

Digitaliseringshvidbog for DAL TVA - Dansk Lokal TV Arkiv

Digitalisering af analoge videobånd

Dansk Lokal TV Arkiv står overfor at skulle digitalisere ca. 3000 videobånd á gennemsnitligt 1 times varighed. Dette dokument er en opsummering af foretaget research omkring valg af formater til digitalisering af analoge videobånd til arkivering og publicering, samt beskrivelse af praktisk fremgangsmåde og heraf følgende behov for datalagerkapacitet til arkivering af de digitaliserede filer.

For at give læseren en mulighed for at vurdere dokumentets anbefalinger indeholder dokumentet overordnede beskrivelser af teknologier som transkodning i forbindelse med fremstilling af digitale filer, samt NAS og RAID-systemer til opbevaring af store datamængder. Disse beskrivelser er ikke fyldestgørende ud fra et teknisk fagligt standpunkt, men udelukkende ment som en vejledende introduktion af begreberne.

Dokumentet indeholder desuden vejledende prisoverslag for indkøb af harddiske, udstyrsliste til digitalisering og endeligt en gennemgang af forslag til navngivning af de producerede filer.

Overordnet problemstilling

Der stilles forskellige krav til et digitalt videoformat, afhængigt af om det er til arkivering eller publicering. Dette dokument identificerer derfor 2 formater til hhv. arkivering og publicering, samt et arbejdsformat der benyttes midlertidigt i processen. Disse 3 formater benævnes herefter:

- Digitaliseringsformat
- Arkivformat.
- Publiceringsformat.

Digitaliseringsformat / rå-fil: Udgangspunktet for digital arkivering af analoge videobånd er en konvertering af det analoge audio/video-signal til en ukomprimeret datastrøm i fuld opløsning, der digitaliseres så alle detaljer fra det analoge signal kan bearbejdes i det digitale domæne.

Den resulterende fil af en sådan digitalisering vil være uhensigtsmæssig stor. Formatet der digitaliseres i vil typisk være afhængigt af producenten af det benyttede digitaliseringsudstyr og kræve installation af den hardware og software der er blevet brugt under digitaliseringen for at filerne kan behandles. Af disse grunde er den digitaliserede rå-fil ikke optimal til arkivbrug.

Arkivformat: Formatet bør være interoperabelt så det er muligt at udveksle filer imellem forskellige platforme. Formatet bør gemme i så høj en kvalitet, at den resulterende fil indeholder alle informationer det er muligt at trække ud af de analoge bånd.

I tilfælde af komprimering bør dette ske "lossless" eller med et minimalt datatab. Moderne formater kan komprimere digitale filer uden praktisk kvalitetstab, hvilket betyder en ikke uvæsentlig pladsbesparelse ved arkivering, i forhold til arkivering af de ukomprimerede digitaliserede rå-filer.

Publiceringsformat: Formatet bør fylde så lidt som muligt og være alment afspilbart på alle gængse platforme. Pladskravet indebærer i praksis at filerne komprimeres, hvorfor der smides en del af de originale data væk. Denne komprimering gør publiceringsformater uegnet som arkivformat, da data som er smidt væk under komprimering, ikke kan genskabes i fremtiden.

Transkodning

Transkodning er en digital proces hvor man ændrer en fil fra et format til et andet. Som det fremgår af ovenstående, vil DAL TVA skulle arbejde i mindst 3 forskellige filformater.

I vores tilfælde skal den digitaliserede rå-fil transkodes til et arkivformat og til et publiceringsformat, hvilket medfører et behov for at kunne transkode filer imellem de 3 formater.

Transkodning fra et højopløsningsformat kan altid ske til et lavopløsningsformat uden kvalitetstab, men ikke omvendt. Ved at arkivere en højopløst arkivfil, kan der altid transkodes fra denne fil til nye formater, der i fremtiden måtte afløse de nuværende formater.

Transkodning foretages typisk ved at man eksporterer en fil fra et redigeringsprogram. Alternativt kan man bruge et selvstændigt stykke software til at transkode filer fra et format til et andet.

Fra teori til praksis

I det følgende beskrives hvordan DAL TVA i praksis vil opnå de ovenfor beskrevne mål:

- Udstyrslisten beskriver de komponenter DAL TVA har sammensat. Denne liste er kun vejledende og til inspiration. Andre kan sammensætte deres setup af andre komponenter.
- Step-by-step gennemgås processen med digitalisering, fra det analoge signal til færdige digitale filer.
- Efter grundig research har DAL TVA valgt de digitale filformater vi mener bedst tjener vores formål. I det følgende gennemgås de enkelte formater samt de overvejelser der har ført til valget af netop disse formater.

Udstyrsliste

Hardware

Grabber PC: HP Z800, XEON processor, 18GB RAM

Analog til SDI konverter: AJA HD 10 AVA

Grabberkort: Blackmagicdesign: Decklink SDI 4K

Editor controler: Blackmagicdesign Speed Editor

Lokal NAS: GNAP TS-410 – 4 bay NAS med 4 x Seagate Exos 18TB harddiske. I alt 48TB effektiv kapacitet

Software

Grabbersoftware og redigeringssoftware: Da Vinci Resolve Studio

Konverteringssoftware: FFmpeg

Praktisk digitalisering – step-by-step

1. Via analog-til-SDI konverteres det analoge signal til en digital datastrøm – et SDI-signal.
2. Denne datastrøm – SDI-signalet – digitaliseres i real-time på en PC med et SDI-grabberkort.
3. Den digitaliserede rå-fil transkodes herefter med programmet FFmpeg til hhv. et arkivformat og et publiceringsformat.
4. De fremstillede filer overføres til arkiv/publikations-servere. Herefter kan den digitaliserede rå-fil slettes.
5. I forbindelse med digitalisering indscannes/affotograferes fysiske indlæg og evt. båndrapporter for hvert bånd.

Digitale formater

Digitaliseringsformat: Ukomprimeret RGB 4:4:4 10 bit eller YUV 4:2:2 10 bit.

Arkivformat: Apple ProRes: Vores research har fastslået at det p.t. mest universelt anvendte og anerkendte arkivformat er "Apple ProRes 4444" for HD/UHD-materiale, og "ProRes 422 HQ" for SD-materiale. (HD = High Definition, UHD = Ultra High Definition(4K), SD = Standard Definition). ProRes bruges og anbefales af bl.a. DRs arkiver.

Da DALTVA's arkiver består af SD-materiale, er ProRes 422 HQ valgt som arkivformat. Dette format forstås af alle moderne redigeringsprogrammer i dag.

Apple ProRes er et proprietært format der er ejet af Apple og der er således ingen garanti for at det vil blive supporteret i al fremtid. Apple kan vælge at nedlægge supporten til formatet og nye programmer vil i så fald ikke kunne få licens til at bruge det. I skrivende stund er dette dog kun en teoretisk mulighed og sandsynligheden for at det sker i overskuelig fremtid er lille.

Alternativt arkivformat: FFV1.mkv: Dette format er et lossless Open Source format der er frit tilgængeligt og ikke er behæftet med licensbetaling. Formatet er udviklet for at lave et fælles format til udveksling af videomateriale, uafhængigt af enkelte softwareleverandører.

Formatet er endnu ikke endeligt standardiseret i internationale organer, og er derfor stadig at betegne som delvist eksperimentelt. Der er dog flere store internationale broadcastere samt det amerikanske kongresbibliotek (at sammenligne med vores rigsarkiv), der er begyndt at bruge formatet allerede.

DALTVA vil løbende følge med i udviklingen af dette format, da foreningen på sigt ønsker at arkivere i et åbent og producentuafhængigt format.

Publiceringsformat: MPEG4 H264: Dette format tilbyder en meget effektiv komprimering der gør filerne praktisk anvendelige til internetoverførsel / streaming, uden at den oplevede kvalitet forringes væsentligt.

Efterfølgeren H265 tilbyder endnu bedre kompression, men er endnu ikke så udbredt og byder derfor på udfordringer på mange af de afspillere den almindelige bruger har til rådighed i dag. DALTVA har valgt at publicere i H264 indtil H265 er mere udbredt.

Indscannings- / affotograferingsformater: .pdf og .jpg

Datahåndtering

Pladsforbrug: En times video i arkivformatet ProRes 422 HQ fylder ca. 28GB. 3000 timers video fylder ca. 80TB. En times video i H264 mp4 fylder ca. 13GB. 3000 timers video fylder ca. 38 TB.

Totalt pladskrav for 3000 timers video i arkivformat og publiceringsformat udgør således ca. 120TB.

Aktuelt er de største diske i handlen på 18TB. Reelt giver det 16TB anvendelig plads per disk.

RAID-systemer: Ved at samle flere harddiske i et RAID-system kan man opbygge en virtuel større harddisk, end ved brug af enkelte diske.

Et RAID-system kan konfigureres så det er det muligt at genskabe data fra en fejlet harddisk, ud fra data på de tilbageværende diske i RAID-systemet. For at kunne genskabe data fra en fejlet harddisk, gemmes der på de enkelte diske ikke kun de data den aktuelle disk indeholder, men også data om de data de andre diske i systemet indeholder. Dette reducerer den totale kapacitet af systemet.

Genetablering af data (regenerering) foregår i et RAID-system i praksis ved, at når systemet melder at en disk fejler, erstatter man fysisk den fejlede disk med en ny disk af samme type. Herefter regenererer systemet de data der var på den fejlede disk, ud fra data på de eksisterende diske. Ofte konfigurerer man RAID-systemer med en ekstra disk monteret rent fysisk i systemet (en såkaldt "spare-disk"), så man ikke fysisk behøver at skifte den fejlede disk ud med en ny, før man kan påbegynde regenerering, men kan styre dette på afstand og derfor ikke behøver haste afsted for at udskifte den fejlede disk.

Ved RAID6 kan op til to diske i systemet fejle uden datatab, forudsat man når at erstatte de fejlede diske og regenerere systemet, før flere diske fejler.

Praktisk regneeksempel: 10 stk. 18TB harddiske (anvendelig plads = 16TB per disk) i RAID6-konfiguration, giver 128TB diskplads (og ikke $10 \times 16\text{TB} = 160\text{TB}$), da en del af diskens samlede kapacitet som sagt bruges til at spejle data imellem diskene, så det er muligt at genskabe data i tilfælde af disknedbrud.

NAS: Et Network Attached Storage (NAS) er en server der er lavet til at indeholde flere harddiske som er konfigureret i et RAID-system. Et NAS tilgås af dets brugere via netværk. På de enkelte brugeres computer "mappes" et NAS-drev, der herefter optræder i brugerens liste af tilgængelige drev, så brugeren udover et C: drev og evt. andre lokale drev, også har f.eks. et X: drev der kan læses/skrives data fra/til.

Anbefalet system: Et NAS med 12 harddiskpladser med 11 stk. 18TB harddiske i RAID6 plus en spare-disk, giver en samlet kapacitet på 144TB og høj sikkerhed.

Datasikkerhed

RAID-systemer giver en stor grad af sikkerhed for datatab, skulle en harddisk fejle. RAID er dog ikke fejlfrit og i uheldige tilfælde kan man risikere at flere diske fejler, før man kan nå at regenerere alle data fra den disk der fejlede. RAID-systemer bør derfor ikke opfattes som en erstatning for back-up af data.

Man bør også tage højde for fysisk uoprettelig skade på det samlede RAID-system, f.eks. ved brand.

I tilfælde hvor man opbevarer uerstattelige data – som her hvor man ikke kan være sikker på at båndene kan digitaliseres igen hvis data går tabt – bør man derfor have data spejlet over (mindst) to fysiske lokationer, så man ikke mister alle data hvis ens servere ødelægges af f.eks. fejlstrøm ved lynnedslag eller i en brand.

Datalagre

Lokalt datalager: I forbindelse med digitaliseringsprocessen og den efterfølgende bearbejdning af de digitaliserede filer, skal der være adgang til et arbejdslager fra den eller de PC'er der foretager digitalisering og transkodning. Dette kan enten være lokalt på den enkelte digitaliserings PC (i form af interne harddiske) eller ved etablering af et lokalt NAS (Network Attached Storage) med RAID-system på arbejdsstedet.

Arkivlager, sikkerhed og backup: Det anbefales at der etableres to spejlede RAID-systemer med hver 11-datadiske og en sparedisk. I alt 24 harddiske á 18TB, med en total datakapacitet på 2 x 144TB i RAID6 konfiguration. Der bør derudover indkøbes ekstra erstatningsdiske der opbevares på de respektive lokationer, så de hurtigt kan isættes som erstatning i tilfælde af disknedbrud.

Det ene af disse systemer kan med fordel etableres i DALTVAs lokaler og fungere dels som lokalt data- og arbejdslager og dels som backup-server, hvorfra filerne replikeres til eksternt backup-arkiv og udgivelses-system.

Eksternt datalager – backup lokation: Der etableres et eksternt datalager som backup for det lokale lager, på en fysisk anden lokation.

Lagerpladsen etableres i form af et "ekspansionskabinet" af mærket QNAP UX-1200-U med 12 frie harddiskpladser tilkøbet en QNAP NAS-server.

Det vil være muligt at rumme det samlede Roskilde-arkiv på dette datalager, hvis det udbygges fuldt ud med 12 harddiske (se tidligere udregning).

Dette ekspansionskabinet vil kunne afkobles fra den eksterne NAS-server og tilkobles en anden kompatibel NAS-server (QNAP), uden at man skal flytte data fra diskene. Det er naturligvis også muligt at kopiere data fra dette datalager til et andet datalager, hvis man på et senere tidspunkt ønsker at benytte NAS-servere fra en anden producent.

Det eksterne datalager etableres i første omgang med 25TB. Lagerpladsen kan udvides efter behov ved indkøb af flere harddiske. Det eksterne datalager er tilsluttet internettet med en 1000/1000Mbit internetforbindelse.

Harddisk priser

Aktuel dagspris for en 18TB harddisk er ca. kr. 2.500,-

Diskforbrug til arkivering af Roskilde-arkivet udgør 2 RAID-systemer á 12 diske, plus 1 løs spare disk per system, plus 4 arbejdsdiske. I alt 30 diske.

Fuldt udbygget er DALTVA således oppe i en pris i størrelsesordenen 30x2.500,- = **kr. 75.000,-** for harddiske til arkivering af Roskilde-arkivet – hertil kommer NAS-servere der skal indeholde diskene.

Udgifter til NAS-servere er ikke omfattet af dette dokument, da disse er en del af DALTVAs overordnede projekt omkring arkivering af al dansk lokal-TV.

Det anses ikke for rimeligt at det overordnede projekt skal afholde udgifter der alene vedrører Roskilde-arkivet (harddiske til Roskilde-arkivet), hvorfor disse udgifter er beregnet selvstændigt og indeholdt i dette dokument.

Løbende omkostninger: Harddiske til NAS-brug har en gennemsnitlig levetid på 5-7 år. Der skal således indregnes en årlig driftsomkostning på ca. kr. 12.500,- med de nuværende priser (kr. 75.000,- over 6 år), til vedligehold af harddiske, når Roskilde-arkivet er fuldt udbygget og i drift.

Prisen på harddiske har dog historisk vist sig at falde drastisk i takt med den teknologiske udvikling, og der er ingen grund til at tro at dette ikke vil fortsætte. Det er således ikke urealistisk at samme lagerkapacitet kan erhverves for et væsentligt mindre beløb, når det bliver tid at udskifte diskene.

Uanset prisudviklingen, er løbende udgifter til udskiftning af diske med 5-7 års mellemrum en post der bør indgå i fremtidige langsigtede budgetter for den løbende drift af Dansk Lokal TV Arkiv.

Forfatteren af dette dokument har ikke foretaget en historisk research på prisudviklingen for harddiske pr. TB – den anslås dog at være omkring 50% på en 5-årig periode.

Navngivning og filstruktur

Alle bånd i DALTVA's båndarkiv er forsynet med båndnumre. Disse bør indgå som en del af filnavnet på de digitaliserede filer.

Udsendelserne på båndene har altid en titel – denne bør også indgå i filnavnet.

Udsendelsesbånd indeholder ofte flere programmer og dermed flere titler. Udsendelser i en programrække vil typisk være identificerbare i form af et ugenummer eller dato for udsendelse.

Eksempel: Bånd 23037 er udsendelsesbånd for uge 8 1987 og indeholder udsendelserne "Aktuelt kvarter" og "Rundt om Roskilde". Udsendelsesdatoen fremgår af programoversigt på båndet som 6 marts 1987.

Båndet digitaliseres i sin helhed og de heraf følgende filer navngives:

"23037 19870306 Uge 8 1987 – aktuelt kvarter – Rundt om Roskilde" plus fil-suffix (.mxf, .mp4, .pdf)

Klippes de enkelte udsendelser på et senere tidspunkt ud af masterfilen, navngives de *"23037 19870306 Aktuelt Kvarter"* hhv *"23037 19870306 Rundt om Roskilde"*.

Bemærk amerikansk datoformat, der betyder at filerne sorteres forløbende efter år, måned og dag.

Det vil ikke altid være muligt at identificere den oprindelige udsendelsesdato.

Det skal overvejes om der skal oprettes en undermappe for hvert enkelt bånd, navngivet efter masterfilen, indeholdende alle filformater og scanninger af det aktuelle bånd.

Et eksempel herpå kunne være:

Navn på undermappe

23037 19870306 Uge 8 1987 – aktuelt kvarter – Rundt om Roskilde

Filer i undermappe

23037 19870306 Uge 8 1987 – aktuelt kvarter – Rundt om Roskilde.mxf

23037 19870306 Uge 8 1987 – aktuelt kvarter – Rundt om Roskilde.mp4

23037 19870306 Uge 8 1987 – aktuelt kvarter – Rundt om Roskilde_scan1.pdf

23037 19870306 Uge 8 1987 – aktuelt kvarter – Rundt om Roskilde_scan2.pdf

23037 19870306 Uge 8 1987 – aktuelt kvarter.mxf

23037 19870306 Uge 8 1987 – aktuelt kvarter.mp4

23037 19870306 Uge 8 1987 – Rundt om Roskilde.mxf

23037 19870306 Uge 8 1987 – Rundt om Roskilde.mp4

Bemærk: Ovenstående er kun et forslag og oplæg til videre diskussion. Det anbefales at der inddrages bibliotekskyndige fagpersoner og databasekyndig programmør, til vejledning i navngivning og filstruktur, før DALTVA ligger sig endeligt fast på et navngivningsformat og en filstruktur.

Odsherred 7/1 2022

/Erik Lange

Referencer:

Apple Prores:

https://en.wikipedia.org/wiki/Apple_ProRes

FFV1:

<https://en.wikipedia.org/wiki/FFV1>

Arkivformater generelt:

https://www.scart.be/?q=en%2Fcontent%2Fshort-guide-choosing-digital-format-video-archiving-masters&fbclid=IwAR3GbleZrYECDYTgXtHAZf4_SUpydFZSvp5xT9XnwLYTwr2COVJ19KMV5c

H264 MPEG4:

https://en.wikipedia.org/wiki/H.264/MPEG-4_AVC_products_and_implementations

DR tekniske standarder (ENG):

https://www.dr.dk/om-dr/eksterne-partnere/eksterne-producenter/retningslinjer-vejledninger-og-skabeloner/tekniske?fbclid=IwAR1qt5_RYmsgXOu4sn_XFtekW_H4cmsXrVUmV-9_1kbhyJzVtt6koGteF7Y

DR tekniske standarder (DK):

https://www.dr.dk/static/documents/2021/06/07/tekniske_standarder_-_august_2020_7683b630_7f78647d.pdf

PRORES OR: HOW WE LEARNED TO STOP WORRYING AND LOVE COMPRESSION:

<https://gammaraydigital.com/blog/prores-or-how-we-learned-stop-worrying-and-love-compression>

Video space calculator:

<https://www.digitalrebellion.com/webapps/videocalc>

RAID-calculator:

<http://www.raid-calculator.com/default.aspx>

RAID-systemer:

<https://en.wikipedia.org/wiki/RAID>

NAS – Network Attached Storage:

https://en.wikipedia.org/wiki/Network-attached_storage

Analog til SDI konverter: AJA HD 10 AVA:

https://www.aja.com/products/hd10ava?gclid=EAlaIqobChMI9vzx7Kfo9AIVhkhGAB1XbAnUEAAYASAAEgKs8vD_BwE

Blackmagicdesign Decklink 4k SDI:

<https://www.blackmagicdesign.com/products/decklink/techspecs/W-DLK-11>

Blackmagicdesign Speed Editor:

<https://www.blackmagicdesign.com/products/davinciresolve/>

Da Vinci Resolve Studio:

<https://www.blackmagicdesign.com/products/davinciresolve/>

FFmpeg konvertering til ProRes 422 HQ:

<https://ottverse.com/ffmpeg-convert-to-apple-prores-422-4444-hq/>

ProRes dataforbrug:

Dimensions	Frame Rate	ProRes 422 (Proxy)		ProRes 422 (LT)		ProRes 422		ProRes 422 (HQ)		ProRes 4444 (no alpha)	
		Mb/s	GB/hr	Mb/s	GB/hr	Mb/s	GB/hr	Mb/s	GB/hr	Mb/s	GB/hr
720 x 486	24p	10	4	23	10	34	15	50	23	75	34
	60i, 30p	12	5	29	13	42	19	63	28	94	42
720 x 576	50i, 25p	12	6	28	13	41	18	61	28	92	41
960 x 720	24p	15	7	35	16	50	23	75	34	113	51
	25p	16	7	36	16	52	24	79	35	118	53
	30p	19	9	44	20	63	28	94	42	141	64
	50p	32	14	73	33	105	47	157	71	236	106
	60p	38	17	87	39	126	57	189	85	283	127
1280 x 720	24p	18	8	41	18	59	26	88	40	132	59
	25p	19	9	42	19	61	28	92	41	138	62
	30p	23	10	51	23	73	33	110	49	165	74
	50p	38	17	84	38	122	55	184	83	275	124
	60p	45	20	101	46	147	66	220	99	330	148
1280 x 1080	24p	31	14	70	31	101	45	151	68	226	102
	60i, 30p	38	17	87	39	126	57	189	85	283	127
1440 x 1080	24p	31	14	70	31	101	45	151	68	226	102
	50i, 25p	32	14	73	33	105	47	157	71	236	106
	60i, 30p	38	17	87	39	126	57	189	85	283	127
1920 x 1080	24p	36	16	82	37	117	53	176	79	264	119
	50i, 25p	38	17	85	38	122	55	184	83	275	124
	60i, 30p	45	20	102	46	147	66	220	99	330	148
	50p	76	34	170	77	245	110	367	165	551	248
	60p	91	41	204	92	293	132	440	198	660	297