and ideas!