No Time to Wait: Standardizing FFV1 and MKV for Preservation

Source: Ashley Blewer's report on the MediaConch website

Introduction

No Time to Wait!: Standardizing FFV1 & Matroska for Preservation was a symposium intentionally overlapping with Internet Engineering Task Force's 96th meeting, held in Berlin. No Time To Wait! was held on 17-19 July, 2016 and hosted by Deutsche Kinemathek, Zuse Institute Berlin, and MediaArea. The symposium was designed to bring together audiovisual archivists and audiovisual format designers with a focus on the standardization of a preservation-grade audiovisual file format combination package. The structure of this symposium was contingent heavily on the CELLAR working group and the initial meeting of this working group at IETF and first rounds of RFCs submitted to the organization.

Kieran O'Leary from the Irish Film Institute shows format conversion and colour space info.Photo credit: CC BY-SA Erwin Verbruggen.

Why these formats?

After introductions, the first talk of the symposium was from Erwin Verbruggen (Netherlands Sound & Vision), who gave a summary of the PREFORMA project. And with it, the insight and history into the decision-making behind the selection of these open formats and how they compared against other potential options to use in the development of a conformance checker for preservation-grade audiovisual formats. As obvious by this symposium, Matroska and FFV1 (and LPCM) were chosen by PREFORMA. The Internet Engineering Task Force working group, formed last year, adopted Matroska and FFV1 but chose to

focus on FLAC.

Steve Lhomme was able to attend the symposium and the IETF meeting, which is amazing because he is one of the founding developers of the Matroska format. His continual input on the CELLAR listserv and during the conference was absolutely invaluable. By the end of the symposium, Steve also had a thorough understanding of the unique needs of archivists and he was happy to assist in the required mapping work to ensure his format is suitable for this use case.

Fun fact: Steve original came up with the Matroska format because he was trying to catch Jacques Chirac, at the time President of France, lying on television. The origins of this format seem very archivally-minded, even if that context was not known or considered at the time.

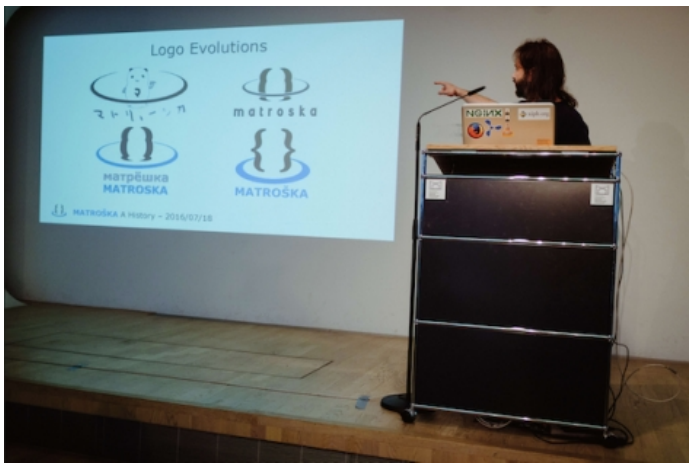
Peter Bubestinger gave a personal overview of the history of FFV1 as he sees it (and as it relates to archives at large and in specific, from his time working at Austrian Mediathek. When Peter was giving his talk on the history of FFV1, he made note that he was hesitant, despite FFV1 being an incredibly good idea to implement technically, of moving forward with it as an archival format because the specification was listed as ?experimental.? Eventually they decided that even with this marker, it was worth moving forward with FFV1 implementation for archival assets. He emailed a core developer of FFV1, Michael Niedermayer, and the developer told him that it was left on the website by accident and that the standard had been stable for at least 3 years at that time.

It was a treat to have Kate Murray from the United States Library of Congress join the symposium and give a talk on her many years of work developing AS-07, which focuses on but is not limited to the MXF format. This gave insight into the larger issues around standardization and conformance within the context of audiovisual preservation. She was kind enough to give us the ?lessons learned? from this work ? including the hazards of waiting until ?everything is perfect? before showing anyone publicly. She referenced the paper User Needs and MXF options for deciding if MXF is the right choice for one's institution.

Kieran O'Leary from the Irish Film Institute shows format conversion and colour space info.Photo credit: CC BY-SA Erwin Verbruggen.

How does this relate to IETF?

Tessa Fallon, co-chair of the CELLAR working group, gave a talk about the standardization track and the rules associate with the event. One afternoon of the symposium, held at the Zuse Institute Berlin, was dedicated to watching a livestream of the CELLAR working group. Steve Llome, Dave Rice, and Jerome Martinez attended the meeting, hosted by Tessa Fallon and Tim Terriberry. They were able to give updates on the work done thus far on the specification of Matroska and FFV1 and ask questions of the crowd. Afterwards, they were able to get feedback from other IETF members and tips on making progress collaboratively while working out the details.



Steve Lhomme presents Matroska logo history.Photo credit: CC BY-SA Erwin Verbruggen.

Leading Issues

There is a gap between the format designer work and the communication relayed to archivists, which is something this symposium sought to find an increased resolution or progress towards. The title of the symposium is apt; there is very little time when it comes to media formats that exist in crisis. There is an urgency that digitization begins as soon as possible, and as much as possible. Magnetic media formats are increasingly unable to be played back due to their inherently fragile nature. Moreso, the machines required to play back these formats are increasingly becoming obsolescent. The technicians with the ability to fix format players are increasingly less and less available too.

Daniel Borosa (Croatian Radiotelevision) gave a talk which covered the urgency of these formats, especially in the context of the Balkans region of Eastern Europe. He stressed that collaboration and assessment are important, but it is a challenge to acquire funding to be able to process this audiovisual material.

Overall funding was an important theme, especially during the preservation working group. The current work done on Matroska and FFV1 are thanks to the PREFORMA initiative, which is funding the MediaConch conformance checker for Matroska, FFV1, and LPCM. Some takeaways from the preservation working group and advocacy unconference group is to seek out local funding networks, and to do advocacy work not just outside of one's own institution but internally as well. Another loose topic was that of ?outreach,? which evolved a bit into the importance of collaboration among archivists. Several conversation streams and talks focused on the importance of being okay with writing bad code and releasing it publicly in order to get feedback and help from other archivists. It is important to be comfortable asking for and getting help. There is a desire for more technical workshops that empower archivists by giving them a situation in which it is ?safe to fail.?

Switching up a bit ? many of the talks and large portions of some of the meetings emphasized the importance of putting patrons first. Igor Wiedler gave a lightning talk on the ?morality of software,? emphasizing that software is never neutral, that all code written is inherently political, and that should be considered when writing any kind of software. Natalie Cadranel gave a brief lightning talk on working with OpenArchive, a mobile application that facilitates access to archives while also maintaining a user's privacy.

An issue coming heavily out of the FFV1 and preservation working groups, despite meeting separately, was the need for high-level understanding of the significance of FFV1 and why it is good for preservation. The technically-savvy people wanting to implement FFV1 can see the benefits ? but that may not be apparent to people in management-level positions. Two unconference groups worked on what the issues were surrounding FFV1's lack of attention in the archival community. One group came up with some guidelines to form an ?Executive Summary? which would convey the reasons why people are choosing FFV1 in a way that is easy to understand (and somewhat marketing-heavy). They concluded that FFV1 also needs a webpage explaining these features? and maybe a logo. Any takers?

Other issues: How to move the conversation forward for something perceived still as experimental (even if it isn't)? How does one convince IT staff that a ?market? solution isn't always the best? Investing in open source is a long-lasting commitment. There is a

problem with the ?free? in open source being considered ?free as in no money? rather than ?free as in freedom.? Hence the need and use of the acronym F(L)OSS, which stands for Free (Libre) Open Source Software, emphasizes that the ?free? mean ?freedom.?.

The unconference group focusing on Matroska's technical specs also focused on its ?weaknesses? which can be resolved through disambiguation while working on the RFC. One issue is timecodes ? Matroska's elements related to time codes are not necessarily in line with archivist's expectations. The CELLAR mailing list is very active regarding this issue.

Dave Rice and Peter Bubestinger showed the timeline of FFV1 development.Photo credit: CC BY-SA Erwin Verbruggen.
Cross-collaboration
So much of the intent of the symposium revolved around getting cultural heritage institutions, format designers, preservation-focused digitization vendors, and computer engineers in the same room and speaking the same language (or at least being able to come up with a shared pidgin-language). Michael Bradshaw, a YouTube engineer working on webm and its relation to its parent structure, Matroska, led a working group and an unconference group to dig into the technical details of the formats. Much of the data from his talk can be found on the YouTube Engineering Blog here. It was great to be able to link the preservation interests with web initiatives happening within large web-focused organizations like Google and Mozilla.

Tobias Rapp, NOA GmbH, gave a talk on his institution's research into what they recommend as vendors, and compared what he knew about AVI (the limitations and positive aspects) and what he knew about Matroska. A lot of the ?question marks? found in his slides were later resolved during the Matroska working group meeting

Kieran O'Leary's unconference group was focused on building or ensuring creation of tools that facilitate not just archivists but filmmakers actively making films so that if the original source can be something archive-worthy. This included namedropping a lot of different tools in their share-out summary: ?shotcut, vlc, natron, virtual dub, avi demux. VL MC Avid Sorenson Squeeze?. A current need is for someone to write an Adobe Premiere plugin that will allow the software to be able to import (read) and export (write) FFV1.

Group photo at the close of the No Time to Wait! symposium in Berlin.Photo credit: CC BY-SA Erwin Verbruggen.
Future of formats, next steps, future of CELLAR work
The symposium spurred conversations about the future of the specific formats, not just in their current specification forms but how they will comply with emerging media formats. How will we map METS/MODS or other metadata standards into Matroska tags? How will these formats deal with emerging audiovisual technology like 360* virtual reality video mappings, or mapping onto a plane or sphere? What kind of forethought can we have now to adequately comply with future standards, read future types of formats, and archive these formats?

Reto Kromer and Kieran O'Leary (Irish Film Institute) gave a talk on integrating FFV1 into film-scanning workflows and working with DPX, emphasizing that FFV1 may seem to be popular to video audiovisual formats but can comfortably handle film, too. There was interest in what small steps would be necessary to have film scans encode directly into FFV1 rather than having to be normalized later, after digitization work. Peter Bubestinger and Kieran Kunhya both spoke on ?stress testing? FFV1.

Peter's work can be found here and Kieran's work can be found here.

Work has and will continue to be active on the CELLAR mailing list.

Reminder: The CELLAR working group can be extremely technically specific, with conversations circling around specific details associated with Matroska and FFV1 and FLAC. However, plenty of work can be done for anyone with the time and willingness to help out. Feedback is crucial for the RFC (it is called Request for Comments for a reason) and even correcting typos within the specifications makes a real impact on the development of these formats. If you are not sure where to begin, don't hesitate to send an email to the CELLAR listserv asking what can be of most use or contact Ashley Blewer directly for help.

Hermann Lewetz from the Austrian Mediathek during the lightning talks.Photo credit: CC BY-SA Erwin Verbruggen.
Read more

- For another take on the symposium, read Erwin Verbruggen's report on the Sound and Vision blog]
- For more visuals and social chatter, see the Storify chronicling the symposium]
- All presentations will be added to the No Time to Wait repository on GitHub during the next few days]
- Background info on the PREFORMA projects is available on the open source portal and the MediaConch website]
- Revisit the PrestoCentre webinar Choose your file format from 2014 to hear Emanuel Lorain from PACKED speak about file format selection and Carl Feischhauer from Library of Congress (now retired) on the AS?07 standardisation process]
- For more background on the symposium title, No Time to Wait, see the blog post where it originated from on The Signal]
- For a list of broadcast-related open source tools, take a look at the EBU's ?amazingly awesome open source resources?]