DET KGL. BIBLIOTEK
Royal Danish Library

9. januar 2019
J.nr. 19/00133

Det Kgl. Biblioteks
Strategi for digital bevaring

Indhold

Bilag

Indledning

Strategi for digital bevaring på Det Kgl. Bibliotek beskriver arbejdet med digital bevaring og de visioner og målsætninger, der er væsentlige for dette arbejde. Strategien er første strategi efter fusionen mellem Statsbiblioteket og Det Kongelige Bibliotek og er dermed en fusionsstrategi, hvorfor den ret grundigt beskriver forskellige tekniske forudsætninger for arbejdet med digital bevaring.

Relation til andre dokumenter

Strategien er en del af et kompleks af politikker, strategier og retningslinjer for arbejdet med det digitale materiale på Det Kgl. Bibliotek. Dette omfatter således ikke kun dokumenter, der relaterer sig direkte til digital bevaring, men også fx både Accessionspolitik og Adgangspolitik. Tilsammen skal dokumenterne ses som en større helhed, der viser Det Kgl. Biblioteks arbejde med det digitale materiale.

Strategien er direkte relateret til Det Kgl. Biblioteks Politik for digital bevaring og udmønter de rammer og principper, der udstikkes i politikken, som aktive beslutninger om bevaringsrelaterede aktioner og procedurer.

Hvad strategien omfatter

Strategien omfatter alle Det Kgl. Biblioteks digitale kulturarvssamlinger som beskrevet i bevaringspolitikken og relaterede politikker.

Bevaringsarbejdet med de digitale kulturarvssamlinger udgør bibliotekets digitale repository, hvilket vil sige, at strategien omfatter både organisation og økonomi, relevante it-systemer og tilhørende software, og selve de digitale objekter bestående af data og metadata.

Strategien opstiller en række visioner og mål, som rækker op til fem år ud i fremtiden. Strategien skal dog revideres hvert tredje år for at sikre, at der tages højde for de eksterne og interne forandringer, som uvilkårligt vil komme til at ske.

Formål og vision

Digital bevaring er en nødvendig forudsætning for at kunne give adgang til digitale samlinger på lang sigt. Udføres der ikke digital bevaring, vil de digitale materialer over tid forgå, enten på grund af fejl i filerne eller manglende software til fortolkning af filerne.

Formålet med Strategien er at synliggøre og skabe overblik over Det Kgl. Biblioteks arbejde med og målsætninger for digital bevaring. Samtidig kan strategien fungere som reference for det daglige arbejde med digital bevaring.

Visionen er at sikre bibliotekets digitale materialer bedst muligt, så også fremtidens brugere vil kunne få glæde af dem.

Baggrund

Det Kgl. Bibliotek opstod i 2017 som en fusion af Statsbiblioteket og Det Kongelige Bibliotek, som hver især har arbejdet med digital bevaring gennem mange år - på nogle punkter hver for sig, på nogle punkter i fællesskab. Ved fusionen opstod behovet for en fælles tilgang til digital bevaring, hvorfor en arbejdsgruppe på tværs af de to tidligere organisationer har lavet en ny politik og strategi for digital bevaring.

En af bibliotekets fornemmeste opgaver er at sikre, at også eftertiden har mulighed for at tilgå kulturarven. Dette er ikke blevet lettere med årene og de mange digitale muligheder giver både meget mere materiale og sværere arbejdsbetingelser for håndteringen.

Den nuværende strategi er lavet på baggrund af tidligere strategier for hhv. Statsbiblioteket og Det Kongelige Bibliotek samt Netarkivet, som er et fælles projekt, der har kørt i en årrække. Med den nuværende strategi ensartes arbejdet med digital bevaring på tværs af det tidligere organisatoriske skel.

Strategiske mål

1. **I 2021 er alle pligtafleverede data, som er modtaget før 2020, under bevaring**
   Dette omfatter bitbevaring, overvågning og fyldestgørende bevaringsplaner

2. **I 2021 er der procedurer for det daglige arbejde med digital bevaring**
   Dette omfatter alt fra bestemmelse af indhold i en bevaringsplan, igangsætning af aktiviteter til sikring af fremtidig tilgang og løbende teknologiovervågning og risikostyring

3. **I 2023 opfylder al bitbevaring krav til geografisk, organisatorisk og softwaremæssig uafhængighed**
   Dette omfatter også etablering af en udenlandsk kopi for relevante materialer

4. **Det Kgl. Bibliotek fastholder løbende sin status som aktiv bidragsyder inden for digital bevaring både nationalt og internationalt**
   Dette indebærer, at Det Kgl. Bibliotek opretholder sine resultater såvel forskningsmæssigt som ved deltagelse i forskellige netværk og standardiseringsarbejde og dermed bevarer sin status som en anerkendt samarbejdspartner

Risikostyring

Vision:

- Risikostyring skal være et dagligt redskab for alle medarbejdere, som har ansvar for digitale samlinger, digital bevaring og de it-systemer, som håndterer data og metadata.

Mål:

- I 2019 udarbejdes en risikovurdering af det digitale repository
- Inden udgangen af 2023 udarbejdes en risikovurdering for alle pligtafleverede samlinger
- Det Kgl. Bibliotek vil løbende sikre, at alle relevante medarbejdere får den nødvendige kompetenceudvikling i risikostyring

Risikostyring omfatter:

- Risikovurdering: Identificering, registrering, analyse og evaluering af de potentielle risici, der truer bibliotekets digitale arkiv. Vurderingen skal revideres løbende for at identificere ændringer i arkivets rammer og vilkår (tekniske, organisatoriske, politiske eller juridiske), som kan ændre det samlede risikobillede.
- Risikohåndtering: Udarbejdelse af plan for håndtering af risici og igangsættelse af relevante aktiviteter, som kan minimere de negative konsekvenser ved aktuelle risici.
- Risikoaccept: Fastlæggelse af kriterier for risikotolerance

Risikostyring er Det Kgl. Biblioteks primære metode til bevaringsplanlægning. Risikostyring sker i samarbejde med Det Kgl. Biblioteks sikringsudvalg, og risikovurderingen skal udarbejdes med reference til audit- og certificeringsstandard ISO 16363 og standarden om informationssikkerhed, ISO 27001.

Risikostyring foregår for Det Kgl. Biblioteks digitale repository som helhed, og omfatter både bitbevaringsløsningen, resource- og metadatahåndterings-systemer og de platforme, der giver direkte adgang til det bevarede materiale samt de organisatoriske og økonomiske rammer. Det samlede risikobillede for repository kræver derfor input og tværgående arbejde fra organisatoriske enheder og medarbejdere med forskellige kompetencer, i områder som omfatter it-drift, it-fagsystemer, samlingsspecifikt kendskab og digital bevaring samt kompetencer til at evaluere juridiske, politiske, og økonomiske rammer og risici.

Teknologiovervågning

Mål:

- I 2019 er der fastlagt procedurer for, hvordan teknologiovervågning gennemføres internt på Det Kgl. Bibliotek
- I 2020 er det fastlagt, hvordan og i hvilket omfang teknologiovervågningen kan foretages ved hjælp af internationalt samarbejde

Det Kgl. Bibliotek foretager løbende overvågning af en lang række aktiviteter og udviklingsområder for på den måde at imødegå trusler mod dets digitale samlinger.

På det overordnede plan overvåger Det Kgl. Bibliotek løbende dataformater, metadataformater, softwaresystemer relevante for digital bevaring, auditstandarder, samt andre standarder, der er relevante for digital bevaring.

I forbindelse med bitbevaring overvåger Det Kgl. Bibliotek eventuelle trusler fra den hardware og software, som anvendes i de tilhørende systemer, samt organisering af bitbevaringsarbejdet.

Når det gælder teknologiovervågning for logisk bevaring vil Det Kgl. Bibliotek, både internt og i samarbejde med sine internationale partnere, etablere en løbende overvågning. Det vil ske både på formatniveau og på et mere overordnet niveau for fx digitale personarkiver og software. Et eksempel på en trussel for software er, at et computerspil ikke længere kan afvikles på moderne platforme. Teknologiovervågningen foregår både i forhold til de materialer, som findes i samlingerne, og til de materialer, som indlemmes.

Det Kgl. Bibliotek vil foretage en løbende teknologiovervågning af eventuelle trusler mod sekundært materiale (software, licenser, dokumentation, mv.), der skal anvendes i forbindelse med brugen af emulering som logisk bevaringsstrategi.

Udover ovenstående områder vil Det Kgl. Bibliotek også foretage løbende overvågning inden for en lang række andre områder, blandt andet:

- Standarder
- Metadataformater
- Karakteriseringssoftware
- Valideringssoftware
- Migreringssoftware
- Systemer til understøttelse af bevaring

Generelt vil Det Kgl. Bibliotek overvåge alle trusler - store som små - mod dets digitale samlinger. Når truslerne er identificeret, vil de udløse en tilhørende risikohåndtering, ud fra hvilken det videre forløb besluttes.

Standarder

Vision:

- Det Kgl. Bibliotek benytter i vides muligt omfang internationalt anerkendte standarder i alle dele af arbejdet med digital bevaring
- Det Kgl. Bibliotek bidrager til udvikling af standarder, som er relevante for bibliotekets arbejde inden for digital bevaring

Mål:

- I 2019 udarbejdes et samlet overblik over og beskrivelser af de mest anvendte standarder

Det Kgl. Bibliotek evaluerer løbende internationalt anerkendte standarder inden for digital bevaring for at kunne drage fordel af og anvende de bedst egnede i det daglige arbejde med digital bevaring. Ved en evaluering ses bl.a. på, hvor gennemskuelig standarden er, om der er et community omkring den, som gør den udbredt inden for digital bevaring, og om der er tilknyttet brugbare værktøjer. Brugen af standarder inden for både værktøjer, metadata, dataudveksling mv. bidrager til at ensarte og dokumentere det bevarede materiale og forenkle procedurer og arbejdsgange. Dermed øges gennemskueligheden og troværdigheden for brugeren, og samtidig gives større sikkerhed for, at det bevarede digitale materiale også vil kunne forstås i fremtiden.

De følgende standarder er med til at sætte rammerne for al digital bevaring på Det Kgl. Bibliotek:

- ISO 14721:2012 (2012): *Space data and information transfer systems -- Open Archival Information System (OAIS) – Reference model*
- ISO 16363:2012 (2012): *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*

Af andre standarder, som er vigtige for arbejdet med digital bevaring, kan nævnes

- ISO/IEC 27001:2013: *Information technology -- Security techniques -- Information security management systems -- Requirements*
- ISO 690:2010 (2010): *Information and documentation -- Guidelines for bibliographic references and citations to information resources*
- ISO 28500:2017: *Information and documentation – WARC file format*
- *PREMIS Data Dictionary for Preservation Metadata*

Biblioteket søger at opnå indflydelse på den fremtidige udvikling ved selv at deltage i standardiseringsarbejdet for flere af disse standarder, herunder OAIS, WARC, PREMIS, ISO 690 om bibliografiske referencer og PWID URN for webarkiv referencer.

Bevaringsplaner

Vision:

- Det Kgl. Bibliotek sikrer, at der altid ligger opdaterede bevaringsplaner for alle digitale materialer

Mål:

- I løbet af 2019 er lavet en skabelon for bevaringsplaner
- I løbet af 2019 er oprettet et system til håndtering af bevaringsplaner og samlingsbeskrivelser
- I løbet af 2021 er der udarbejdet eller revideret bevaringsplaner for alle pligtafleverede materialer indlemmet før 2020

For at sikre dokumentation og overblik over den digitale bevaring på Det Kgl. Bibliotek oprettes der for alle digitale materialer en bevaringsplan. En bevaringsplan kan dække én samling, flere samlinger eller en eller flere delsamlinger, så længe materialerne har samme bevaringsniveau (bitsikkerheds- og fortrolighedsniveau) og karakteristika i forhold til, hvordan de skal bevares.

Bevaringsplaner er et vigtigt værktøj for det daglige arbejde med digital bevaring og indeholder oplysninger om bl.a. materialernes tilstand, bevaringsniveau, risikobillede og handlingsplan for det fremadrettede arbejde. Der arbejdes i øjeblikket (2018) på at harmonisere bevaringsplaner over hele Det Kgl. Bibliotek, så de formmæssigt er ens og lettere kan bruges sammen med samlingsbeskrivelser på tværs af organisationen. Bevaringsplanerne udarbejdes i tæt samarbejde mellem de ansvarlige for digital bevaring og de samlingsansvarlige.

Åben og proprietær software

Vision:

- Det Kgl. Bibliotek tilstræber at fastholde kompetencer i at forstå, anvende og videreudvikle åben software
- Det Kgl. Bibliotek bidrager til fortsat udvikling af relevante open source-produkter

Det Kgl. Bibliotek ønsker at have kontrol med sine løsninger til digital bevaring. Biblioteket ønsker at anvende software og løsninger, der er baseret på åbne, standardiserede teknologier med klar evidens for løsningernes kvalitet, fx i form af dokumentation, test coverage mv. Biblioteket benytter løsninger, der minimerer risikoen for vendor lock-in, sådan at vi kan være sikre på, at vi efter behov kan flytte data til nye bevaringsløsninger.

For at sikre bevaringsindsatsen er det vigtigt, at den software, der anvendes, er åben og gennemskuelig og giver mulighed for justeringer og indpasninger af komponenter udviklet af os selv eller andre. Derfor anvender institutionen i videst muligt omfang åben software (open source) og software, der understøttes og anvendes af et bredt, internationalt fællesskab af institutioner, der er sammenlignelige med Det Kgl. Bibliotek.

På en række områder findes der ikke åben software, der ressource-, sikkerheds- og funktionalitetsmæssigt kan måle sig med kommercielle softwareprodukter, og institutionen anvender derfor også proprietær software. I det omfang det er muligt, deltager biblioteket aktivt i videreudviklingen af open source-produkter, ligesom vi ønsker at være en aktiv bidragyder i brugerfora og lign. for de kommercielle produkter, vi anvender.

I overvejelser omkring udvikling, anskaffelse og implementering af nye softwareløsninger indgår både åbenhed, økonomi, sikkerhed og eksistensen af internationale fællesskaber. Softwareløsninger underkastes risikovurderinger, både ved anskaffelse og løbende derefter.

Det Kgl, Bibliotek vil i den aktuelle strategiperiode benytte både open source og proprietær software til håndtering af forskellige dele af bevaringsprocesserne for vores digitale samlinger. Bibliotekets produktportefølje modsvarer en væsentlig del af bibliotekets aktuelle krav til digital bevaring. Vi overvåger løbende øvrige produkter og teknologier med henblik på i tide at kunne udbygge porteføljen og udskifte løsningerne, når der opstår behov herfor.

Indlemmelse

Vision:

- Det Kgl. Bibliotek sikrer, at alle modtagne digitale materialer overføres korrekt og komplet i modtagelsesprocessen
- Det Kgl. Bibliotek sikrer, at alle materialer har universelle, unikke og persistente identifikatorer, så materialet kan genfindes

Mål

- I løbet af 2020 er der fastlagt en procedure, som sikrer, at man ved indlemmelse så vidt muligt får de relevante metadata, fx om en digitaliseringsproces

Det Kgl. Bibliotek tilstræber at modtage samlinger i så korrekt form som muligt. Som en del af sikringen af data-indsamling anvendes bl.a. checksummer til sammenligning af data. Hvis muligt genereres checksummerne hos leverandøren, dvs. før indlemmelse på Det Kgl. Bibliotek. Alternativt genereres de straks ved indlemmelse. Hvor det vurderes væsentligt - og praktisk muligt - gennemgår samlingerne en automatisk validering og karakterisering ved indlemmelse.

Samlinger bevares af princip i det/de format(er), de er modtaget/indsamlet i. Der vil dog være undtagelser (fx Netarkivet). Ved undtagelse udarbejder Funktionsgruppen for Digital Bevaring en begrundet indstilling, som skal godkendes af samlingsejer og derefter dokumenteres i bevaringsplanen. Ved en evt. senere migrering, bevares både original og migreringskopi i det omfang, det er muligt. Har Det Kgl. Bibliotek mulighed for at påvirke valg af format, fx ved digitalisering af egne samlinger eller gennem aftale med en leverandør, vil Funktionsgruppen for Digital Bevaring træffe beslutning om bedst egnede format ud fra kriterier som formatets udbredelse, bevaring af signifikante egenskaber og økonomi.

Bevaringsprincipper

Det Kgl. Bibliotek tilstræber at følge international best practice i udførelsen af både bit- og logisk bevaring. Det Kgl. Biblioteket bruger risikostyring til at sikre, at repository-ændringer, som fx anskaffelse eller nedlæggelse af software- og hardware-systemer, eller oprettelse af nye samlinger, sker i forhold til disse principper. De ansvarlige for en given ændring udarbejder en risikovurdering i samarbejde med Funktionsgruppen for Digital Bevaring.

Bevaringsprincipper skal anvendes både på repository- og samlingsniveau. På repositoryniveau skal biblioteket leve op til anerkendte standarder inden for software-kvalitet, sikkerhed og procedurer.

Hver samling har sine egne egenskaber angående indhold, struktur, rettigheder osv. For hver samling sikres, at der afklares:

- Formater for data
- Samlingens overordnede struktur, dvs. relationer mellem de enkelte elementer i samlingen
- Bitbevaringsstrategi
- Logisk bevaringsstrategi

For hver samling vælges også formater for de forskellige typer af metadata ud fra de vedtagne procedurer og guidelines, med argumentation for evt. afvigelser. Disse oplysninger specificeres i bevaringsplanen og skal være så detaljerede, som det er praktisk muligt.

Bitbevaring

Vision:

- Det Kgl. Bibliotek tilstræber at bitbevare samlinger på så ensartet og optimal måde som muligt med hensyn til antal kopier, uafhængighed mellem kopier, samt integritetstjek med hensyntagen til krav om fortrolighed, tilgængelighed og bitsikkerhed
- Det Kgl. Bibliotek er aktiv i internationale diskussioner om problematikker vedrørende bitbevaring blandt andet via artikelbidrag og deltagelse i relevante workshops

Mål:

- I løbet af 2019 er alle bitbevaringsløsninger baseret på softwaren bitrepository.org
- I løbet af 2020 er bitbevaringskrav til uafhængighed opfyldt - med hensyn til geografi (i Danmark) og organisation
- I løbet af 2020 er risikostyringsprocedurer for bitbevaring etableret
- I løbet af 2020 er teknologiovervågningsprocedurer for bitbevaring etableret
- Biblioteket vil arbejde på at få et udvidet nordisk samarbejde om bitbevaring inden 2021
- I løbet af 2021 er der idriftsat en bitbevaringsløsning for fortrolige, personlige arkiver
- Senest i løbet af 2021 er alle pligtafleverede data modtaget før 2020 under bitbevaring

Det Kgl. Bibliotek implementerer bitsikkerhed ud fra de basale principper om, at der er:

- Et antal kopier af data
  Hvor der er nok kopier til at kunne identificere og genetablere mistede eller skadede kopier
- Uafhængighed mellem kopier af data
  Hvor uafhængigheden skal sikre, at den samme fejl/skade ikke kan ramme flere kopier
- Regelmæssige integritetstjek af og mellem kopierne
  Hvor fejl/skade kan opdages tidsnok til at kunne rettes, inden det medfører tab af data

Fastlæggelse af forholdet mellem disse parametre sker ud fra en kombineret risikovurdering og omkostningsanalyse. De risici, som skal tages i betragtning, skal som minimum omfatte organisatoriske, lovgivningsmæssige, teknologiske og geografiske forhold.

Det Kgl. Biblioteks strategi er at understøtte flere bevaringsløsninger for at kunne honorere samlingernes og brugernes forskellige krav til:

- bitsikkerhed
- fortrolighed
- tilgængelighed

Implementeringen af understøttelsen skal så vidt muligt honorere disse krav, når de sammenholdes med rammerne for at implementere bevaringsløsninger (herunder økonomi og teknologi). Disse differentierede implementeringer kommer til at omfatte løsninger for såvel ikke-fortrolige som hemmelige data, og med bevaringsniveauer fra opbevaring for materialer med fysisk kopi (dvs. en kopi med backup) til aktiv bitbevaring med flere kopier. Ved aktiv bitbevaring forstås løbende integritetscheck og fejlretning.

Overordnede bitbevaringsløsninger
En bitbevaringsløsning er karakteriseret ved, hvilket bitbevaringsniveau og fortrolighedsniveau den opfylder, men kan også have karakteristika fx i forhold til tilgængelighed og økonomi per TB. Det Kgl. Bibliotek tilstræber at holde antallet af bitbevaringsløsninger på et minimum, uden at gå på kompromis med krav til bevaringsniveau. Som udgangspunkt tilstræber biblioteket at have én løsning til hvert af de bevaringsniveauer, der er behov for, ved at se på krævet niveau af bitsikkerhed og fortrolighed. Undtagelser for dette princip skal kunne argumenteres ud fra økonomiske eller tekniske betragtninger.

For de løsninger, som indgår i Det Kgl. Biblioteks bitbevaring, vedligeholdes specifikationerne i henhold til løbende risikovurderinger på det overordnede plan. Der findes fast etablerede rutiner for reetablering og logning i overensstemmelse med Det Kgl. Biblioteks sikringspolitik.

Det Kgl. Bibliotek tilstræber at samarbejde med andre organisationer for at opnå de bedste og økonomisk mest fordelagtige betingelser for bitbevaring. Biblioteket vil blandt andet satse på et udvidet nordisk samarbejde.

Den løbende teknologiovervågning og risikostyring for bitbevaringsløsningerne skal bidrage til at sikre udførelse af relevante justeringer af bitbevaringsløsningerne. Dette gøres for at opnå og fastholde, at Det Kgl. Bibliotek benytter de mest optimale bitbevaringsløsninger i henhold til rammerne udstukket af politikken.

Risikostyring i det daglige arbejde med langtidsbevaring af bits er med til at sikre, at nødvendige ændringer og justeringer foretages tidsnok til at undgå datatab. En del af risikostyringen er også at eskalere diskussioner og beslutninger om risikomæssige hændelser til styregruppe og Sikringsudvalg, jf. afsnittet om Organisation, når fx en løsning vil kræve en større økonomisk eller ressourcemæssig indsats, eller der er risiko for et større datatab. Kendte eksempler på risici er:

- **Komprimering**, som kan resultere i, at data mistes, fx hvis de ikke kan dekomprimeres igen.
  Som udgangspunkt benytter det Kgl. Bibliotek ikke komprimering for digitalt bevarede materialer. Hvor det af praktiske eller økonomiske grunde besluttes at anvende komprimering af data i bevaringsprocesserne, anvendes reversible komprimeringsalgoritmer. Dette gælder kun komprimering efter modtagelse, dvs. det omfatter ikke komprimeringsalgoritmer, som er del af formatet for et modtaget materiale, eller som led i skrivning på et medie for en enkelt kopi på en replikaenhed.
- **Kryptering**, som kan resultere i, at data mistes, fx hvis nøgle til dekryptering mistes.
  Som udgangspunkt benytter det Kgl. Bibliotek ikke kryptering for digitalt bevarede materialer.
- **Sletning**, som kan give tab, hvis der er fejl forbundet med, hvad der slettes.
  Som hovedregel må der ikke slettes materiale, som er under bitbevaring. Sletning kan dog være nødvendig, hvis: 1) der er fejl forbundet med, at materialet blev bitbevaret, eller 2) materialet er udtaget til kassation. Ved sletning skal procedure for sletning

følges. Proceduren sikrer, at sletning ikke kan foretages af én person alene, ligesom det altid vil involvere en ledelsesgodkendelse, før en sletning må foretages.
- **Manglende kobling mellem identifiers og bitsekvenser**, som kan give tab, hvis bitsekvenser ikke kan genfindes ud fra deres tildelte identifier.
  Identificering sikres ved enten at pakke identifier med bitsekvenser (i WARC) eller ved at bevare checksum for filer sammen med deres identifier, så checksummen kan bruges til at genberegne kobling i tilfælde af tab.
- **Pakkeformater**, som kan give tab, hvis de fraviger generelle bevaringsprincipper.
  I det omfang det Kgl. Bibliotek benytter pakkeformater, bruges WARC, som opfylder krav til systemuafhængighed og til bevaringsformater.

Fremadrettet pakkes alle metadata i WARC for at sikre identificering og ensartet pakning.

Valg af bitbevaringsløsninger
Det er målet, at der er ensartede, letforståelige procedurer for, hvordan valg af bitbevaringsløsning for et materiale kan findes og beskrives. Et sådant valg er baseret på krav om bitsikkerhed, fortrolighed og evt. specialkrav til økonomi og/eller tilgængelighed. Procedurerne indeholder blandt andet en tjekliste af potentielle risici, som skal overvejes, samt vejledning om hvornår Funktionsgruppen for Digital Bevaring bør kontaktes.

Nedenfor beskrives mulige bitsikkerhedsniveauer, fortrolighedsniveauer og eksisterende bevaringsløsninger til understøttelse af krav.

Bitsikkerhedsniveauer
Bitsikkerhedsniveauer udtrykker, i hvilket omfang der bruges bitbevaringsprincipper for at sikre, at bitsekvenser vedbliver at være læsbare og ikke ændrer sig.

Niveauerne er fastsat ud fra de generelle bitbevaringsprincipper om antal kopier, uafhængighed og integritetstjek. Ved uafhængighed forstås såvel geografisk, organisatorisk som teknisk uafhængighed. Geografisk uafhængighed skal sikre, at samme begivenhed ikke kan skade for mange kopier (naturkatastrofer, terrorangreb, krig osv.). Organisatorisk uafhængighed skal sikre, at enkeltpersoner ikke kan skade for mange kopier. Eksempler på afhængigheder omkring teknologi er alt fra operativsystemer, hardware, software til håndtering af kopien og til forskellige checksumsalgoritmer i kombination.

For detaljer om niveauernes betydning og implementering, se Bilag 1: Bitsikkerhedsniveauer.

Fortrolighedsniveauer
Fortrolighedsniveauer udtrykker generelt, i hvilket omfang der skal laves adgangsrestriktioner til de bitbevarede data. Fortrolighed ses her som en sikkerhedsinformation (beskrevet i ISO 27000 serien), der er vigtig at have med i bitbevaringsprincipper, da redskaber til sikring af bitsikkerhed kan være i modstrid med redskaber til sikring af fortrolighed (mange kontra få kopier, brug af kryptering osv.).

For detaljer om niveauernes betydning og implementering, se Bilag 2: Fortrolighedsniveauer.

Bevaringsløsninger

Herunder vises en oversigt over bitbevaringsløsninger, som er i drift, og som dækker samlinger med samme karakteristika (angivet med store bullets), - samt manglende bitbevaringsløsninger, som skal implementeres, for at fortrolige samlinger kan bitbevares. Sidstnævnte løsninger forventes idriftsat inden for en overskuelig fremtid (angivet med kursiv og med stiplede, hvide bullets). De "uerstattelige digitale data" fra København er dog for det samlede overbliks skyld repræsenteret under en implementering, selvom flere meget ens løsninger p.t. eksisterer.

- Netarkivet. I 2018 ligger data ikke på software fra bitrepository.org.
- Radio/TV
- Uerstattelige digitale data fra København (Uerstat. KBH)
- Uerstattelige digitale data fra Århus (Uerstat. Århus)
- Digitaliserede billeder og bøger (Digi. KBH) med eksisterende fysisk kopi
- Digitaliserede aviser Århus (Digi. Århus) med eksisterende fysisk kopi
- Digitaliseringer med eksisterende fysisk kopi og en kopi hos anden organisation (men ikke under aktiv bitbevaring, fx ProQuest-digitaliseringer som også ligger hos ProQuest)
- *Personlige arkiver med mere (personarkiver)*
- *Hemmelige materialer (hemmelig)*

Bevaringsmatricen - fortrolighed, bitsikkerhed og overordnede mål:

Følgende tabel illustrerer, hvordan samlinger med forskellige karakteristika i dag er placeret med hensyn til bitsikkerhed og fortrolighed.

Bitsikkerhed / Fortrolighed	Meget lav	Lav	Middel	Høj	Meget høj
Minimal					
Meget lav					
Lav		Digi. m. kopier	Digi. Århus, Radio/TV, Digi. KBH, Uerstat. Århus, Uerstat. Kbh.		
Middel					
Høj				Netarkivet	
Meget høj					
Maximal					

Af tabellen fremgår det, at uerstattelige samlinger i henholdsvis Aarhus og København, samt Radio/TV har 'Middel' bitsikkerhed. Som det fremgår senere, burde disse samlinger have mindst 'Høj' bitsikkerhed. Denne uoverensstemmelse skyldes to forhold: 1: efter fusionen er der ikke længere organisatorisk uafhængighed mellem Aarhus og København, og 2: den geografiske afstand mellem kopierne for de aarhusianske samlinger er kun ca. 5 km og derfor ikke tilstrækkeligt stor.

Prioriteringer:

Det har højeste prioritet at få etableret organisatorisk og geografisk uafhængighed for de uerstattelige data, der nu kun har 'Middel' bitsikkerhed. Dette vil blive tilfældet, når Statens IT får overdraget en række kopier i løbet af 2018-2019, som beskrevet i afsnittet om placering af data nedenfor.

Det har også høj prioritet at få personarkiver under bitbevaring.

Efter implementering af disse højt prioriterede områder vil situationen være følgende:

Bitsikkerhed / Fortrolighed	Meget lav	Lav	Middel	Høj	Meget høj
Minimal					
Meget lav					
Lav		Digi. m. kopier	Digi. Århus Digi. KBH	Radio/TV → Uerstat. Århus →	Uerstat. KBH
Middel					
Høj				Netarkivet →	
Meget høj				*personarkiver* →	
Maximal			*hemmelig*		

Grå og sorte pile angivelser udtrykker vision

På nuværende tidspunkt findes der både aarhusianske og københavnske implementeringer for samme bevaringsniveau. Dette er historisk betinget, og på længere sigt (udover denne strategiperiode) er målsætningen, at der fastlægges én løsning for hvert bitbevaringsniveau i matricen ovenfor. Undtaget herfra er samlinger med specielle karakteristika, som fx Radio/TV-samlingen, der har sin egen løsning, fordi de meget store datamængder er for dyre at have på disk, - og derfor udelukkende er lagt på bånd replikaenheder (med forskellig leverandør, software m.m.).

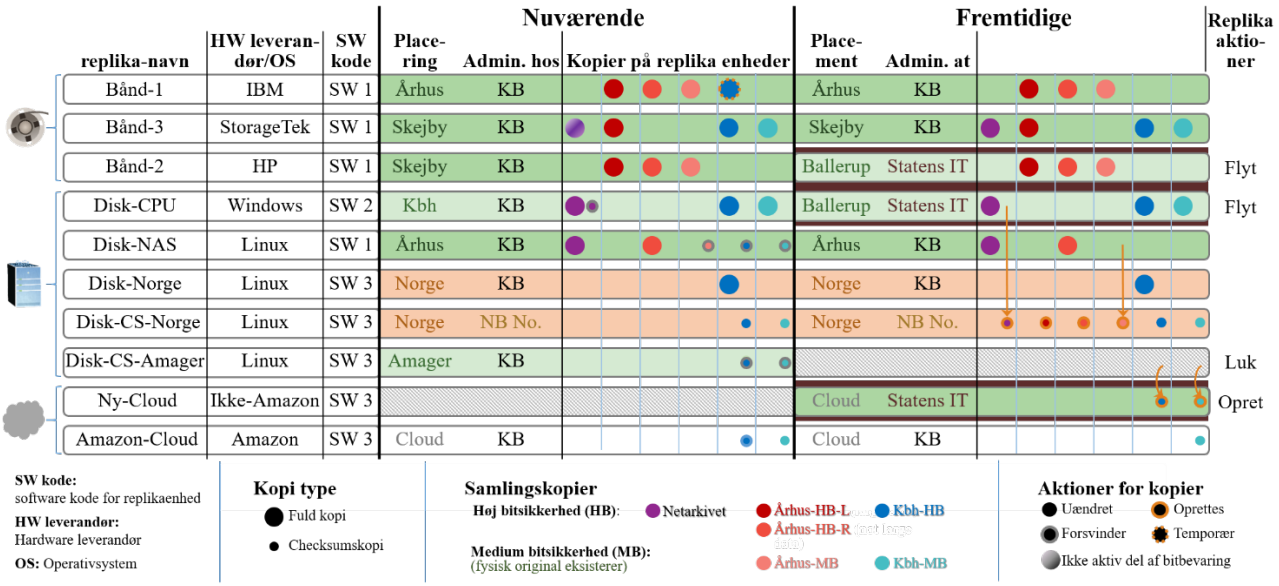
Herefter er det næste prioriterede skridt at etablere 'Meget høj' bitsikkerhed for de mest værdifulde samlinger, der endnu ikke har dette niveau. Figurens sorte pile illustrerer, hvilke samlinger der er tale om. Overgangen fra 'Høj' til 'Meget høj' bitsikkerhed kan for disse implementeringer sikres ved etablereringen af a) en udenlandsk kopi, og b) en bedre softwaremæssig uafhængighed imellem kopierne.

Den grå pil ønskes også iværksat. Dette vil dog kræve, at Radio/TV ikke kun lægges på bånd-medier. Trods forskel på lokation, leverandør og software, kan denne samling ikke opnå 'Meget høj' bitsikkerhed, så længe alle kopier er på samme slags medie. Derfor vil iværksættelse kræve, at en af kopierne kan placeres på et tilsvarende billigt medie. Et sådant medie findes ikke i dag. I stedet satses på bedste 'Høje' bitsikkerhed ved, at denne implementering også forbedres med en udenlandsk kopi og bedre uafhængighed af software.

Det Kgl. Bibliotek ønsker også at bitbevare hemmelige materialer, der dog ikke kan bevares med meget høj bitsikkerhed på grund af fortrolighedsniveau.

Placering af data for at opnå uafhængigheder:
Nedenstående figur viser placeringen af datakopier for de forskellige implementeringer af bitbevaringsniveau. For hver implementering er kopierne placeret på uafhængige replikaenheder (bestående af organisation og teknik for kopien). Uafhængigheden er beskrevet ved angivelse af relevant information i forhold til hver enkelt replikaenhed med henblik på hardware/software, geografisk placering og organisatorisk tilhørsforhold.

Figuren afbilder både den nuværende og den fremtidige situationen, svarende til de samme to fikspunkter, som i de to ovenstående figurer. Den fremtidige situation opnås med de angive replikaaktioner for replikaenheder, samt flytning af checksummer som angivet med gule pile.



Den fremtidige løsning vil understøtte førsteprioriteten med at etablere "Høj bitsikkerhed" for de samlinger, der siden starten af 2017 ikke længere har denne status på grund af den organisatoriske sammenlægning af de to før uafhængige biblioteker. Som angivet i figuren opnås dette ved: Overdragelse af ansvaret for replikaenhederne Bånd-2 og Disk-CPU til Statens IT, både organisatorisk og geografisk, samt 2: Overflytning af to aarhusianske checksumsreplika til Norge. Hermed genindføres både organisatorisk og geografisk uafhængighed for de data, der overdrages til Statens IT.

For at indfri den heraf følgende målsætning om etablering af "Meget høj bitsikkerhed", planlægger det Kgl. Bibliotek på længere sigt at flytte en data-replikaenhed (Bånd-3) til en udenlandsk samarbejdspartner. Dette kunne for eksempel være i form af et udvidet samarbejde med Norge, men hvor der er tale om en bånd-replikaenhed i stedet for en disk-replikaenhed. For Radio/TV vil det yderligere kræve at bånd 2 eller bånd 3 migreres til en anden slags medie.

Parallelt med målopfyldelsen i relation til geografisk og organisatorisk uafhængighed er der konkrete målsætninger relateret til softwarekomponenterne i bevaringsløsningerne:

1. For at understøtte opfyldelsen af det allerede planlagte mål om overgang til 'Høj' bitsikkerhed for alle relevante samlinger, omlægges Netarkivets bitmagasinsoftware til at køre på software fra bitrepository.org (som resten af bitbevaringen anvender). Denne omlægning er en forudsætning for at flytte data-replikaenheden Disk-CPU til Statens IT.
2. På længere sigt ønsker biblioteket at få udviklet ny software, som kan overtage styringen af nogle af de replikaenheder, der i dag anvender SW1 og SW3.
3. På længere sigt ønsker biblioteket også at se på mulighederne for at reducere antallet af online replikaenheder, da disse er væsentligt dyrere end offline bånd-replikaenheder.

Logisk bevaring

Mål:

- I 2019 er der gennemført test af en række emuleringsframeworks som fx Emulation as a Service (EaaS)
- I 2020 er der gennemført valg og etablering af prototype af et emuleringsframework
- I 2020 er indsamling af sekundært materiale (software, licenser, dokumentation, mv.) til understøttelse af emulering igangsat
- I 2023 er analyse og test af mulighederne for migreringer af store samlinger gennemført

Begrebet "logisk bevaring" omfatter en lang række aktiviteter, som sikrer, at digitale materialer kan anvendes nu og i fremtiden i en form, som er en troværdig replikering af deres originale indhold og funktionalitet. Logiske bevaringsaktiviteter, i denne bredde forstand, starter allerede ved indlemmelse - med indsamling af tekniske metadata, data-karakterisering og -validering.

Strategien for logisk bevaring varierer fra bevaringssamling til bevaringssamling, og det er derfor et vigtigt emne i udarbejdelsen af bevaringsplanen for hver bevaringssamling at identificere de væsentligste aktuelle risici for en given bevaringssamling og den eller de mest relevante risikohåndteringsstrategier. De to risiko-håndteringsstrategier for logisk bevaring, der anvendes på Det Kgl. Bibliotek, er migrering og emulering.

Under bearbejdelsen af bevaringsplanen og tilhørende risikovurdering for hver samling udpeger Det Kgl. Bibliotek en foretrukken logisk bevaringsstrategi (emulering eller migrering) for samlingen. Samtidig vurderes, om den eller de nuværende aktuelle risici er alvorlige nok til, at der skal sættes handlinger i gang for at anvende den foretrukne strategi.

Et centralt begreb i logisk bevaring er aktiv teknologiovervågning for at sikre, at man følger best practice og håndterer trusler fra teknologiændringer. Se afsnittet om Teknologiovervågning for flere detaljer.

Når løbende risikovurdering viser, at en trussel mod logisk bevaring er blevet aktuel, skal Funktionsgruppen for Digital Bevaring lave en anbefaling for, hvilke konkrete handlinger der bør igangsættes for at imødegå truslen.

Migrering og emulering

Migrering vil typisk være den anbefalede strategi for samlinger bestående af enkeltstående mediefiler såsom billeder, lyd, og video. Hvis resultatet af risikovurderingen fører til, at materialet migreres, bevares både original og afledte udgaver i det omfang, det er muligt.

Emulering vil typisk være den anbefalede strategi for samlinger, som inkluderer eksekverbare elementer, fx computerspil, tekstdokumenter (med makroer) og Netarkivet. I webarkiver er det stadig normalt at bruge nutidens webbrowsere til at vise historisk indhold, men der er stigende interesse for brug af emulerede historiske browsere. Computerspil (både til PC og konsol) er generelt ekstremt sårbare over for teknologisk forældelse, og her er emulering allerede udbredt.

For at understøtte anvendelsen af emulering vil Det Kgl. Bibliotek i samarbejde med andre bevaringsorganisationer løbende indsamle sekundært materiale (software, licenser, dokumentation, mv.). Denne indsamling foretages for, at emuleringsmiljøer, både nu og i fremtiden, kan opbygges med den nødvendige software.

Adgang til bevarede digitale materialer

Vision:

- Det Kgl. Biblioteks digitale bevaring understøtter i videst muligt omfang den fremtidige adgang til materialerne

Mål:

- I 2019 laves en procedure for at lave stikprøver på præsentationskopier til tjek af, at de stemmer overens med bevarede versioner

Formidlingskopier bidrager i et vist omfang til at højne kvaliteten af den digitale bevaring, idet aktiv brug kan være med til at validere og tjekke materialernes integritet. Den feedback, der modtages fra interne og eksterne brugere af de digitale samlinger, kan bruges aktivt i bevaringen og fx være med til at afklare, om et materiale er ufuldstændigt eller fejlbehæftet - og der derfor bør laves en ny udgave, hvor det er muligt -, eller om metadata er utilstrækkelige eller forkerte - og derfor bør opdateres, hvor det er muligt.

Det tilstræbes at logge enhver intern og ekstern adgang til samlingerne, dels så et audittrail for bevaringen kan følges og dels for at sikre, at fx personfølsomme data ikke tilgås af uvedkommende.

Det Kgl. Biblioteks bevaringsunderstøttende systemer skal samtidig også understøtte de krav, der måtte være til adgang til det bevarede materiale i forbindelse med formidling, levering, bevaringsaktiviteter og kuratering.

Data, metadata og dataformater

Vision

- Det Kgl. Bibliotek bitbevarer data og metadata i en form, der til enhver tid kan forstås og fortolkes i fremtiden. Dette indebærer følgende for data og metadata

Mål:

- Løsning til bevaring af metadata fra Kuana er klar til implementering med udgangen af 2019
- I 2019 påbegyndes en analyse af mulighederne for en samlet bevaringsløsning af metadata fra alle Det Kgl. Biblioteks systemer til understøttelse af bevaring
- I 2021 er en samlet bevaringsløsning for metadata fra alle Det Kgl. Biblioteks systemer til understøttelse af bevaring igangsat. Arbejdet baseres på input fra den ovenstående analyse
- I 2022 er alle bevarede data og metadata repræsenteret i en form som opfylder krav til datamodel og standardisering

For at data og metadata til enhver tid kan forstås og fortolkes i fremtiden iværksættes en række tiltag i forbindelse med datamodel for data og metadata, samt metadata og dataformater generelt. Dette omfatter anvendelse af en så simpel datamodel for data som muligt, under hensyntagen til variation og kompleksitet af data og metadata. Derudover tilstræbes det, at alle former for formater har stor sandsynlighed for at være forståelige i fremtiden. Dette vurderes ud fra i hvilket omfang de er åbne, standardiserede og internationalt anerkendte.

En delvis standardisering af data- og metadataformater under bevaring er ønskelig for at gøre bevaringsprocessen så effektiv og billig som mulig. Det skal dog understreges, at på grund af bredden af indhold, som ønskes bevaret, vil det aldrig være muligt med en fuldstændig standardisering. For eksempel er der mange værktøjer, som eksporterer metadata i ikke-standardiserede formater. Endvidere er der historisk blevet brugt forskellige standarder på tværs af biblioteket til at beskrive indhold. Bagudrettet fastholdes de allerede anvendte formater, da det at konvertere metadata mellem forskellige formater er en ikke-triviel opgave, som medfører risiko for tab af betydning.

Det Kgl. Bibliotek søger at bevare sine digitale samlinger i så få dataformater som muligt. Der vil dog aldrig ske en fuldstændig normalisering af dataformater under bevaring. For dataformater for modtagne materialer skyldes dette at biblioteket modtager materiale uden indflydelse på materialers dataformater - og biblioteket ønsker ikke at foretage migrering, før det er nødvendigt. For metadataformater skyldes dette, at der er mange specialiserede formater for forskellige typer af digitale materialer (fx er MIX kun for still-billeder) og at metadata kan være eksporteret i givne formater som nævnt ovenfor.

Datamodel

Datamodellen for de bitbevarede materialer er udformet, så den vil kunne bruges på langt sigt. Modellen kan understøtte leverance af data til forskellige applikationer, uanset hvilke dele af data de benytter, og uanset hvilke datamodeller applikationerne benytter - med andre ord, modellen understøtter, at alle relationer kan genetableres ved gennemlæsning af de bitbevarede data. Endvidere har datamodellen til formål at strukturere data på en sådan måde, at de kan forstås på længere sigt. Derfor er den skitserede model en simpel datamodel uden hensyntagen til optimering af forskellige frontend brugsscenarier.

Den skitserede datamodel er en forenkling af andre modeller kendt fra sammenhænge som fx PREMIS og modellen fra Planets-projektet (delvist brugt i Preservica). Efter en kort gennemgang af modellen er der givet et konkret eksempel samt referencer til eksempler på modellering.

Grundlæggende datamodel

Datamodellen har tre grundlæggende entiteter (illustreret i nedenstående figur):

- **Digital Intellektuel Entitet**
  Denne udtrykker det øverste niveau for et digitalt objekt, som entydigt identificerer et digitalt materiale, som ønskes identificeret. Objektet skal kunne identificeres, uanset hvilke bevaringsaktioner, tilrettelser eller transformationer der har været lavet. En Digital Intellektuel Entitet adskiller sig fra en FRBR Intellektuel Entitet ved, at forskellige udgaver af et værk, som er skabt uafhængigt af hinanden, vil opfattes som forskellige Digitale Intellektuelle Entiteter.

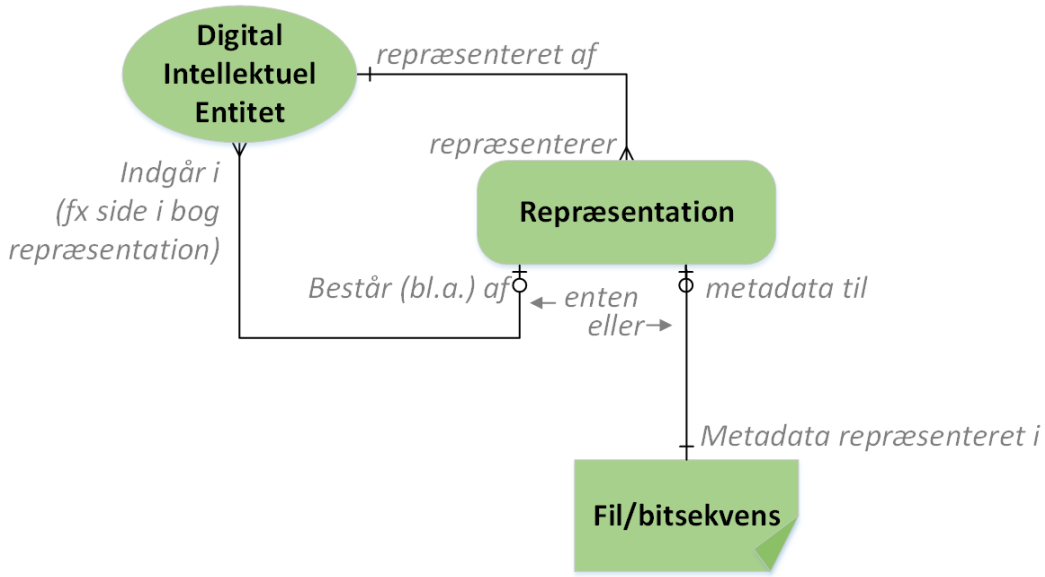
  En Digital Intellektuel Entitet vil altid bestå af en eller flere repræsentationer, som repræsenterer forskellige udgaver/versioner af den Digitale Intellektuelle Entitet.

  Bemærk, at der kan være specielle tilfælde, hvor det ikke er entydigt, om en ændring til en repræsentation for en eksisterende Digital Intellektuel Entitet resulterer i en ny repræsentation af den Digitale Intellektuel Entitet eller resulterer i en ny Digital Intellektuel Entitet. Det vil være op til en kurator at beslutte dette på basis af en vurdering af, hvorvidt man ønsker at se det som to enheder frem for to versioner af samme enhed. Et eksempel er opdaterede udgaver/versioner af e-bøger.

- **Repræsentation**
  Denne udtrykker en Repræsentation af en Digital Intellektuel Entitet, dvs. den repræsenterer en bestemt version/udgave af en bestemt Digital Intellektuel Entitet. En Repræsentation kan kun repræsentere én Digital Intellektuel Entitet.

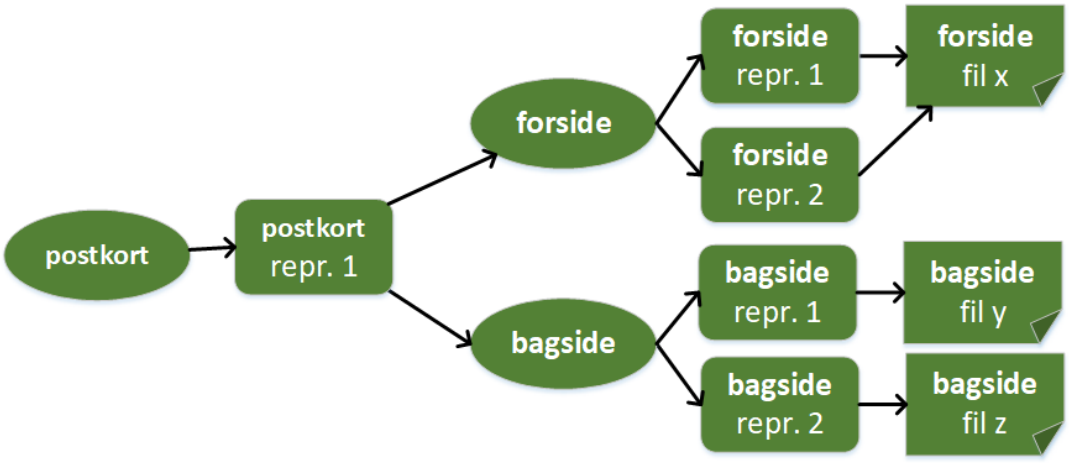
  Indholdet af Repræsentationer kan være forskelligt. Enten kan en Repræsentation indeholde reference til en fil og indeholde denne fils metadata, eller den refererer til fragmenter, som identificeres via fragmenternes Digitale Intellektuelle Entiteter, samt indeholder metadata om fragmenternes indbyrdes relation for at udgøre Repræsentationen (fx rækkefølgen af sider i en bog, hvor fragmenterne er bogens sider)

- **Fil/Bitsekvens**
  Denne er blot en fil/bitstrøm, som bevares i et bitmagasin (dataobjektet).



Selve afgørelsen af, hvor mange niveauer der benyttes i en modellering, fastlægges ved ingest til bevaring, som det fx er blevet gjort for data, som ingestes via Kuana, SB-DOMS og Cumulus.

Eksempel på modellering af et digitaliseret postkort
Et eksempel er forsiden og bagsiden af et digitaliseret postkort. Selve digitaliseringen er lagt i filen "forside fil x" og filen "bagside fil y", hvor bagsiden senere er gendigitaliseret og lagt i filen "bagside fil z".



Repræsentationerne for filerne indeholder de relevante metadata til filerne, inklusiv metadata fra fx karakterisering eller digitaliseringen. I dette eksempel har filen for forsiden på et senere tidspunkt fået opdateret metadata, som er lagt i en ny Repræsentation af den samme fil, og bagsiden har fået en ny fil med tilhørende nye metadata (fx i forbindelse med en migrering).

Der er Digitale Intellektuelle Entiteter for henholdsvis for- og bagside, som de forskellige Repræsentationer hører til, og det er disse Digitale Intellektuelle Entiteter, som Repræsentationen for postkortet peger på.

En mere detaljeret beskrivelse af disse relationer kan findes i Bilag 3: Detaljeret postkort datamodel eksempel.

En beskrivelse af, hvordan datamodellen implementeres (inkl. relation til Det Kgl. Biblioteks nuværende systemer til understøttelse af bevaring Kuana, Cumulus og SB-DOMS) kan findes i Bilag 4: Detaljerede implementeringseksempler.

Metadata

Det Kgl. Bibliotek tilstræber så vidt muligt at bitbevare metadata for de digitale materialer, der skal bevares.

Bevaring af metadata er vigtigt, da det er grundlæggende for al brug og bevaring af vores digitale materialer. Derudover udgør metadata ofte en større investering af tid og ressourcer, der gør dem både svære og dyre at genskabe, hvis de går tabt.

Når der foretages ændringer/opdateringer af digitale objekter og/eller metadata, skal dette logges, dvs. der skal være et audittrail for både materialer og metadata. Disse audittrails er at betragte som metadata og skal også bitbevares.

Overordnet set ønsker Det Kgl. Bibliotek at bevare metadata af følgende typer:

- *Beskrivende metadata*
  Med informationer der beskriver, hvad det digitale objekt er

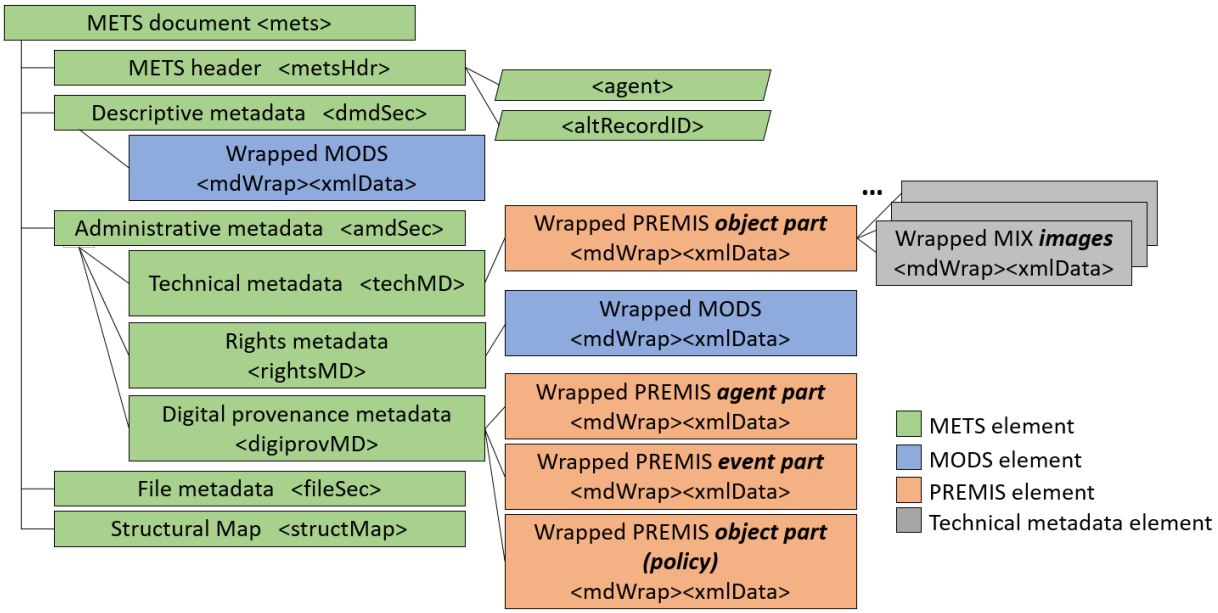
- *Administrative metadata*
  Med nødvendig information til forvaltning af det digitale objekt, hvilket omfatter:
  - *Tekniske metadata*, fx filformat og informationer fra digitalisering
  - *Bevaringsmetadata*, som omfatter metadata, der er nødvendige for bevaringsaktioner, fx bevaringsniveau og bevaringsstrategi
  - *Digital proveniens*, som indeholder audittrails for, hvad der er foretaget på det digitale objekt
  - *Rettighedsmetadata*, som kan bruges til at finde frem til, hvem der har lov til at tilgå materialet
- *Strukturelle metadata*
  Med informationer om strukturer der måtte være for det digitale objekt (for fx en bog vil dette være referencer til sider)

Det er et ønske og arbejdspunkt for Funktionsgruppen for Digital Bevaring, at man på tværs af huset anvender de samme metadatastandarder på en så standardiseret måde som muligt. Dette medvirker både til mere effektiv bevaring og gør det nemmere at give adgang til materialet.

Der findes i dag et eksisterende workflow, der lægger metadata fra Cumulus systemet i en METS container og derefter pakker filen i en WARC pakke, som bitbevares. Der skal udvikles et tilsvarende workflow til bitbevaring af metadata fra Kuana.

Bitbevaring af metadata vil ske gennem jævnlig eksport af metadata, inklusiv audittrails vedrørende de enkelte objekter, fra repository systemerne (som KUANA og Cumulus) til en standardiseret METS profil, som vist i illustrationen nedenfor. Dette gøres for at sikre en ensartet struktur for de bevarede metadata så uafhængigt af det system, de kommer fra, som muligt. Som en del af projektet om bitbevaring af metadata fra Kuana bliver nedenstående model opdateret, så metadata fra Kuana mappes ind i modellen.

Disse bitbevarede metadata gemmes pakket som WARC-filer. Denne pakning foregår uden for bitmagasinet.



For at sikre at bibliotekets metadata er brugbare på lang sigt, skal metadataprofil og de metadatastandarder, der anvendes i vores digitale samlinger, være tilgængelige, også hvis den oprindelige placering for fx et metadataskema ikke længere er tilgængelig. Ansvaret herfor ligger hos Funktionsgruppen for Digital Bevaring, og det sker konkret på id.kb.dk, som med jævne mellemrum bevares i Netarkivet.

Metadata i forbindelse med bevaring

Bevaring af Det Kgl. Biblioteks materiale forudsætter tekniske metadata om de enkelte materialer, der skal beskrive, hvad materialet er (fx et jpeg billede). Derfor skal alt materiale, der bevares, gennemgå en karakterisering og hvor muligt også en validering så tidligt i materialets livscyklus som muligt, og output fra disse processer gemmes som tekniske metadata for materialet.

Derudover bør et objekts tilblivelse og historik beskrives i metadata, fx skannet på skanner med serienr. xx, derefter bearbejdet med software xx, eller det kan være metadata som omdrejningshastighed, pickup-type og afspiller ved en pladedigitalisering. Et tredje eksempel kunne være, hvilken software og hardware der er anvendt til at lave et image af en disk fra et privat arkiv. Disse oplysninger udgør proveniensen for bibliotekets digitale materialer og er vigtige for deres fortolkning.

Ved digitalisering foretaget hos eksterne leverandører gælder som udgangspunkt de samme krav om metadata fra digitaliseringsprocessen, men her har biblioteket som institution mindre indflydelse på, hvordan leverandøren kan levere disse metadata, da dette er meget afhængigt af leverandørens tekniske setup. Det skal derfor være et fast punkt i ethvert eksternt digitaliseringsprojekt at afklare, hvad en leverandør kan levere af metadata og i hvilke formater.

I alle processer, der genererer indhold til de digitale samlinger, bør checksummer laves på så tidligt et stadie i materialets livscyklus som muligt. Dette gælder også, når der digitaliseres ved eksterne partnere.

De tekniske metadata for materialer i Det Kgl. Biblioteks samlinger kan også beskrive et materiales struktur, rettigheder, samt en historik over hvilke bevaringshandlinger et objekt har gennemgået.

Valg af værktøjer til karakterisering og validering vil være afhængigt af materialet, men et fælles krav er, at de metadata, der genereres, skal gemmes i en struktureret form, og disse metadata skal angive, hvilket værktøj og i videst muligt omfang hvilken version af værktøjet der er brugt.

Konkrete eksempler på, hvordan filer med metadata er bitbevaret (inkl. relation til Det Kgl. Biblioteks nuværende systemer til understøttelse af bevaring Kuana, Cumulus og SB-DOMS), kan findes i Bilag 4: Detaljerede implementeringseksempler.

Dataformater

Det Kgl. Biblioteks samlinger indeholder digitale materialer i mange forskellige formater, nogle egnede til bevaring, andre ikke. Formater, der er egnede til bevaring, er karakteriseret ved at være åbne, standardiserede, udbredte eller internationalt anerkendte som bevaringsformater og gerne bredt understøttet af standard software-værktøjer.

Det Kgl. Bibliotek bevarer som udgangspunkt materiale i det format, det er modtaget eller indsamlet i, og laver ikke om på materialer (via migrering eller pakning) ved indlemmelse i bibliotekets systemer til understøttelse af bevaring. Undtaget herfra er dog Netarkivet, der pakker og samler objekter i WARC-filer, hvilket er et resultat af den teknologi, der anvendes til indsamling.

Funktionsgruppen for Digital Bevaring har som en af sine opgaver at have et samlet overblik over, hvilke dataformater der produceres ved interne processer, og gruppen rådgiver i forhold til eksterne digitaliseringsprojekter og ved overdragelse/levering af samlinger til biblioteket.

Teknisk infrastruktur

Vision:

- Det Kgl. Bibliotek tilstræber at ensarte bevaringprocesser på en så optimal måde som muligt med hensyntagen til krav om fortrolighed, tilgængelighed og kompleksitet i de digitale materialer
- Det Kgl. Bibliotek er aktiv i internationale diskussioner om problematikker vedrørende bevaringsprocesser blandt andet via artikelbidrag og deltagelse i relevante workshops

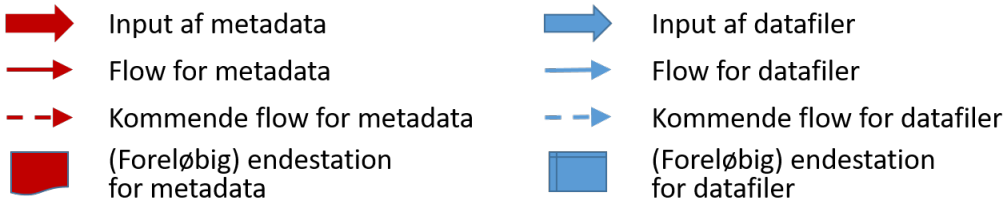
Mål:

- I 2020 er der idriftsat bevaring af metadata fra KUANA
- Senest i løbet af 2021 er der etableret et fælles pre-ingest-område
- I første halvdel af 2022 er der etableret bitbevaring af alle metadata fra Netarkivet
- I løbet af 2022 er der etableret en type metadata-warehouse, som kan repræsentere og tilgå bitbevarede metadata for alle bevarede materialer
- I løbet af 2022 er SB-DOMS udfaset
- I løbet af 2023 er der lavet integration mellem bevaringsadministration (fx med bevaringsplaner) og metadata-warehouse

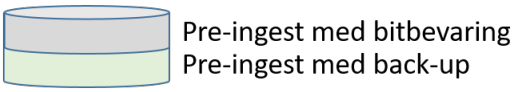
Nedenfor er arkitekturen gennemgået i tre trin:

- den nuværende arkitektur,
- den fremtidige arkitektur, som Det Kgl. Bibliotek vil stræbe efter at have på plads inden for en 3-5 årig periode
- den arkitektur, som det er visionen at nå frem til

De tre arkitekturtegninger vil benytte sig af følgende symboler:



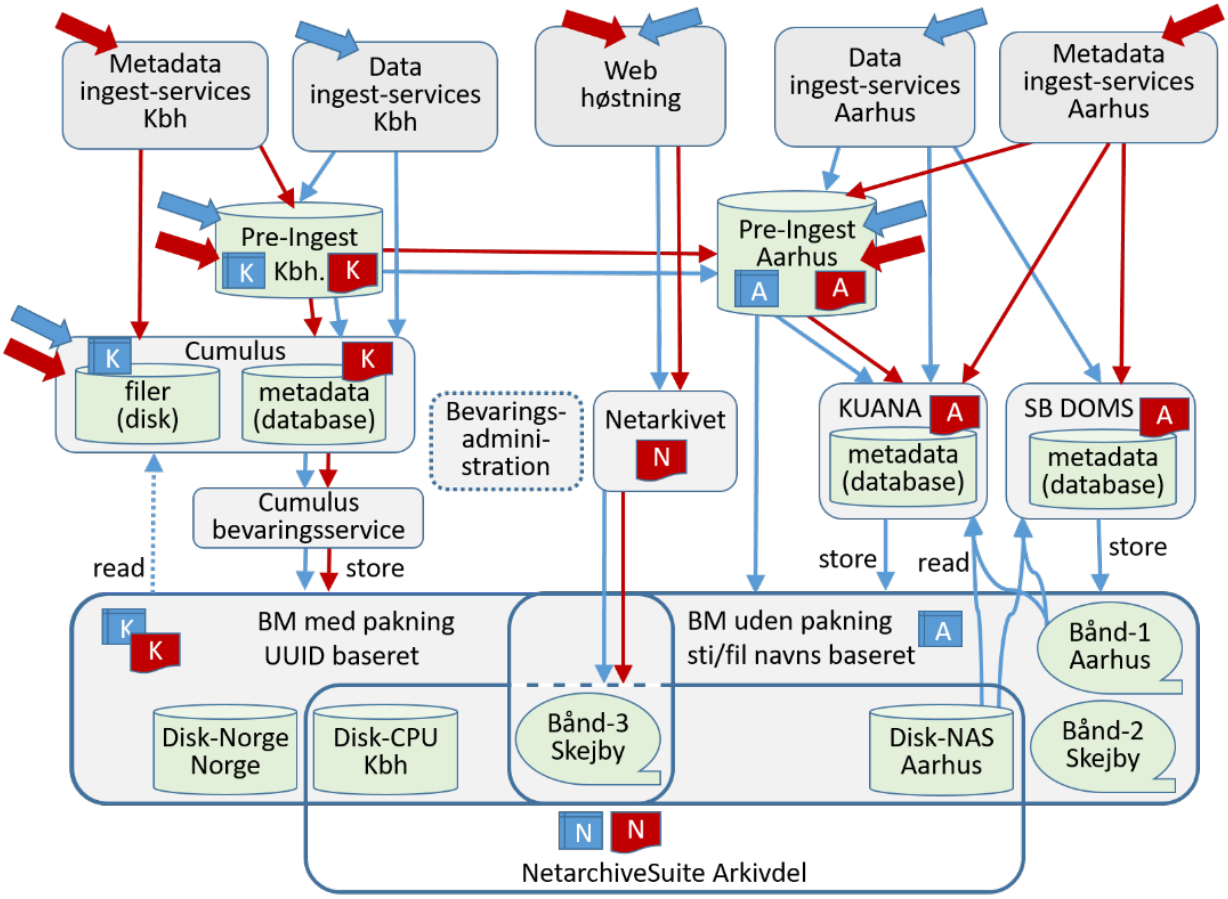
Hvid skrift angiver om det er Kbh (K), Netarkivet (N) eller Aarhus (A)



Den nuværende arkitektur

Nedenstående figur illustrerer den nuværende arkitektur, som bærer præg af, at Det Kgl. Bibliotek er resultatet af en fusion mellem de to tidligere nationalbiblioteker, hvor der er en del i København (primært afbildet i venstre side) og en del i Aarhus (primært afbildet i højre side), og et fælles Netarkiv (afbildet i midten).

Arkitekturen, der herunder omtales, omfatter udelukkende systemer til bevaring, hvilket vil sige, at formidlings- og adgangsplatforme ikke er medtaget (fx API til forskere, Kulturarvscluster, OAI-PMH udtræk fra KUANA, API'er på bitmagasin). Af samme grund er materialer, som endnu ikke er kommet i de systemer, der understøtter bevaring (materialer fra fx Github repositories, ADL og Søren Kirkegaard data) heller ikke taget med.



I den nuværende arkitektur er der forskellige metadata og data ingest-services. Eksempler på sådanne services er Teracom for radio/tv data, Infomedia, Elivagar og Ritzau for metadata, og GoAnywhere (og indirekte Publizon), som både dækker data og metadata. Disse services leverer data til de såkaldte repository-systemer til understøttelse af bevaring eller til Pre-ingest områder med data og metadata, som endnu ikke er klar til at blive ingestet i repository-systemerne.

Der er en række data og metadata input-pile, som ikke går gennem ingest-systemer. Disse pile illustrerer, at der er data og metadata, som fx visse computerspil, der lægges direkte i den københavnske Pre-ingest, og digitaliseringer som lægges direkte ind i Cumulus. Tilsvarende er der data, som lægges direkte i den aarhusianske Pre-ingest.

Der er for nuværende to Pre-ingest systemer. Det ene ligger i København med interface primært til Cumulus. Det andet ligger i Aarhus og interfacer til hhv. KUANA og SB-DOMS. I begge pre-ingest systemer sker sikring af data udelukkende via backup. Linjerne mellem Pre-ingest i København og Aarhus afspejler, at e-bøger og lydbøger er på vej til at blive ingestet i KUANA via Pre-ingest i Aarhus. Netarkivet får både data og metadata fra indsamlings- og ingest-systemer indbygget i NetarchiveSuite og WebDanica samt Archive-It indsamlinger.

De fire repository-systemer understøtter i hovedtræk følgende områder af digital bevaring:

- Cumulus laver strukturering af data (tilgået via disk) og metadata (via indlejret database) og giver mulighed for editering af metadata.
  Der er et separat program (Cumulus bevaringsservice), som sørger for, at både metadata og filer bliver bitbevaret med det bevaringsniveau, som er specificeret i metadata.
  Der er ikke nogen automatiseret måde at få data fra bitmagasinet til Cumulus, men data fra bitmagasinet kan inspiceres via webgrænseflade til bitmagasinet. Dette er grunden til, at pilene fra bitmagasinet til Cumulus er stiplede.
- Netarkivet er med NetarchiveSuite softwaren en integreret del med ingest-systemerne til overførsel af data til bitbevaring (dog mangler bitbevaring af nogle metadata om høstningsjobs). Derudover er der et overførselssystem til overførsel af data indhentet via Archive-It.
- KUANA laver karakterisering og til dels strukturering af data, hvor filer lægges direkte i bitmagasinet som del af ingest, og hvor metadata gemmes i en intern database.
  Der er i den nuværende løsning ikke overførsel af metadata til bitmagasinet fra KUANA.
  De bitbevarede filer kan inspiceres direkte i KUANA.
- SB-DOMS (som her også omfatter tilhørende workflowsystemer og GUI) laver karakterisering af data via workflows, som sørger for, at data lægges i bitmagasin, og metadata lægges i den interne database.
  Der er i den nuværende løsning ikke overførsel af metadata til bitmagasinet fra SB-DOMS.
  De bitbevarede filer kan i visse tilfælde stilles til rådighed via SB-DOMS.
  SB-DOMS bygger på en forældet teknologi og er allerede planlagt at skulle udfases.

Bevaringsadministration, som blandt andet omfatter bevaringsplaner på samlingsniveau og filformatniveau, er angivet med stiplet kant, da den er under konstruktion. Bevaringsadministration vil i første omgang fungere uafhængigt af resten af bevaringsapplikationerne.

Bitmagasinet er i illustrationen angivet som ét bitmagasin, men er i realiteten flere forskellige bitmagasiner:

- En københavnsk opsætning og brug af bitrepository.org frameworket, som pakker i WARC og bruger UUID identifiers - hvor bitbevaringen er baseret på WARC-pakker
- En århusiansk opsætning og brug af bitrepository.org frameworket, som bruger fil/sti identifiers - hvor bitbevaringen er baseret på filer (metadata er endnu ikke med)
- NetarchiveSuite's bitmagasin software, suppleret med en backup (stiplet linje til Bånd-3, da denne er en separat back-up-kopi, som ikke indgår i den aktive bitbevaring).

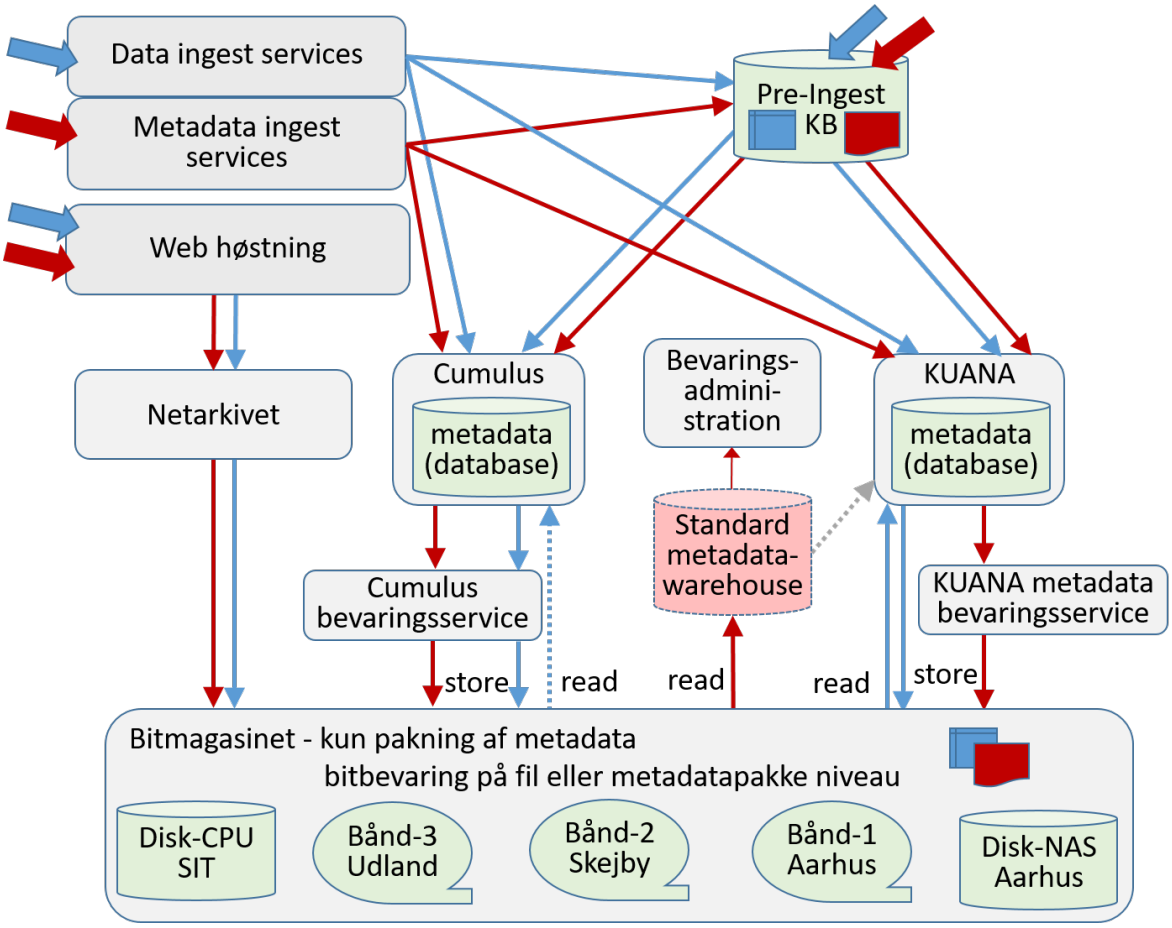
Detaljer om dette kan findes i afsnittet om bitbevaring.

Tilgang til data i bitmagasinerne sker primært via API-kald med undtagelse af en speciel masseprocesseringsplatform for Netarkivet.

Den fremtidige arkitektur (helst om 3 - senest om 5 år)

Der arbejdes hen imod at optimere infrastrukturen så meget som muligt for at få bedre overblik, færre systemer der skal vedligeholdes, og en bedre sikring af data. I nedenstående skitse af arkitekturen på kort sigt er denne strategi afspejlet ved,

- at ingest foregår igennem langt færre knudepunkter
- at alle data enten er bevaret eller i pre-ingest
- at antallet af repository-systemerne, der understøtter bevaring, er reduceret
- at der er indført en form for standard metadata-warehouse, som kan give overblik over samtlige metadata på ens og systemuafhængig vis.



I denne arkitektur vil input af data og metadata være strømlinet i størst muligt omfang via samlede data og metadata ingest-services. Realistisk set vil dette aldrig kunne omfatte alle data og metadata, da der altid vil kunne opstå pukler eller nye former for data og metadata, som må afvente implementeringer eller udvidelser af kapacitet. Derfor viser illustrationen, at der også vil være input af data og metadata til Pre-ingest. Målet er at få minimeret den allerede eksisterende pukkel af data i Pre-ingest så meget som muligt.

Alle data og metadata, som i dag eksisterer i systemer uden at være bevaret, vil være under bevaring. Dvs. alle data og metadata er enten bevaret eller ligger på pre-ingest området og

venter på at komme til bevaring gennem systemerne. Der vil her kun vil være ét Pre-ingest-system (understøttet af backup).

Antallet af repository-systemer, som understøtter bevaring, er indsnævret til tre systemer, idet SB-DOMS forventes udfaset som planlagt. Netarkivet er bibeholdt, fordi det afviger kraftigt fra de andre systemer. KUANA forventes på sigt at skulle kunne rumme alle informationer om bevarede data, men der er flere grunde til, at dette skal ske gradvist, og Cumulus antages derfor at være i spil på en 3-5 årig horisont, som bygger på følgende observationer:

- KUANA kan ikke i tilstrækkelig grad supportere forvaltningsfunktioner for fx billeder, og det forventes ikke, at KUANA ændrer sig i forhold til denne udfordring på en kort tidshorisont.
- Cumulus kunne være i spil, som et pre-ingest system, - men dette vil kræve udvikling både med hensyn til at få styring af bitbevarede filer ind i KUANA og dels styring af kompliceret fletning af metadata, som på forskellig vis kan ændre sig i de to systemer.
- Der er en stor og voksende mængde data på pre-ingest, som det er vigtigt at få ingested (primært) i KUANA så hurtigt som muligt. Opsætning af workflows for dette kræver udviklingsressourcer.

Ud fra disse betragtninger er det vigtigt at holde fokus på at få bedre sikring af flere data end at lave en konvertering af et fungerende system til et system, hvor eksisterende processer besværliggøres. Endvidere er det ikke hensigtsmæssigt at bruge kræfter på kompliceret synkronisering med Cumulus som ingestsystem, da dette også vil betyde fjernelse af fokus fra at få bevaret de ophobede mængder af data. Dette skal også ses ud fra en betragtning om at data og metadata bevares på ensartet måde, uanset om de bevares via KUANA eller Cumulus.

De ophobede data vil primært skulle ingestes i KUANA, da dette er det strategisk valgte hovedsystem til understøttelse af bevaring. Der vil dog være data, som har afhængigheder til formidlingsworkflow, der er bundet op på Cumulus, eller data som afhænger af funktionaliteter i Cumulus (primært billeder). I sådanne situationer afgør funktionsgrupperne for bevaring og adgang, hvilken placering der vil være mest fordelagtig.

For at kunne skabe bedre overblik over bevaringen på en måde, som er uafhængig af understøttende systemer (ikke netarkiv), er der behov for at kunne tilgå alle bevarede metadata samlet i standardiseret og normaliseret form. Dette er skitseret som et metadata-warehouse. Mere specifikt kan dette understøtte:

- et samlet overblik over alle bevarede metadata i normaliseret form, fx til overblik over filformater der skal dækkes af bevaringsadministrationen
- fastholdelse af systemuafhængighed for bitbevarede data og metadata til understøttelse af systemer med statistik og overbliksfunktionalitet, som ikke understøttes andre steder
- bedre mulighed for at lave exit-strategier for proprietære systemer med bevaringsdata
- at fungere som en aggregator for **alle** bevarede metadata
    - med minimum af kompleksitet i forbindelse med, at der kører flere repository-systemer, der understøtter bevaring

    - med al historik for metadata (fra alle systemer de har eksisteret i), hvilket højst sandsynligt ikke vil være i et eksisterende system

Generelt for alle eksisterende bevaringsløsninger er der forskellige lokalt udviklede metadata-formater, hvilket også er tilfældet for Cumulus og KUANA (og SB-DOMS). Som beskrevet i metadata-afsnittet, er der behov for at få så standardiserede og systemuafhængige metadata som muligt. Den vedvarende standardiserede metadata-struktur (som beskrevet i afsnit om datamodel) er den, som skal være repræsenteret i metadata-warehouse. Dermed fås en ensartet basis for udtræk af informationer om bevarede metadata.

Bevaringsadministration med bevaringsplaner vil være baseret på data fra metadata-warehouset, hvor det er relevant.

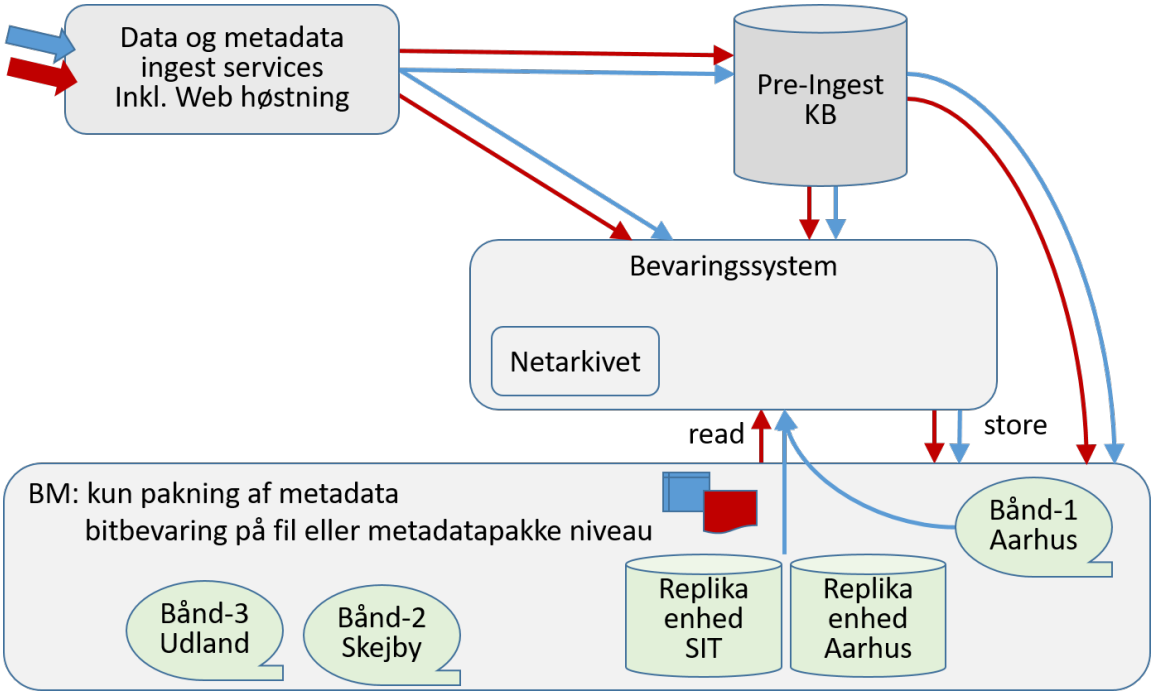
For at ensarte bitbevaring og interface til bitbevaring er det strategien, at:

- Selve bitbevaringsoperationerne fremadrettet udelukkende skal baseres på filer, uanset om filerne er pakket i WARC eller ej. Det skal bemærkes, at dette ikke gælder metadata (som fortsat skal pakkes) og heller ikke høstede Netarkiv-data (da WARC-formatet indeholder de nødvendige metadata om et web-element, for at elementet senere kan tilgås som webdata).
- Fremadrettet skal filer ikke pakkes i WARC (med ovennævnte undtagelser). Det vides ikke, om der kommer fremtidige scenarier, som vil kræve pakning med den tids teknologi, derfor skal dette løbende evalueres og kun revideres, såfremt der er behov for det. Det skal overvejes, om de nuværende pakkede filer (ikke undtagelserne) skal udpakkes på et senere tidspunkt som led i harmoniseringen af måder, det Kgl. Bibliotek bitbevarer på.
- Identifiers, som bruges i bitbevaringsoperationer og i interface, er i dag enten UUID'er eller identifiers konstrueret af fastfrossen metadata (eller høstningsinformationer). Disse valg tjener forskellige formål, hvor hidtidige analyser endnu ikke kan pege på en harmoniseret løsning. Strategien er derfor at ignorere forskellene på identifiers og forsætte analysearbejdet, som senere kan danne baggrund for vedtagelse af en mere harmoniseret løsning.

Bitmagasinet vil udelukkende være baseret på bitrepository.org frameworket uden afkoblede backups.

Den ønskede arkitektur (visionen)

Visionen for Det Kgl. Bibliotek er at have ét (på overfladen) bevaringssystem, med underliggende ingest-system og bitbevaringssystem. Ensartetheden skal bidrage til at give overblik og et fast udgangspunkt for overvågning, planlægning og udførelse af de nødvendige bevaringsaktiviteter. Systemet vil højst sandsynligt indeholde flere delsystemer for at kunne understøtte forskellighederne, som også vil betyde forskelligheder i behandlingen. Dette er illustreret herunder.



I den ønskede arkitektur er der stadig en Pre-ingest. Dette skyldes, at der realistisk set altid vil være behov for Pre-ingest, som det er forklaret i afsnittet om fremtidig arkitektur. Visionen er, at Pre-ingest ikke kun er under back-up, men også har tilknyttet midlertidig bitbevaring, indtil materialet er ingested i bevaringssystemet.

Netarkivet er i denne illustration set som en del af bevaringssystemet, da indsamling i stigende grad også sker via høstning af internettet, og der derfor med høj sandsynlighed vil blive et behov for et samspil mellem Netarkivet og det øvrige arkiv.

Forskning, videndeling og kompetenceudvikling

Vision:

- Det Kgl. Bibliotek ønsker at fastholde et højt forsknings- og vidensniveau inden for digital bevaring for at sikre bedst mulig bevaring af det digitale materiale

Mål:

- Fortsat aktiv tilstedeværelse i relevante faglige organisationer som fx OPF, RESAW og IIPC
- Deltagelse i relevante konferencer og workshops, og optagelse af mindst et bidrag på iPRES-konferencen hvert år
- Løbende præsentation af Det Kgl. Biblioteks erfaringer og viden inden for digital bevaring til det faglige, internationale netværk via webinarer, artikler og lignende
- Løbende præsentation af Det Kgl. Biblioteks erfaringer og viden inden for digital bevaring til et bredere, nationalt publikum via medierne og digitalbevaring.dk

Det Kgl. Bibliotek holder fokus på videndeling og kompetenceudvikling inden for digital bevaring ved blandt andet at deltage aktivt i faglige organisationer såsom IIPC og OPF og i relevante konferencer, workshops og lignende, hvor medarbejderne dels kan være medarrangører, dels fremlægge egne erfaringer med digital bevaring via papers og posters, og dels holde sig opdaterede med udviklingen inden for feltet. Den vigtigste konference inden for digital bevaring er iPRES, hvor Det Kgl. Bibliotek så vidt muligt er repræsenteret med et indlæg hvert år.

Biblioteket er og vil fortsat være en anerkendt spiller inden for international digital bevaring og bidrager gennem forskning og andre aktiviteter til at skabe bedre standarder, bevaringshandlinger, værktøjer, metoder og best practices, mv., som er med til også at forbedre bibliotekets eget arbejde inden for digital bevaring.

Biblioteket har som en af de førende danske institutioner inden for digital bevaring også en forpligtelse til at formidle egne og andres resultater i Danmark og på dansk, hvilket bl.a. sker gennem hjemmesiden digitalbevaring.dk i samarbejde med Rigsarkivet, og gennem interviews, avisartikler og networking.

Samarbejde om bevaringsaktiviteter

Vision:

- Det Kgl. Bibliotek samarbejder nationalt og internationalt om bevaringsaktiviteter, i det omfang det er ressourcemæssigt og økonomisk fordelagtigt.

Mål:

- I 2019 overtager Statens IT driften af en replikaenhed i bitmagasinet for Det Kgl. Bibliotek for at sikre organisatorisk uafhængighed i bitbevaringen
- I 2023 har Det Kgl. Bibliotek et udvidet samarbejde med en udenlandsk partner om bitbevaring
- Det Kgl. Bibliotek deltager aktivt i udviklingen af software under OPF, i Preservica og evt. andre relevante communities
- Det Kgl. Bibliotek deltager i samarbejde om udvikling af standarder, som er vigtige for biblioteket (se afsnit om standarder)

Samarbejdsaktiviteter om digital bevaring skal udvælges og prioriteres ud fra det overordnede princip, at de skal øge Det Kgl. Biblioteks muligheder for at udføre sine forpligtelser i forhold til langtidsbevaring af Danmarks kulturarv og for at styrke Det Kgl. Biblioteks internationale placering inden for digital bevaring. Sådanne samarbejder kan blandt andet omfatte driftfællesskaber, softwareudvikling, software user communities, og kompetenceudvikling og -udveksling.

Trustworthy Digital Repository

Vision:

- Det Kgl. Bibliotek vil opfattes som et Trustworthy Digital Repository

Mål:

- Der gennemføres et selv-audit af bibliotekets digitale bevaring inden udgangen af 2021
- Selv-audit laves som minimum hvert fjerde år

Det Kgl. Bibliotek skal kunne demonstrere, at organisationen opfylder de nødvendige tekniske, organisatoriske og økonomiske kriterier for at overholde sine forpligtelser til at bevare det danske kulturarv nu og for eftertiden.

Med tre-fire års mellemrum gennemføres en selv-audit proces, som dokumenterer - over for biblioteket selv og øvrige interessenter (fx offentlige myndigheder og internationale partner) - at man lever op til kravene til et Trustworthy Digital Repository (Troværdigt Digitalt Arkiv). Audit-processen omfatter både det digitale materiale, teknisk infrastruktur, organisation og økonomi. Selv-audit processen iværksættes og styres af Funktionsgruppen for Digital Bevaring, som også har ansvar for koordinering af arbejdet og færdiggørelse af den endelige audit-rapport, som skal behandles og godkendes i styregruppen jvf. organisation. Selv-audit udføres som en gennemsigtig og offentligt-dokumenteret proces, der lever op til anerkendte internationale standarder og praksis, herunder *ISO16363:2012 Space data and information transfer systems -- Audit and certification of trustworthy digital repositories.*

Organisation

Vision:

- Det Kgl. Bibliotek ønsker at skabe en robust organisatorisk forankring af arbejdet med digital bevaring, således at bevidstheden om og ansvaret for denne opgave opbygges og fastholdes som en af husets kerneopgaver

Mål:

- Ved udgangen af 2019 har Det Kgl. Bibliotek organiseret arbejdet med digital bevaring således, at der er klare ansvarsfordelinger mellem de samlingsansvarlige, funktionsansvarlige og IT-ansvarlige på biblioteket

Det Kgl. Biblioteks arbejde med digitale samlinger finder sted i et samarbejde mellem en række afdelinger og funktionsområder, der hver især har ansvar for dele af de digitale samlingers livscyklus. Arbejdet med digital bevaring finder sted i alle "led" i denne cyklus, fra udvælgelse og indlemmelse af materiale til udførelse af "rene" bevaringshandlinger, som fx formatmigrering. Derfor er det afgørende, at der tages ansvar for den digitale bevaring alle steder i organisationen.

For arbejdet med det digitale materiale er oprettet funktionsgrupper, som arbejder med hhv. indsamling, bevaring og adgang. På tværs af funktionerne er udpeget materialeansvarlige, som hver især er vidensholdere for en bestemt materialetype og tager sig af den generelle datahåndtering og rådgivning på tværs af hele materialets livscyklus.

Funktionsgruppen for digital bevaring er bemandet på tværs af de afdelinger, som til dagligt arbejder med digital bevaring, dvs. DKU og ITU. Gruppen er det sted, man som ansat på Det Kgl. Bibliotek henvender sig, hvis man har spørgsmål til arbejdet med digital bevaring. Gruppens funktioner er at skabe overblik over den digitale bevaring på Det Kgl. Bibliotek og sikre, at politik og strategi for digital bevaring udmønter sig i, at beslutninger om digital bevaring kan træffes på et kvalificeret grundlag. Samtidig sikrer gruppen, at der findes nedskrevne procedurer for de aktiviteter, der handler om digital bevaring.

Gruppen træffer således de faglige beslutninger om digital bevaring. I tilfælde, hvor dette ikke er muligt på grund af fx de økonomiske eller ressourcemæssige konsekvenser ved en foretrukket beslutning, eskaleres diskussionen og afgørelsen til styregruppen. Denne består af sektions- og afdelingsledere samt vicedirektører for de relevante organisatoriske enheder for digital bevaring.

Funktionsgruppen for digital bevaring fungerer samtidig som et underudvalg til Det Kgl. Biblioteks Sikringsudvalg. Risikovurderinger og sikkerhedsmæssige hændelser rapporteres hertil, ligesom disse rapporteres til styregruppen, som også inddrages, hvis er er risici i forhold til fx økonomi eller datatab.

Funktionsgruppen har et tæt samarbejde med materiale-, og samlingsansvarlige og de tilsvarende grupper for hhv. Indsamling og Adgang, ligesom IT Drift, styregruppen og Sikringsudvalget også er vigtige samarbejdspartnere.

Administration af dokumentet

Nærværende dokument er første version af Det Kgl. Biblioteks Strategi for digital bevaring. Denne strategi erstatter de tidligere strategier for hhv. Det Kongelige Bibliotek [1], Netarkivet [2] og Statsbiblioteket [3].

Opdatering af dokumentet

Opdatering af dokumentet sikres af Funktionsgruppen for digital bevaring og godkendes af styregruppen for denne. Dokumentet opdateres hvert tredje år eller med højere frekvens, hvis der er behov på grund af ændrede rammer for digital bevaring.

Formidling af dokumentet

Dokumentet publiceres via Det Kgl. Biblioteks hjemmeside i en dansk og en engelsk version med det formål at understøtte nationalt og internationalt samarbejde om digital bevaring.

Referencer

1. Strategi for langtidsbevaring af digitalt samlingsmateriale på Det KongeligeBibliotek, 2014, urn:pwid:netarkivet.dk:2015-03-09Z17:27:08Z:part:http://www.kb.dk/export/sites/kb_dk/da/kb/downloadfiler/StrategiLangtidsbevaringDigitaltMateriale-KB-DK-2014.pdf
2. Strategi for langtidsbevaring af materiale indsamlet til Netarkivet ved Det Kongelige Bibliotek og Statsbiblioteket, 2014, urn:pwid:netarkivet.dk:2015-09-10T10:02:07Z:part:http://netarkivet.dk/wp-content/uploads/2015/03/Netarkivet_Strategi_Langtidsbevaring_1.0_150115.pdf
3. Statsbibliotekets strategi for digital bevaring, 2016, urn:pwid:netarkivet.dk:2016-12-30T13:20:04Z:part:http://www.statsbiblioteket.dk/nationalbibliotek/digital-bevaring/strategi-for-digital-bevaring

Bilag 1: Bitsikkerhedsniveauer

I beskrivelsen af niveauerne er det indforstået, at det tilstræbes, at eventuelle checksumskopier er geografisk adskilt fra hinanden og fra de fuldstændige kopier. I de tilfælde, hvor niveauet er middel eller højere, er det også indforstået, at én enkelt person ikke kan have adgang til et flertal af komplette kopier (dvs. fraregnet checksumskopier), ej heller til et flertal af samtlige kopier (checksumskopier såvel som komplette kopier).

Frekvensen for integritetstjek varierer for forskellige løsninger og er derfor ikke fast for et enkelt niveau. Årsagen er, at de forskellige medier, som indgår i løsningerne, kræver forskellige former for integritetstjek. Frekvensen af integritetstjek og checksumsberegninger lokalt på en kopi skal derfor fastsættes ud fra en risikovurdering og en omkostningsanalyse, som tager højde for medietype, mediets forventede levealder, samt statistik fra logninger på alle fejl.

Uafhængighed mellem kopier kan også være af forskellige grader, med forskellig betydning og forskellig sandsynlighed for tab alt efter hvilke andre uafhængigheder der er. Derfor er den specifikke grad af uafhængighed for forskellige områder ligeledes fastsat efter ud fra en risikovurdering og en omkostningsanalyse.

Nedenfor beskrives de forskellige bitsikkerhedsniveauer (ud fra basale principper). Hvert niveau indeholder en beskrivelse af

- den vedvarende forståelse af niveauet

og alle anvendte niveauer indeholder desuden en beskrivelse af

- fortolkning af niveau i form af krav til løsninger ud fra nuværende rammer og
- et eksempel på, hvornår niveauet kan benyttes.

Maksimal bitsikkerhed

Tager højde for alt andet end jordens undergang. Dette niveau benyttes ikke.

Meget høj bitsikkerhed

Tager højde for så mange risici som muligt inkl. krig, besættelse, større terroraktioner og større naturkatastrofer, men hvor der kun er beskeden sikring mod de sidstnævnte større risici.

En løsning, der kan give dette niveau af bitsikkerhed, kræver:

- Minimum 3 komplette kopier samt minimum 2 checksumskopier
- Organisatorisk, teknologisk, politisk og geografisk uafhængighed. Politisk uafhængighed kan opnås med udenlandsk kopi. Geografisk uafhængighed fastlægges ud fra en risikovurdering af, hvorvidt samme hændelse (fx terroraktion eller naturkatastrofe) vil kunne ramme alle kopier
- Hyppige tjek af integritet mellem kopier, og hyppige tjek hvor checksum genberegnes på de bitbevarede data

Tildeles i princippet alle materialer, som vil gå tabt, hvis den digitale udgave mistes og hvor tab af data vil have konsekvenser for den danske kulturarv og/eller bibliotekets renommé. Dette omfatter dermed fx alt digitalt født pligtafleveret materiale samt substitutionsdigitaliseret materiale, hvor den fysiske kopi vil forsvinde inden for en kort tidshorisont. Undtagelsesvis kan der vælges 'høj bitsikkerhed', hvis rammerne for bitbevaringen umuliggør valg af "meget høj bitsikkerhed", fx ved manglende mulighed for at lave aftale om en udenlandsk kopi, der opfylder de juridiske og sikkerhedsmæssige krav.

Høj bitsikkerhed

Tager højde for alle former for risici på nær krig, besættelse, store terroraktioner og store naturkatastrofer.

En løsning, der kan give dette niveau af bitsikkerhed, kræver:

- Minimum 3 komplette kopier samt minimum 2 checksumskopier
- Organisatorisk, teknologisk og geografisk uafhængighed, hvor mindst to komplette kopier er lagt med en afstand på mindst 150 km
- Hyppige tjek af integritet mellem kopier, og hyppige tjek hvor checksum genberegnes på de bitbevarede data

Tildeles de materialer, som i princippet burde tildeles "meget høj bitsikkerhed", men hvor rammerne for bitbevaring umuliggør det meget høje bitsikkerhedsniveau.

Middel bitsikkerhed

Tager højde for at de risici der kan forårsage tab holdes på et minimum, dvs. kun mindre tab kan tolereres.

En løsning, der kan give dette niveau af bitsikkerhed, kræver:

- Minimum 2 komplette kopier og 1 checksumskopi
- Organisatorisk, teknologisk og geografisk uafhængighed, hvor mindst to komplette kopier er lagt med en afstand på mindst 150 km
- Hyppige tjek af integritet mellem kopier, og hyppige tjek hvor checksum genberegnes på de bitbevarede data

Dette niveau af bitsikkerhed kan tildeles materialer, hvor der eksisterer en fysisk kopi under bevaring, hvorfra den digitale kopi kan genetableres, eller hvor tab af data er vurderet til af have lille konsekvens for biblioteket.

Lav bitsikkerhed

Tager højde for risici, der kan forårsage tab af mange data på en gang. Dvs. tab kan tolereres, fx fordi sandsynligheden for at kunne genskabe digitaliseret materiale fra andre kilder er relativt høj (og dermed mindre risiko for større re-digitaliseringsopgaver), eller at det er acceptabelt at skulle investere i re-digitalisering af større mængder data.

En løsning, der kan give dette niveau af bitsikkerhed, kræver:

- 1 komplet kopi med minimum 1 backupkopi
- Minimum en gang i kvartalet laves backup
- Geografisk uafhængighed mellem kopierne med en afstand på mindst 150 km

Tildeles digitaliserede materialer til formidling, hvor der eksisterer fysiske udgaver under fysisk bevaring, hvorfra det digitaliserede materiale kan genetableres.

Meget lav bitsikkerhed

Tager højde for, at et tab af en enkelt kopi ikke nødvendigvis betyder, at disse data er tabt (svarer til en normal backupløsning uden geografisk uafhængighed). Dette niveau benyttes ikke.

Minimal bitsikkerhed

Tager ikke højde for noget. Dette niveau benyttes ikke.

Bilag 2: Fortrolighedsniveauer

Nedenfor beskrives de forskellige fortrolighedsniveauer. Hvert niveau indeholder en beskrivelse af

- den vedvarende forståelse af niveauet
- fortolkningen af niveauet i form af krav til løsninger ud fra nuværende rammer og
- et eksempel på, hvornår niveauet kan benyttes

Det skal bemærkes, at de restriktioner, som gælder i forhold til fortrolighed, skal forstås akkumulativt. Dvs. restriktioner for et lavere niveau vil også gælde for det pågældende niveau. For eksempel er krav om, hvorvidt en læselig kopi kan lægges i udlandet, kun beskrevet som en mulighed under 'lav fortrolighed', men vil gælde for 'middel fortrolighed' til 'maksimal fortrolighed'.

Maksimal fortrolighed

Betyder, at det gøres så svært som muligt for uvedkommende at kende til eksistensen af og få adgang til læsning af data, og at kun meget få personer inden for biblioteket kan godkendes til at have adgang. Det vil på dette niveau være vigtigere at opretholde fortrolighed fremfor at sikre at data ikke mistes. Med andre ord, hellere miste data end bryde fortrolighed.

En løsning, der kan give dette niveau af fortrolighed, kræver, at:

- alle kopier er off-line
- der er et minimalt antal kopier (maksimalt to kopier suppleret med mindst en checksum)
- alt er krypteret
- krypteringsnøglen skal bevares på en måde, hvor procedurer sikrer, at adgang til nøglen kun kan ske efter godkendelse på direktionsniveau

Tildeles de materialer, som er vurderet til, at kendskab til eksistensen kan have konsekvenser for nationale sikkerhedsaspekter, personer omtalt i materialet eller deres efterkommere.

Meget høj fortrolighed

Betyder, at det skal gøres så svært som muligt for uvedkommende at få adgang til læsning af data, og at kun meget få personer inden for biblioteket kan godkendes til at have adgang. Fortrolighed skal sikres i så høj grad som muligt, hvor der stadig tages højde for at data ikke må mistes. Med andre ord, hellere bryde fortrolighed end miste data.

En løsning, der kan give dette niveau af fortrolighed, kræver, at:

- det digitale materiale bitbevares i krypteret form under det ønskede bitbevaringsniveau og sikret på samme niveau som materialer med 'middel fortrolighed'
- en ukrypteret komplet kopi gemmes in-house, off-line og fysisk sikret - suppleret med minimum to uafhængige checksummer

- sikring mod uvedkommende adgang til læsning af data er maksimal for den ukrypterede kopi
- krypteringsnøglen skal bevares på en måde, hvor procedurer sikrer, at adgang til nøglen kun kan ske efter godkendelse på direktionsniveau

Tildeles materialer, hvor brud på fortroligheden vil få store konsekvenser. Dvs. dette omfatter fx donationer fra kulturpersonligheder, hvor der er krævet fortrolighed i en årrække, og hvor brud på fortroligheden kan have konsekvenser for personer omtalt i materialet eller for kulturpersonens efterkommere.

Høj fortrolighed

Betyder, at det skal gøres så svært som muligt at få fat i bits for personer, som ikke er godkendt til det. Godkendte personer behøver ikke nødvendigvis at være ansat ved biblioteket. Adgangsrestriktioner i forhold til materialet må ikke ske på bekostning af bitsikkerheden. Med andre ord, hellere bryde fortrolighed end miste data.

En løsning, der kan give dette niveau af fortrolighed, kræver, at:

- alle komplette kopier er sikret, og at adgang kun tildeles personer, der er godkendt hertil
- adgangsrestriktioner for personfølsomme data er overholdt
- der er maksimalt tre komplette kopier af data

Tildeles materialer, hvor der Det Kgl. Bibliotek er forpligtet til at behandle data fortroligt, men hvor det er muligt at give borgere godkendt adgang. Et eksempel er Netarkivet, som er fortroligt i den forstand, at det indeholder personfølsomme data og kan bruges til sammenkædning af data med konsekvenser for borgere, men som samtidig skal kunne stilles til rådighed for forskningsprojekter.

Middel fortrolighed

Betyder, at der er lavet tiltag, så uvedkommende ikke kan få adgang til data.

En løsning, der kan give dette niveau af fortrolighed, kræver, at:

- adgang til kopierne er sikret med speciel adgangsbegrænsning
- krypterede kopier eller checksumskopier må placeres i udlandet. Placering af læsbare kopier i udland skal der gives dispensation for

Tildeles materialer, som typisk kun skal kunne tilgås af KB-personale, eller hvor adgang til materialerne styres af KB-personale.

Lav fortrolighed

Betyder, at der er lavet tiltag, så uvedkommende ikke umiddelbart kan få adgang til data.

Løsninger er lavet med almindelig adgangsstyring til kopierne

Tildeles materialer, hvor biblioteket har ret til at give adgang til borgere. Det bør bemærkes at det er uvedkommende for fortrolighedsniveauet i bitbevaringen om der eventuelt er restriktioner i forhold til, hvordan data stilles til rådighed.

Meget lav fortrolighed

Betyder at det er åbent for forskellige større grupper, såsom et lukket forum f.eks. et intranet.

Anvendes ikke i forbindelse med bevaring (bruger 'lav fortrolighed' da der altid vil være adgangsstyring til bitbevarede data)

Minimal fortrolighed

Betyder at der ikke er nogen krav og materialet er offentligt tilgængeligt

Anvendes ikke i forbindelse med bevaring (bruger 'lav fortrolighed' da der altid vil være adgangsstyring til bitbevarede data)

Bilag 3: Detaljeret postkort datamodel eksempel

Et detaljeret eksempel af et digitaliseret postkort er givet i dette bilag. Denne digitalisering er i første omgang kun lavet for forsiden af postkortet.

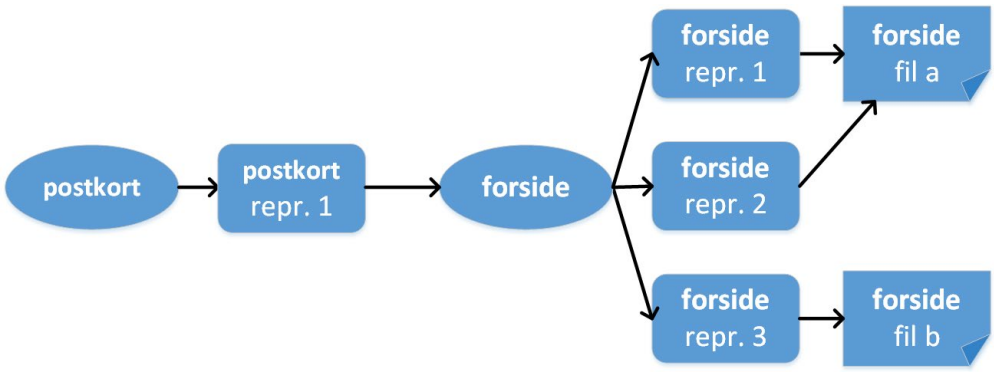
Selve digitaliseringen er lagt i Filen "forside fil a". De metadata, som stammer fra fx karakterisering eller digitaliseringen, er lagt i Repræsentationen "forside repr. 1". Det digitale objekt for forsiden er modelleret som den Digitale Intellektuelle Entitet "forside". Selve postkortet med denne forside er repræsenteret i Repræsentationen "postkort repr. 1", hvor det digitale objekt for postkortet er modelleret som den Digitale Intellektuelle Entitet "postkort".

I dette eksempel er der gennem tiden forskellige opdateringer til forsiden, hvilket er illustreret i den næste figur.

Den første ændring er en ændring af metadata, fx i form af re-karakterisering. Der er ikke nogen ændringer til selve filen, og det er derfor en ny Repræsentation "forside repr. 2" af den Digitale Intellektuelle Entitet "forside", hvor Repræsentation "forside repr. 2" peger på den samme fil som Repræsentation "forside repr. 1".

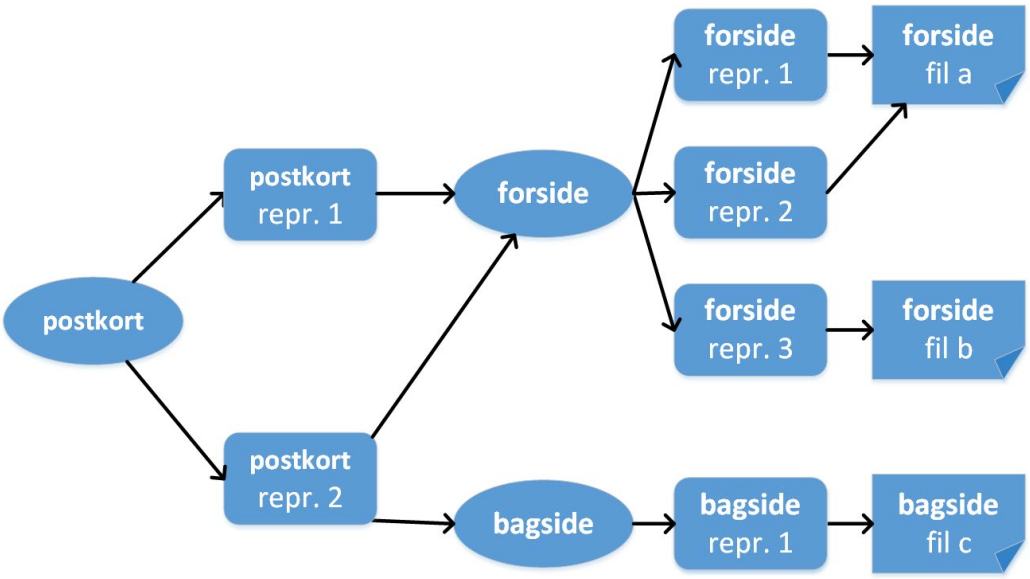
Den anden ændring, der sker, er, at der er opdateringer til filen (fx en gendigitalisering eller filmigrering). Da det er en helt ny fil "forside fil b", vil der også være nye metadata til filen. Derfor er der en ny Repræsentation, "forside repr. 3", som også repræsenterer den Digitale Intellektuelle Entitet "forside".

Det skal bemærkes, at dette ikke påvirker postkortets Digitale Intellektuelle Entitet og Repræsentation. Dette er grunden til, at det er nødvendigt at bruge det Digitale Intellektuelle Entitetsniveau for forsiden. Hvis postkortets Repræsentation pegede direkte på forside-repræsentationen, ville der også skulle laves en ny Repræsentation for postkortet, og alle andre Repræsentationer, som måtte pege på Repræsentationer af forsiden.



Der kan også være eksempler på ændringer, som kun vedrører postkortet og ikke selve forsiden. I det simple tilfælde kunne det være opdatering af bibliografiske metadata til postkortet, som ville resultere i en ny Repræsentation (denne er ikke illustreret i figuren).

En mere avanceret ændring af postkortet ville være, hvis bagsiden sidenhen også blev digitaliseret (fx i tilfælde hvor der er relevante noteringer på bagsiden). Dette er illustreret i nedenstående figur, hvor der kommer en ny Repræsentation "postkort repr. 2" af den Digitale Intellektuelle Entitet "postkort". Denne peger på både Digitale Intellektuelle Entitet for "forside" og en ny Digitale Intellektuelle Entitet "bagside", som er skabt for for den nye digitalisering af bagside, som er lagt i filen "bagside fil c". Repræsentation "postkort repr. 2" kan da også indeholde metadata om forsiden og bagsidens indbyrdes relation.



Der er en masse identifiers involveret i denne model. Den vigtigste af disse er identifieren for de Digitale Intellektuelle Entiteter, da det er denne, der vil være den samme over tid for det digitale objekt repræsenteret på forskellige måder over tid. Eventuelle alternative identifiers vil ligge i metadata og er derfor henvist til repræsentationerne i denne model. Det skal her bemærkes, at dette ikke burde være noget problem, idet et indeks for identifier og alternative identifiers til enhver tid kan genetableres fra informationerne i datamodellen.

Bilag 4: Detaljerede implementeringseksempler

Dette bilag indeholder to implementeringseksempler for brug af den generelle datamodel og modellen for metadata:

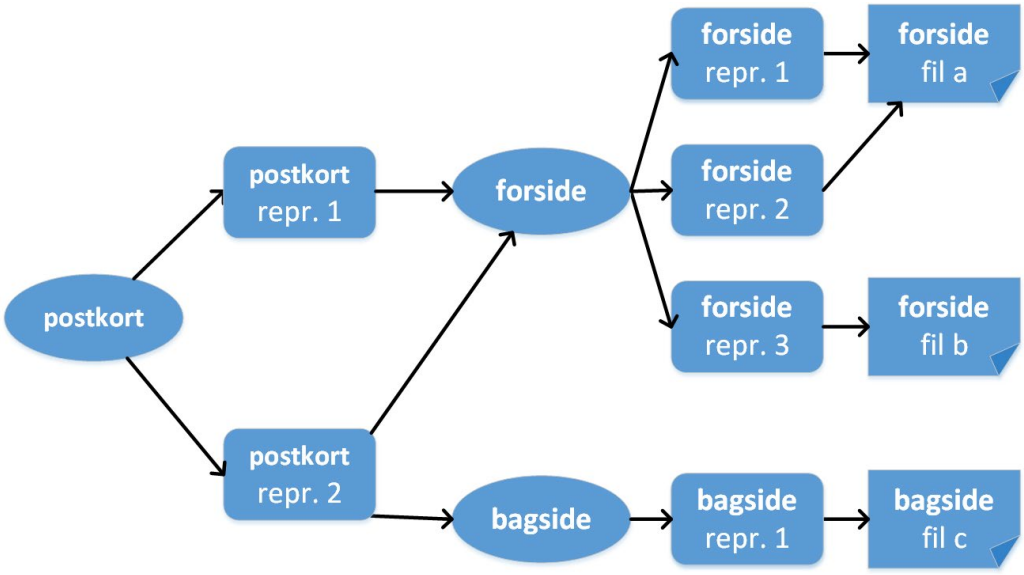
- et generelt eksempel
- et eksempel ud fra et KUANA perspektiv

Grunden til, at der kun er angivet eksempel fra KUANA, er, at datamodellen allerede er implementeret for data i Cumulus, - og SB-DOMS vil blive konverteret til Kuana.

Generelt implementeringseksempel

Dette detaljerede eksempel bygger på eksemplet med for- og bagside af et digitaliseret postkort, som er beskrevet i Bilag 3: Detaljeret postkort datamodel eksempel. Dette er opstået gennem følgende tidsforløb:

1. Den første digitalisering af forsiden er lagt i filen "forside fil a" og med metadata i repræsentationen "forside repr. 1".
2. Senere har der været opdatering af metadata, som er repræsenteret i "forside repr. 2".
3. Forsiden er senere igen blevet gendigitaliseret, hvor resultatet er lagt i "forside fil b" og "forside repr. 3" (med nye metadata - fx. de tekniske metadata).
4. Endnu senere er bagsiden også digitaliseret, og resultatet lagt i "bagside fil c" og "bagside repr. 1", hvilket også har betydet en ny repræsentation af postkortet, da dette nu har både for- og bagside.



For hvert step på tidslinjen er der blevet genereret hver sin del af metadata, som er afbilledet i tegningen. Filer er bevaret for sig og kan identificeres via deres identifier, hvor checksummen for filen i metadata kan anvendes, hvis sammenhæng mellem identifier for fil og filen selv er gået tabt. Metadata bygges op efter principperne beskrevet i strategien og er pakket i WARC.

De WARC-records, som bitbevares for de producerede metadata, er angivet for hvert step i de følgende underbilag:

- WARC-records for første digitalisering af forsiden til postkort
- WARC-records for opdatering af forside metadata
- WARC-records for gen-digitalisering af forsiden
- WARC-records for tilføjelse af bagside til postkort

Det skal bemærkes, at der i eksemplerne er en ekstra WARC-record, som binder Digitale Intellektuelle Entiteter sammen med repræsentationer (og evt. filer). Disse er i princippet overflødige, idet de ville kunne skabes ved gennemlæsning af arkivet. Deres eksistens er derfor blot optimering, så det vil være nemt at lave et indeks, hvor alle repræsentationer for en Digital Intellektuel Entitet kan findes (søgning på identifier for den Digital Intellektuel Entitet, hvor datoerne for dem vil være forskellig for hver af repræsentationerne).

Eksempel ud fra et KUANA perspektiv

I dette afsnit bruges postkort-eksemplet i KUANA, og derefter beskrives de muligheder, der er for at udtrække og bitbevare dette KUANA-eksempel. Endvidere er det beskrevet, hvordan specialtilfælde i KUANA håndteres.

Generelt om forskelle mellem datamodellen og KUANA

Eftersom både Preservicas datamodel og denne datamodel er inspireret af datamodellen fra EU Planets-projektet (afsluttet i 2010), er der også store ligheder mellem KUANAs datamodel og den generelle datamodel. Dog er der brugt forskellig terminologi, og der er mindre forskelle i betydningen.

Preservica har en anden terminologi og lidt anden betydning for entiteterne i deres datamodel, end det er beskrevet for den generelle datamodel, som anvendes for bitbevarede data. Der er dog mange lighedspunkter, da de tager udgangspunkt i nogle af de samme tidligere datamodeller (fx fra Planets-projektet).

En af de store forskelle på datamodellerne er, at Preservicas datamodel kun tager højde for ændringer i filer og dermed ikke omfatter ændringer i metadata.

Termerne er:

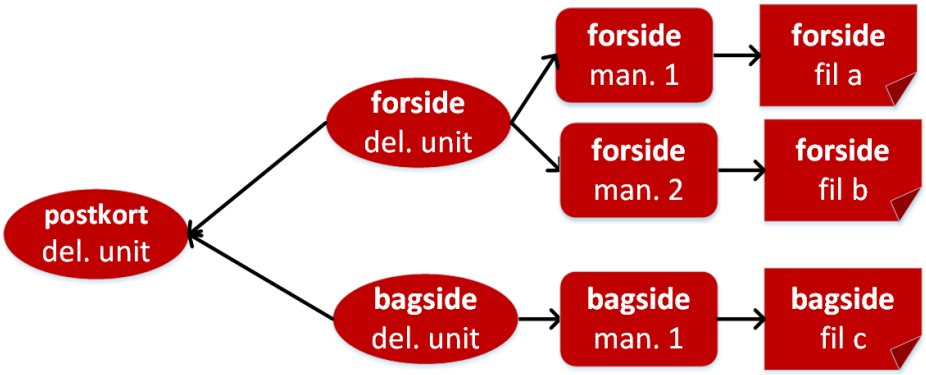
- **Filer** i Preservica indeholder både fil og metadata til filen, hvor metadata ændre sig. Mapningen til den generelle model gøres ved at se selve filen som en fil og metadata (og evt. ændrede metadata) som Repræsentationer for filerne for den samme Digitale Intellektuelle Entitet, som repræsenterer filens udformning over tid.

- **Manifestationer** i Preservica svarer nogenlunde til Repræsentationer i den generelle model. Der er dog to punkter, hvorpå de adskiller sig. Det første er, at metadata kan ændre sig. Det andet punkt er, at Preservica kun kan have lineær versionsudvikling, dvs. der kan ikke være to forskellige, 'aktive' bevaringsrepræsentationer, som det fx kan være krævet for en powerpoint, der både migreres til pdf (for at bevare egenskaber for udseende) og til en nyere version af powerpoint (for at bevare egenskaber i forbindelse med animationer, lyd med mere).

- **Deliverable Units** i Preservica bruges for ingested materiale og svarer nogenlunde til datamodellens Digitale Intellektuelle Entiteter. Dog kan en Deliverable Unit have metadata, og disse metadata kan ændre sig over tid. En Digital Intellektuel Entitet kan ikke ændre sig og kan derfor heller ikke have metadata. Metadata for Deliverable Units og ændringer i disse over tid vil i stedet være repræsenteret i Repræsentationer af den tilsvarende Digitale Intellektuelle Entitet. Normalt består en Deliverable Unit af manifestationer, men i interfacet kan det se ud, som om den også kan indeholde filer. I praksis skaber Preservica manifestationer for sådanne filer, og derfor er der ingen forskel, så længe en Deliverable Unit kun består af en fil. Hvis en Deliverable Unit består af flere Deliverable Units og filer, da er der kun et implicit Deliverable Unit lag for filerne.
- **Collections** i Preservica er samlinger af Deliverable Units, hvor en Deliverable Unit kan tilhøre én, og kun én, Collection. Ud over denne restriktion, fungerer Collection på samme måde som en Deliverable Unit, der udelukkende består af Deliverable Units.

Det generelle eksempel som det kunne se ud i KUANA

Eftersom Preservica tillader ændringer i metadata på alle entiteter i datamodellen (Collections, Deliverable Units, Manifestationer og Filer), så ser datamodellen en del enklere ud. Derfor vil ovenstående digitaliseringseksempel se ud som følger i en KUANA datamodel.

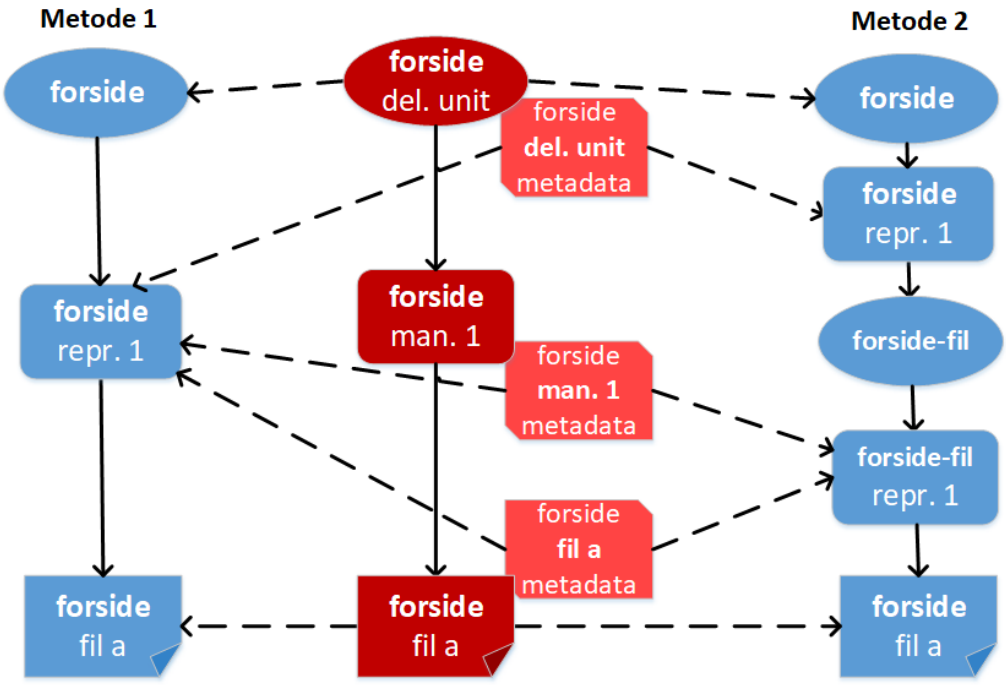


Hvis dette sammenlignes med ovenstående figur af den generelle model, så fremgår det, at "forside Man. 1" dækker over både "forside repr 1" og "forside repr 2". På samme måde kan det ikke ses i Preservica, at postkortet har ændret sig ved at få tilføjet en bagside.

De manglende informationer i modellen kan dog udtrækkes fra KUANA baseret på oplysninger, som KUANA kan give om ændrede metadata. Den sværeste del vil her være at kunne se, at postkortet har fået tilføjet en bagside, eftersom Preservicas datamodel kun indeholder disse informationer på de underliggende elementer - i dette tilfælde bagsiden.

Transformation for repræsentationer med filer

Denne lettere del med at transformere KUANA metadata og information om ændringer i metadata til normaliserede data i den generelle datamodel er illustreret i nedenstående figur. Der vil i praksis være to måder at gøre dette på. Valg af metode vil afhænge af, hvilke metadata der er lagt på de forskellige Preservica entiteter.



Normalt vil metadata på de forskellige Preservica entiteter være forskellige: på Fil-entiteter vil der primært ligge de af Preservica genererede tekniske metadata for filen (fx size, id, format, filnavn, checksum), og det er ikke muligt umiddelbart at se, hvilke af disse metadata der hører til Filen, og hvilke der hører til i Manifestationen, - på den anden side er de beskrivende metadata normalt lagt på Deliverable Unit, da man sjældent ønsker, at disse skal gentages for hver manifestation. Derfor er den illustrerede metode 1 sandsynligvis den metode, der vil bruges mest.

I de tilfælde, hvor der er samme type metadata på Deliverable Unit og Manifestation, vil dette kunne repræsenteres ved at indskyde et ekstra niveau i den generelle model - hvilket er det, der gøres i den illustrerede metode 2.

På basis af dette KUANA eksempel vil der blive transformeret til den generelle model og udtrukket metadata til bitbevaring.

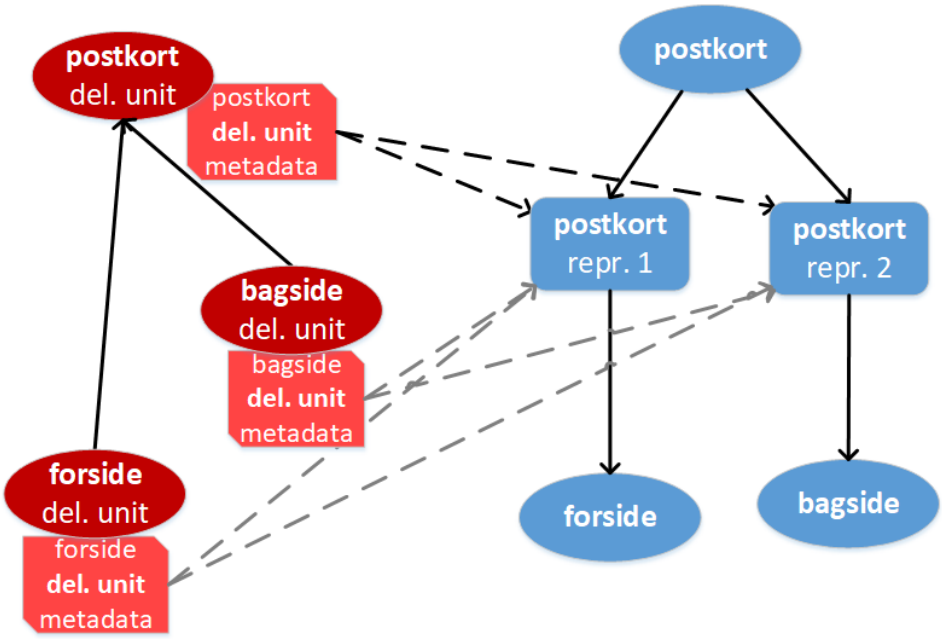
De resulterende bitbevarede WARC-records vil for Metode 1 resultere i samme WARC-filer som for det generelle eksempel, bortset fra at metadata om fx karakteriseringen af filen vil have informationer om hvilke værtøjer KUANA har brugt samt hændelsesinformationer om at metadata er hentet fra KUANA. Selve klumperne af beskrivende metadata ligger for sig i KUANA og kan derfor placeres ind i METS strukturen, Informationer om bevaringsniveau udregnes på basis af KUANAs storage policy, og de strukturelle metadata udregnes på basis of relationer i KUANA eller er direkte specificeret som metadata i KUANA.

Metode 2 vil ikke give meget mening i dette simple eksempel, men vil bestå i de samme principper som Metode 1, hvor der indføres et ekstra niveau og beskrivende "forside del. unit metadata" lægges i forside repræsentationen, mens beskrivende metadata fra manifestationen og filen lægges på repræsentationen forside-fil, sammen med tekniske

metadata og bevaringsmetadata for filen.

Transformation for repræsentationer uden filer

Den følgende figur illustrerer transformation af KUANA informationer for postkort svarende til en Deliverable Unit med andre Deliverable Units.



I Preservica er der kun informationer om, hvad en Deliverable Unit tilhører (Collection eller Deliverable Unit), ikke omvendt. Dvs. for at få metadata om, hvad et postkort består af, så skal man udtrække historik og relationer for de Deliverable Units, der peger på postkort. Dette er angivet med grå streger i figuren (metadata for for- og bagside er del af deres respektive Repræsentationer som beskrevet for forside ovenfor). De resulterende bitbevarede WARC-records vil svare til repræsentationerne i for det generelle tilfælde, bortset fra små forskelle, fx at metadata er taget fra KUANA.

Transformationer af specialtilfælde i KUANA datamodellen

Der er to specialtilfælde

- Transformation af Collections
- Filer, der ser ud til at ligge direkte under en Deliverable Unit

For Collections er dette meget nemt, da der modelleringsmæssigt ikke er nogen forskel på Collections og Deliverable Units, der kun indeholder andre Deliverable Units. Derfor laves Collections på samme måde som beskrevet for postkort ovenfor.

For filer, der ser ud til at ligge direkte under en Deliverable Units, så er dette kun et spørgsmål om, at Preservica illustrerer filerne på denne måde. I praksis er der manifestationer til filerne, som der findes informationer om via API'et til Preservica. Disse tilfælde kræver lidt ekstra omhyggelighed, men kan løses på samme måde som vist for forside - evt. med ekstra niveauer af Deliverable Units, såfremt der er flere "filer".