DET KGL. BIBLIOTEK
Royal Danish Library

Det Kgl. Biblioteks

Strategi for digital bevaring

10. april 2024

Indhold

Indledning 1
Relation til andre dokumenter 1
Hvad strategien omfatter 1
Formål og vision 1
Baggrund 2
Strategiske mål 2
Organisation 2
Risikostyring 3
Trustworthy Digital Repository 4
Teknologiovervågning 5
Standarder 6
Åben og proprietær software 7
Indlemmelse 8
Bevaringsprincipper og -processer 9
Bevaringsplaner 9
Bevaringsplaner for bitbevaring 10
Bevaringsplaner for emulering 11
Bevaringsplaner for migrering 11
Bitbevaring 12
Pre-ingest 13
Overordnet om implementering af bitbevaring 13
Indhold af bitbevaringsplaner 15
Nuværende implementering af bitbevaring for samlingsgrupperne 16
Ny implementering af samlingsgrupper og collections 18
Vedligeholdelse og oprettelse af collections 20
Logisk bevaring 20
Migrering og emulering 20
Datamodel, metadata og dataformater 21
Datamodel 22
Grundlæggende datamodel 22
Metadata 24
Bitbevaring af metadata 24

Metadata i forbindelse med bevaring 25
Dataformater 26
Teknisk infrastruktur 26
Den nuværende arkitektur 27
Arkitektur inden for en 3-5-årig tidshorisont 29
Den ønskede arkitektur (visionen) 31
Forskning, videndeling og kompetenceudvikling 32
Samarbejde om bevaringsaktiviteter 33
Administration af dokumentet 33
Opdatering af dokumentet 33
Formidling af dokumentet 33
Referencer 34

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.

Relation til andre dokumenter

Strategien udmønter de rammer og principper, der udstikkes i Det Kgl. Biblioteks Politik for digital bevaring. Disse danner baggrund for aktive beslutninger om bevaringsrelaterede handlinger og procedurer.

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 Accessionspolitik og Adgangspolitik. Tilsammen skal dokumenterne ses som en større helhed, der viser bibliotekets arbejde med det digitale materiale.

Hvad strategien omfatter

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

Bevaringsarbejdet med de digitale kulturarvssamlinger – og dermed også strategien – omfatter både organisation og økonomi, relevante it-systemer og tilhørende software, samt selve de digitale objekter bestående af data og metadata.

Strategien revideres som minimum hvert femte år for at sikre, at der tages højde for de interne og eksterne forandringer, som uvilkårligt vil ske.

Formål og vision

Digital bevaring er en nødvendig forudsætning for at kunne give adgang til digitale samlinger på langt 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 fungerer strategien som reference for det daglige arbejde med digital bevaring.

Visionen er at sikre bibliotekets digitale materialer bedst muligt, så vores digitale kulturarv også er tilgængelig for fremtidige generationer.

Baggrund

Det Kgl. Bibliotek er resultatet af en fusion i 2017 mellem 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 lavede en politik og strategi for digital bevaring. Nærværende strategi er første revision af Strategi for digital bevaring fra januar 2019.

Strategiske mål

1. **I 2027 er alle pligtafleverede data, som er modtaget før 2023, under bevaring**
   Dette betyder, at data er bitbevarede, samt at der er etableret overvågning og fyldestgørende bevaringsplaner for samlingerne
2. **I 2025 opfylder al bitbevaring krav til geografisk, organisatorisk og softwaremæssig uafhængighed**
   Dog er etablering af en udenlandsk kopi ikke omfattet af dette mål
3. **Det Kgl. Bibliotek fastholder sin status som anerkendt samarbejdspartner og bidragyder inden for digital bevaring**
   Dette indebærer, at biblioteket bidrager forskningsmæssigt og ved aktiv deltagelse i standardiseringsarbejde, relevante netværk og konferencer både nationalt og internationalt.

Organisation

Vision:

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

Mål:

- Funktionsgruppen for Digital Bevaring arbejder aktivt på at udbrede kendskabet til gruppens arbejde, således at kolleger ved, hvordan og hvornår man kan og bør rådføre sig med gruppen.

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 af 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 kulturarvsmateriale er der oprettet funktionsgrupper, som arbejder med hhv. indsamling, bevaring og adgang. På tværs af disse funktioner er udpeget materialeansvarlige, som hver især er videnspersoner for en eller flere materialetyper og tager sig af den generelle datahåndtering og rådgivning på tværs af hele materialets livscyklus.

Funktionsgruppen for Digital Bevaring er sammensat af ansatte ved biblioteket med kompetencer inden for forskellige aspekter af digital bevaring. Hvis en ansat på Det Kgl. Bibliotek har spørgsmål relateret til digital bevaring, kan de stiles til funktionsgruppen. Gruppens primære ansvar er at sikre udmøntningen af politik og strategi for digital bevaring, samt at beslutninger om digital bevaring træffes på et kvalificeret grundlag. Derudover skaber funktionsgruppen overblik over den digitale bevaring på Det Kgl. Bibliotek og sikrer, at der findes nedskrevne procedurer for aktiviteter relateret til digital bevaring.

Funktionsgruppen træffer således de faglige beslutninger vedrørende digital bevaring. I tilfælde hvor fx økonomiske eller ressourcemæssige konsekvenser har indflydelse på en foretrukket beslutning, eskaleres diskussionen og afgørelsen til funktionsgruppens styregruppe.

Risikostyring

Vision:

- Bevaring af bibliotekets samlinger af digital kulturarv sikres ved brug af risikostyring som et dagligt redskab for alle de medarbejdere, der har ansvar for digitale samlinger, digital bevaring og de it-systemer, som håndterer data og metadata.

Mål:

- Den fælles proces om etablering af en ISO/IEC-27001-compliant informationssikkerheds-proces for bibliotekets it-aktiver, skal, hvor det omhandler bevaringssystemer, understøtte langtidsbevaring af kulturarven

Risikostyring omfatter:

- Risikovurdering: Identificering, registrering, analyse og evaluering af de potentielle risici, der truer bibliotekets digitale aktiver. Vurderingen revideres løbende for at identificere ændringer i rammer, vilkår, nyopståede risici eller mitigeringsmuligheder, 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 af it-systemer, aktiver og processer sker i samarbejde med Det Kgl. Biblioteks sikringsudvalg, og disse risikovurderinger udarbejdes med reference til standarden om informationssikkerhed, ISO/IEC-27001.

Derudover foregår der også risikostyring for bibliotekets digitale repository som helhed, hvilket omfatter både bitbevaringsløsningen, ressource- og metadatahåndteringssystemer og de platforme, der giver direkte adgang til det bevarede materiale samt de organisatoriske og økonomiske rammer. Det samlede risikobillede for bibliotekets repository kræver derfor input fra og tværgående arbejde med 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. Denne samlede risikovurdering udarbejdes med reference til audit- og certificeringsstandard ISO 16363 og danner grundlag for et fremtidigt audit af biblioteket som Trustworthy Digital Repository.

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 2026
- 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 den 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 partnere) – at biblioteket lever op til kravene til et Trustworthy Digital Repository (Troværdigt Digitalt Arkiv), dog uden ekstern certificering. 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 auditrapport, som skal behandles og godkendes i styregruppen. Selv-audit udføres som en gennemsigtig og offentligt dokumenteret proces, der lever op til anerkendte internationale standarder og praksisser, herunder *ISO16363: Space data and information transfer systems -- Audit and certification of trustworthy digital repositories.*

Teknologiovervågning

Vision:

- Det Kgl. Bibliotek sikrer, at der løbende foretages teknologiovervågning

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

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

I forbindelse med bitbevaring overvåger biblioteket 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 biblioteket, både internt og i samarbejde med internationale partnere, etablere en løbende overvågning. Det vil ske 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.

Biblioteket foretager 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 foretager Det Kgl. Bibliotek også løbende overvågning inden for en lang række andre områder, blandt andet:

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

Generelt overvåger biblioteket i videst muligt omfang alle trusler – store som små – mod de digitale samlinger. Når en trussel er identificeret, udløser det en tilhørende risikohåndtering, ud fra hvilken det videre forløb besluttes.

Standarder

Vision:

- Det Kgl. Bibliotek benytter i videst 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

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 er vigtig for, at det bevarede digitale materiale vil kunne fortolkes 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 27001:2022: *Information technology -- Security techniques -- Information security management systems -- Requirements*
- ISO 690:2021: *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*
- URN Namespace Registration for Persistent Web IDentifiers (PWID) (https://www.iana.org/assignments/urn-formal/pwid)

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.

Åben og proprietær software

Vision:

- Det Kgl. Bibliotek fastholder kompetencer i at forstå, anvende og videreudvikle åben software (FOSS – Free and Open Source Software)
- Det Kgl. Bibliotek bidrager til fortsat udvikling af relevante FOSS-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 biblioteket selv eller andre. Derfor anvender biblioteket i videst muligt omfang åben software og software, der understøttes og anvendes af et bredt, internationalt fællesskab af institutioner, der er sammenlignelige med Det Kgl. Bibliotek. I det omfang, det er muligt, deltager biblioteket aktivt i videreudviklingen af open source-produkter, ligesom biblioteket ønsker at være en aktiv bidragyder i brugerfora og lign. for de kommercielle produkter, som anvendes.

På en række områder findes der ikke åben software, der ressource-, sikkerheds- eller funktionalitetsmæssigt kan måle sig med kommercielle softwareprodukter, og biblioteket anvender derfor også proprietær software. Hvor biblioteket bruger proprietær software i sit

repository, er det essentielt at bruge software, som gemmer data og metadata i åbne formater, sådan at lock-in kan undgås. Derudover foretrækker biblioteket software, som bruger veldefinerede og helst standardiserede programmerings-interfaces, som gør det muligt at integrere med andre produkter og løsninger. I overvejelser omkring udvikling, anskaffelse og implementering af nye softwareløsninger indgår både åbenhed, økonomi, sikkerhed og evt. udbredelse af internationale fællesskaber. Softwareløsninger underkastes risikovurderinger, både ved anskaffelse og løbende derefter.

Indlemmelse

Vision:

- Det Kgl. Bibliotek sikrer, at alle 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 2024 udarbejdes en vejledning til, hvilke metadata der skal opsamles i forbindelse med digitaliseringsprocesser.

Det Kgl. Bibliotek tilstræber at modtage samlinger i så korrekt form som muligt. Som en del af sikringen af dataindsamling 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 og sådanne undtagelser dokumenteres i samlingens bevaringsplan. 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 og -processer

Vision:

- Det Kgl. Bibliotek sikrer, at tilgængeliggjort materiale er under aktiv bevaring

Mål:

- I løbet af 2024 er der udarbejdet eller revideret bevaringsplaner for alle Det Kgl. Biblioteks digitale kulturarvsdata
- I løbet af 2024 udarbejder Funktionsgruppen for digital bevaring en vejledning for valg af bitbevaringsplan
- Funktionsgruppen for digital bevaring gennemgår bevaringsplanerne for alle Det Kgl. Biblioteks digitale kulturarvsdata hvert andet år

I arbejdet med digital bevaring følger Det Kgl. Bibliotek en række grundlæggende principper. Bevaringsprincipperne anvendes på alle niveauer af bevaringen, fx skal biblioteket leve op til anerkendte standarder inden for software-kvalitet, sikkerhed og procedurer for arbejdet. Biblioteket bruger risikostyring til at sikre de digitale kulturarvsdata bedst muligt, blandt andet ved en årlig gennemgang af systemer og arbejdsgange.

For hver samling af data, som kan bevares på samme måde, afklares

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

Desuden afklares formater for metadata ud fra vedtagne procedurer og guidelines med begrundelse for eventuelle afvigelser. Oplysningerne skal være så detaljerede, som det er muligt, og specificeres i Samlingsdatabasen, som er bibliotekets oversigt over digitale kulturarvssamlinger.

Biblioteket har et pre-ingest-område til opbevaring af materiale, der fx endnu ikke er kurateret, eller som mangler workflows til endelig bevaring. Materialet på pre-ingest-området er således ikke i de endelige systemer til langtidsbevaring. Hvor der i det følgende ikke specifikt står "pre-ingest bitbevaring" betyder "bitbevaring" altid langtidsbevaringen.

Bevaringsplaner

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. Bevaringsplanerne udarbejdes i tæt samarbejde mellem de

ansvarlige for digital bevaring og de samlingsansvarlige og dokumenteres i bibliotekets Samlingsdatabase.

På Det Kgl. Bibliotek oprettes for alle digitale materialer bevaringsplaner for at sikre dokumentation og overblik over den digitale bevaring. Der er tre typer af bevaringsplaner:

- Bitbevaringsplaner
  Alle materialer har en bitbevaringsplan. Materialer, som har samme bitsikkerheds- og fortrolighedsniveau samt karakteristika i forhold til, hvordan de skal bitbevares, er omfattet af samme bitbevaringsplan.

- Bevaringsplaner for emulering
  Disse planer dækker materialer, som har samme plan for emulering. Der findes forskellige planer i forhold til værktøjer eller teknikker til emulering.

- Bevaringsplaner for migrering
  Disse planer dækker materialer, som har samme plan for migrering. Der findes forskellige planer i forhold til hvilket format, der migreres fra og til, hvilke værktøjer, der anvendes, og hvilke signifikante egenskaber, der skal bevares i migreringen.

Målet er, at alle data skal have relevante bevaringsplaner, og der skal laves ensartede procedurer for, hvordan der kan træffes et valg om bevaringsplan for nye data.

Bevaringsplaner for bitbevaring

En bitbevaringsplan er karakteriseret ved det bitbevaringsniveau og fortrolighedsniveau, den opfylder. Hvilken bevaringsplan, der skal tildeles bestemte data, er derfor fastlagt ud fra krav til:

- Bitsikkerhed
  Angives i bitsikkerhedsniveauer, som udtrykker de bitbevaringsprincipper, der anvendes 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
  id.kb.dk/vocabulary/preservationLevels.html

- Fortrolighed
  Angives i fortrolighedsniveauer, alt efter omfang af adgangsrestriktioner for de bitbevarede data. Fortrolighed er vigtig at have med i bitbevaringsprincipperne, 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
id.kb.dk/vocabulary/preservationLevels.html

Disse krav har indflydelse på, hvor mange kopier af data, der bevares, samt sammensætningen af såkaldte pillars, som kopierne fordeles på, og hvor godt de enkelte pillars sikres. En pillar er defineret ved den organisation og teknik, der er anvendt for den enkelte kopi i en samlet bevaringsløsning.

Det Kgl. Bibliotek har følgende bitbevaringsplaner. Planerne er navngivet efter bitbevaringsniveau og fortrolighedsniveau:

H mH	Høj bitsikkerhed – meget høj fortrolighed
H H	Høj bitsikkerhed - høj fortrolighed
H L	Høj bitsikkerhed - lav fortrolighed
M L	Middel bitsikkerhed - lav fortrolighed

Bevaringsplaner for emulering

For komplekse dataformater vil man med fordel kunne anvende emulering som logisk bevaringsstrategi. Det samme er tilfældet, når materialerne er en del af en større sammenhæng som fx et disk image, hvor det så er muligt at emulere helheden inklusiv den originale software, som har været anvendt til at tilgængeliggøre materialerne. I forbindelse med emulering er det vigtigt at sikre, at de nødvendige emuleringsmiljøer er til stede. Emuleringsmiljøerne skal indeholde de nødvendige softwaresystemer i de nødvendige versioner mv., ligesom der skal være dækkende beskrivelser af miljøerne.

Bevaringsplaner for migrering

Hvis det involverede dataformat er simpelt, vil man i de fleste tilfælde med fordel kunne anvende migrering som logisk bevaringsstrategi. For at finde den optimale migrering for et materiale fastlægges materialernes dataformat samt de signifikante egenskaber. Med udgangspunkt i dette undersøges, hvilken migreringsvej der er den optimale, altså hvilket dataformat materialet skal have, efter migreringen er afsluttet. De mulige migreringsveje afdækkes ved at undersøge de teknologiske muligheder i de værktøjer, som tilbyder migreringer mellem dataformaterne.

Bitbevaring

Vision:

- Det Kgl. Bibliotek tilstræber bitbevaring af digitale materialer, så materialerne på langt sigt (100 år) kan genfindes og læses i deres oprindelige form
- Det Kgl. Bibliotek tilstræber at bitbevare samlinger på så ensartet og optimal måde som muligt med hensyntagen til krav om fortrolighed, tilgængelighed og bitsikkerhed
- Det Kgl. Bibliotek tilstræber at have så teknologiuafhængige bitbevaringsløsninger som muligt

Mål:

- I 2024 er der fastlagt procedurer for interne såvel som eksterne direkte adgange til samlinger under bevaring med henblik på at sikre, at både fortroligheds- og sikkerhedsniveauer på materialerne overholdes
- I 2024 laves en bitbevaringspillar på Amager, og flytning af kopier hertil vil forbedre bitbevaringen for størstedelen af materialerne
- Senest i 2026 er der lavet en udredning om muligheder for at opnå bedre uafhængighed i forhold til geografi, organisation, lagringsmedier og politisk/juridisk uafhængighed
- Senest i 2026 er der lavet procedurer og tilretninger, der gør det muligt at bitbevare meget fortrolige data i krypteret form
- Senest i 2027 er en bitbevaringsløsning for pre-ingest-materiale implementeret
- Senest i 2024 er der lavet procedurer for drift og vedligeholdelse af bitmagasinerne
- Senest i 2024 er der lavet risikoanalyse for samlinger med to digitale kopier og en fysisk kopi

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

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

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

- bitsikkerhed
- fortrolighed
- tilgængelighed

Implementering af flere bitbevaringsløsninger sker på basis af en risikovurdering og ud fra hensyntagen til dels de ovennævnte krav, dels de helt overordnede rammer, der er for bevaring på biblioteket i forhold til fx økonomi og teknologi. Når alle disse differentierede implementeringer er gennemført, vil de omfatte løsninger for

- forskellige fortrolighedsniveauer, fx ikke-fortrolige data eller meget fortrolige data,
- forskellige bitsikkerhedsniveauer, hvor der er behov for forskellige antal kopier og lokationer.

Alle kopier af data vil være under aktiv bitbevaring, hvilket vil sige, at der foretages løbende integritetscheck og fejlretning.

Bånd migreres løbende hvert 7. år til nye bånd. Diske migreres i forhold til leverandør-anbefalinger og ved fejl ud fra et forsigtighedsprincip. I forbindelse med migrering af bånd laves der altid integritetstjek. For diske laves løbende integritetstjek, minimum hver 6. måned.

Fastlæggelse af forholdet mellem disse parametre sker ud fra en kombineret risikovurdering og omkostningsanalyse. De risici, som skal tages i betragtning, omfatter som minimum organisatoriske, lovgivningsmæssige, teknologiske og geografiske forhold. Forefindes fysiske kopier, som kan gendigitaliseres, indgår disse i risikovurderingen for de digitale kopier af samme materiale.

Det Kgl. Bibliotek sikrer via software og principper på bitrepository.org, at bitbevaringen kan håndteres så teknologi- og leverandør-uafhængigt som muligt.

Pre-ingest

Materiale, som endnu ikke er klar til at blive lagt i bevaring, lægges på et område kaldet 'pre-ingest'. For at sikre dette materiale bedst muligt etableres en separat installation af bitmagasinet, hvor pre-ingest-materialet kan ingestes. I modsætning til det etablerede bitmagasin vil der efterfølgende kunne slettes fra denne installation, når materialet er ingestet i pre-ingest-bitmagasinet. I denne nye pre-ingest-installation vil alt materiale ligge på samme bitbevarings- og fortrolighedsniveau, og der vil ikke være de samme retningslinjer som for bitmagasinet.

Overordnet om implementering af bitbevaring

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 fra dette princip skal kunne begrundes ud fra økonomiske eller tekniske betragtninger.

Specifikationerne for bitbevaring vedligeholdes i henhold til løbende risikovurderinger. Der er fast etablerede procedurer for reetablering og logning i overensstemmelse med Det Kgl. Biblioteks sikringspolitik.

Biblioteket tilstræber at samarbejde med andre organisationer for at opnå de bedste og økonomisk mest fordelagtige betingelser for bitbevaring.

Den løbende teknologiovervågning og risikostyring skal bidrage til at sikre udførelse af relevante justeringer af implementeringer af bitbevaring. Dette gøres for at opnå og fastholde, at biblioteket benytter de mest optimale implementeringer af bitbevaring i henhold til rammerne udstukket af politikken. Dette indebærer også løbende evaluering af, hvilke nye afhængigheder der kan føre til datatab, fx ny teknologi, der anvendes.

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

Kendte eksempler på risici er beskrevet her, sammen med tiltag for mitigering af disse:

- **Komprimering**, som kan resultere i, at data mistes, fx hvis de ikke kan dekomprimeres.
  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. Denne regel gælder kun komprimering efter modtagelse.
- **Kryptering**, som kan resultere i, at data mistes, fx hvis nøgle til dekryptering mistes.
  Som udgangspunkt benytter biblioteket ikke kryptering for digitalt bevarede materialer, med mindre der er særlige krav om fortrolighed.
- **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.
- **Manglende kobling mellem identifiers og bitsekvenser**, som kan give tab, hvis bit-sekvenser ikke kan genfindes ud fra deres tildelte identifier.
  Identificering sikres ved enten at pakke identifier med bitsekvenser eller ved at bevare checksum for filer sammen med deres identifier, så checksummen kan bruges til at genberegne kobling i tilfælde af tab. Disse koblinger bitbevares også.
- **Pakkeformater**, som kan give tab, hvis de fraviger generelle bevaringsprincipper.
  I det omfang biblioteket benytter pakkeformater, bruges WARC, som opfylder krav til systemuafhængighed og til bevaringsformater.
- **Direkte fysisk og virtuel adgang** til kopier under bitbevaring, som kan kompromittere både fortrolighed og bitsikkerhed
  Adgang til data i bitmagasinerne sker, ud over for bitmagasinsoperationer, kun via etableret read-only adgang og via API-kald. Det tilstræbes at logge enhver intern og

ekstern adgang til kopierne, så et audittrail for bevaringen kan følges, og for at sikre at fx personfølsomme data ikke tilgås af uvedkommende.

Strategien er at pakke alle metadata i WARC for at sikre identificering og ensartet pakning.

Disse tiltag gælder ikke for pre-ingest, eftersom data på pre-ingest er i den form, som de bliver leveret i, og der skal kunne slettes fra pre-ingest.

Indhold af bitbevaringsplaner

På nuværende tidspunkt er bitbevaring implementeret på mange måder for de forskellige samlinger af data, her kaldt collections. Dette skyldes primært, at løsningerne ikke er harmoniseret efter fusionen af det tidligere Statsbiblioteket og det tidligere Kongelige Bibliotek. En anden faktor er, at der er forskellige krav til, hvordan de forskellige collections bitbevares. Disse krav udgøres primært af tilgængelighed, økonomi og tekniske udfordringer.

Herunder vises en oversigt over de eksisterende collections med indhold. Collections er grupperet i samlingsgrupper, som repræsenterer en bestemt måde at bevare collections på. De enkelte samlingsgrupper er i det nedenstående repræsenteret med en farvet kasse. De 12 collections er angivet ved det navn, de har i bibliotekets dokumentation, Samlingsdatabasen.

For hver af samlingsgrupperne er beskrevet, hvilken bevaringsplan de tilhører:

Tre båndkopier (plan: H L Høj bitsikkerhed - lav fortrolighed)

- Radio/tv-samlingen

Webarkivet (plan: H H Høj bitsikkerhed - høj fortrolighed)

- Netarkivet

Tre kopier Kbh (plan: H L Høj bitsikkerhed - lav fortrolighed)

- F: Subst.-digi. samlingsmateriale, ikke fortroligt
- H: Dig. født dokumentationsmateriale
- I: KB institutionsarkiv, digitalt født
- kb-ikke-fortroligt-materiale

Tre kopier Aarhus (plan: H L Høj bitsikkerhed - lav fortrolighed)

- Digital pligtafleverede aviser
- Digital pligtafleverede e-bøger
- Digital pligtafleverede e-musik
- Det administrative bibliotek

Med fysisk kopi Kbh (plan: M L Middel bitsikkerhed - lav fortrolighed)

- K: Digitaliseringsproj. med originaler

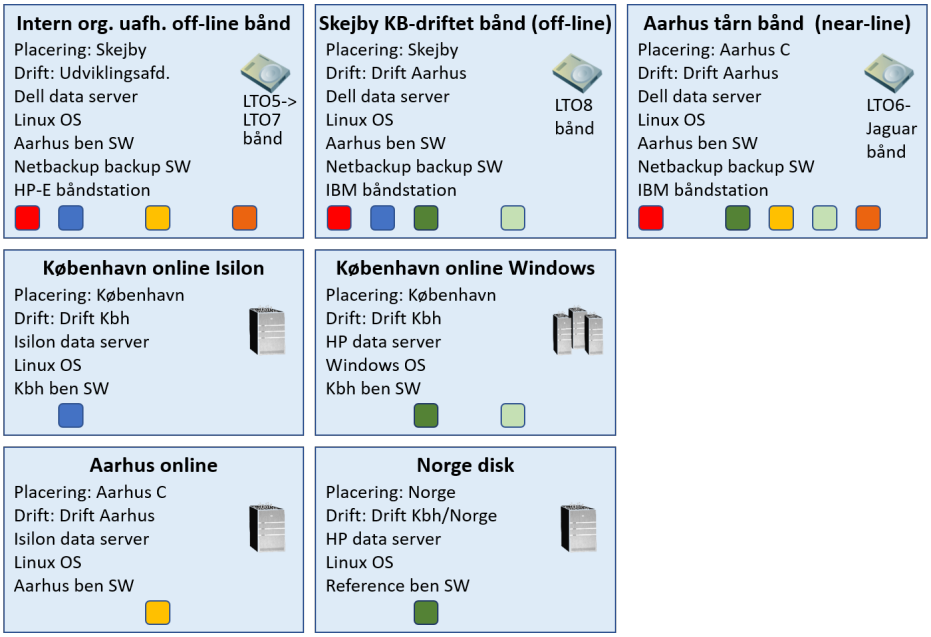
Med fysisk kopi Aarhus (plan: M L Middel bitsikkerhed - lav fortrolighed)

- Avis samlingen (Ninestars)

En del samlinger er endnu ikke under bitbevaring og ligger på føromtalte pre-ingest, blandt andet de meget fortrolige samlinger.

Nuværende implementering af bitbevaring for samlingsgrupperne

De enkelte samlingsgruppers kopier er fordelt på såkaldte pillars. En pillar er teknikken og organisationen omkring en kopi i en samlet implementering af bitbevaring. De enkelte pillars er illustreret i nedenstående tegning:



Figur 1 Nuværende implementering med fordeling af samlingsgrupper på pillars

Figuren viser kun pillars med fulde kopier, dvs. pillars, der alene rummer checksummer, er ikke medtaget her.

Desuden er det forskellige organisatoriske enheder, der tager sig af driften, selvom alle pillars er driftet internt.

Endeligt skal det bemærkes, at denne implementering kun ser på uafhængighed mellem pillars. Der er også andre aspekter, der skal tages højde for, såsom:

- Antal af installationer af bitmagasinløsningen
  De fire installationer er
  - samlingsgruppen *Tre båndkopier*,
  - samlingsgruppen *Webarkivet*,
  - samlingsgrupperne *Tre kopier Kbh* og *Med Fysisk kopi Kbh* og
  - samlingsgrupperne *Tre kopier Aarhus* og *Med fysisk kopi Aarhus*.
- Antal collections

- Bitbevaring af datafiler som selvstændige filer eller pakket med andre data i Warc-filer

Risici ved nuværende bitbevarings-samlingsgrupper
Af historiske grunde er der en del ikke-mitigerede risici forbundet med den nuværende implementering af bitbevarings-samlingsgrupperne. Disse er:

- Afstand: To ud af tre kopier er placeret *i samme bygning* eller kopierne er placeret med *lille afstand*.
- Software: Alle kopier er på pillars, der bruger *samme operativsystem* eller *samme pillar-software*.
- Hardware: Alle kopier er på pillars, der bruger *samme type server* eller *samme medietype.*
- Organisation: Alle kopier er på pillars, der driftes i *samme organisation*.
- Politisk: Alle kopier er på pillars, der er placeret i *samme land*.

I nedenstående tabel vises hvilke samlingsgrupper, der har hvilke risici:

Risici for nuværende samlingsgrupperne						
To af tre af kopier i samme bygning	■	■		■		
Lille afstand	■			■		■
Samme operativsystem	■	■		■		■
Samme pillar-software	■			■		■
Samme backup-software	■					■
Samme hardware	■					■
Samme medietype	■					■
Samme organisation	■	■	■	■	■	■
Samme land	■	■		■	■	■

I det følgende vises, hvordan strategien vil mindske disse risici.

Ny implementering af samlingsgrupper og collections

For at reducere antallet af samlingsgrupper og collections uden at gå på kompromis med kravene til deres bitbevaring arbejdes der på at implementere forenklinger. En forudgående analyse inddrog betragtninger om ressourcebehov samt vedligeholdelsestilstanden af de eksisterende ben. Dette vil i praksis betyde:

- Færre platforme/pillars, der skal driftes, vedligeholdes og administreres
- Nemmere vedligeholdelse og overvågning af indhold fra forretningens side
- Færre måder at tilgå data
- Mitigering af hovedparten af risici for tab af data
- Færre collections
- Færre installationer

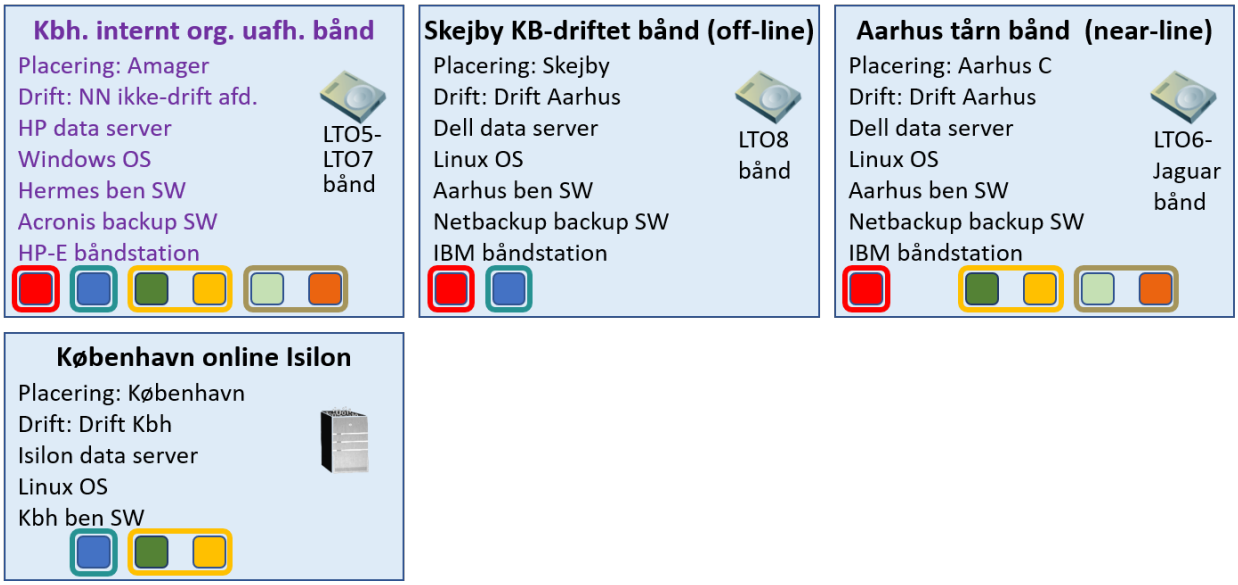
Med implementeringen vil samlingsgrupperne kunne reduceres til minimum de følgende:

Nye samlingsgrupper		Bevaringsplan
☐	**Høj bitsikkerhed - høj fortrolighed - med én online kopi** for tilfælde hvor adgangsløsninger bruger de bevarede data i formidlingen	H H
☐	**Høj bitsikkerhed - lav fortrolighed – med én online kopi** for tilfælde hvor mulighed for hurtig adgang kan skabes til de bevarede data	H L
☐	**Høj bitsikkerhed - lav fortrolighed – billigt lager** for tilfælde hvor omkostninger til lagringsmedier holdes nede pga. store datamængder	H L
☐	**Middel bitsikkerhed - lav fortrolighed – billigt lager** for tilfælde hvor omkostninger til lagringsmedier holdes nede pga. store datamængder, og hvor der findes en fysisk original	M L

Der bliver muligvis flere collections end samlingsgrupper på grund af de føromtalte hensyn til tekniske og kuratoriske udfordringer.

Planen er at reducere antallet af pillars fra syv til fire ved, at den tidligere 'Intern org. uafh. off-line bånd'-pillar migreres over i en ny 'Kbh. internt org. uafh. bånd'-pillar, som har nyt dataserver-fabrikat, ny placering, nyt operativsystem, ny pillar-software og ny backup-software.

Efter gennemførelse af forenklingerne vil billedet af pillars med nye collections se ud som illustreret i figur 2. Figuren viser også, hvordan de nye samlingsgrupper er sammensat af de gamle samlingsgrupper ved, at symboler for samlingsgrupperne er sat ind i symbolerne for de nye collections.



Figur 2 Ny implementering af pillars med angivelse af nye collections og tidligere samlingsgrupper

Bemærk, at som i figur 1 viser figur 2 kun pillars med fulde kopier, dvs. checksumspillars er ikke medtaget her.

Med den nye implementering vil risiko-billedet være reduceret til følgende:

Risici for samlingsgrupperne				
Samme medietype				
Samme organisation				
Samme land				

Der vil på længere sigt også blive set på mitigering af disse tre risici. De fortsatte risici med ”Samme organisation” og ”Samme land” vil relativt nemt kunne mitigeres ved at overføre den nye "Kbh. internt org. uafh. bånd" til en organisation i udlandet. Risikoen med ”Samme medietype” kræver en større ændring, fx etablering af en pillar baseret på optiske bånd.

Der vil også blive set på muligheden for at indføre en collection for meget fortrolige data. I denne collection vil alle kopier være krypterede, eftersom risiko for at bryde fortrolighed vejer tungere end risiko for at miste data. Da bitmagasinet ikke forholder sig til, om bits er krypterede eller ej, skal der ikke laves om i bitmagasin-softwaren. Derimod vil der skulle laves applikationer, som ikke bare ingester data, men også krypterer data inden ingest, og applikationer, hvorfra der gives adgang, og derfor skal kunne dekryptere data, inden de vises.

Vedligeholdelse og oprettelse af collections

Funktionsgruppen for Digital Bevaring har ansvaret for vedligeholdelsen af collections. Dette sker blandt andet ved at have procedurer, der indeholder en tjekliste af potentielle risici, beskriver hvordan integritetstjek foretages osv.

Logisk bevaring

Mål:

- I 2024 er der gennemført valg og etablering af prototyper for et eller flere emuleringsframeworks
- I 2025 er der etableret produktionsmiljøer for et eller flere emuleringsframeworks
- I 2026 er analyse og test af mulighederne for migrering 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 gengivelse af deres originale indhold og funktionalitet. Logiske bevaringsaktiviteter, i denne brede forstand, starter allerede ved indlemmelse – med indsamling af tekniske metadata, data-karakterisering og -validering.

Strategien for logisk bevaring varierer fra samling til samling, og det er derfor et vigtigt emne i udarbejdelsen af bevaringsplanen for samlingen at identificere de væsentligste aktuelle risici og den eller de mest relevante risikohåndteringsstrategier. På Det Kgl. Bibliotek anvendes to risiko-håndteringsstrategier for logisk bevaring, nemlig migrering og emulering.

Under udarbejdelsen af bevaringsplan og tilhørende risikovurdering beslutter biblioteket en foretrukken logisk bevaringsstrategi (emulering eller migrering) for en samling. Samtidig vurderes, om den eller de nuværende aktuelle risici er alvorlige nok til, at der skal sættes handlinger i gang for at bringe den foretrukne strategi i anvendelse.

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 den 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 i simple dataformater. Hvis resultatet af en risikovurdering 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 webarkiver. I webarkiver er det stadig normalt at bruge nutidens webbrowsere til at vise historisk indhold, men der er 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 muligheden for anvendelse 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.

Datamodel, metadata og dataformater

Vision

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

Mål:

- I 2024 er der lavet en opdateret vejledning til valg af dataformat til brug i de processer, hvor biblioteket selv kan bestemme formatet.
- I 2025 er der etableret en bevaringsløsning for metadata i system til understøttelse af bevaring, KUANA, så metadata herfra løbende bitbevares i bitmagasinet.

For at data og metadata til enhver tid kan forstås og fortolkes, 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. Når det er muligt, vælges formater, som er åbne, standardiserede og internationalt anerkendte, da dette sikrer, at data kan fortolkes på langt sigt.

En delvis standardisering af data- og metadataformater under bevaring er ønskelig for at gøre bevaringsprocessen så transparent, vedligeholdelsesvenlig og omkostningseffektiv som mulig. På grund af bredden af indhold, som ønskes bevaret, vil det aldrig være muligt at lave en fuldstændig standardisering. Fx 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.

Det Kgl. Bibliotek bevarer mange forskellige dataformater; dels fordi vi modtager donationer i forskellige formater, dels fordi vi selv indsamler materiale i forskellige formater. Vi søger at benytte så få og så standardiserede formater som muligt i de processer, hvor biblioteket selv har indflydelse på valg af format, fx ved digitalisering. Biblioteket bevarer som udgangspunkt originalformatet, også selvom der senere foretages en formatmigrering.

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 understøtter modellen, 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.

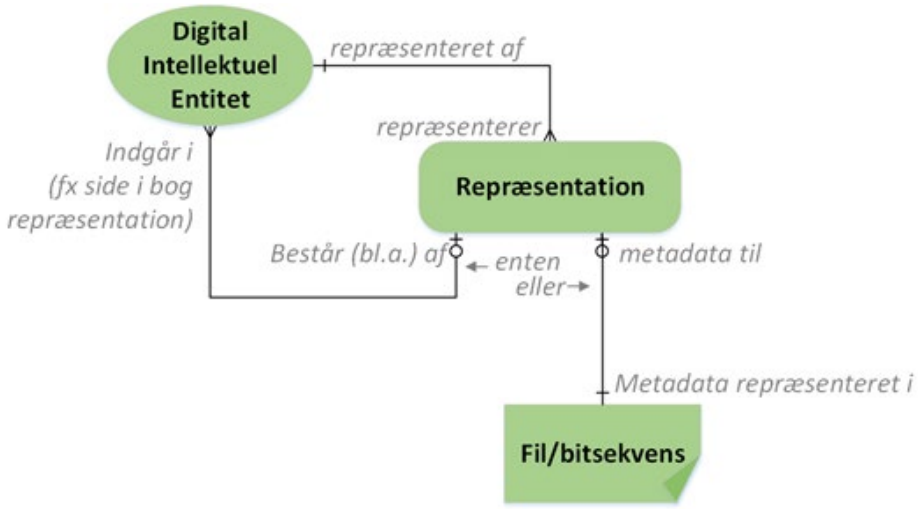
Grundlæggende datamodel

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

- **Digital Intellektuel Entitet**
  Denne udtrykker det øverste niveau for et digitalt objekt og identificerer entydigt et digitalt materiale. Objektet skal kunne identificeres uanset hvilke bevaringsaktioner, tilrettelser eller transformationer, der er lavet.
  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 særlige 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 baggrund 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 kan referere til fragmenter, som identificeres via fragmenternes Digitale Intellektuelle Entiteter og indeholder metadata om fragmenternes indbyrdes relation for at udgøre Repræsentationen (fx rækkefølgen af sider i en bog, hvor fragmenterne er bogens sider)

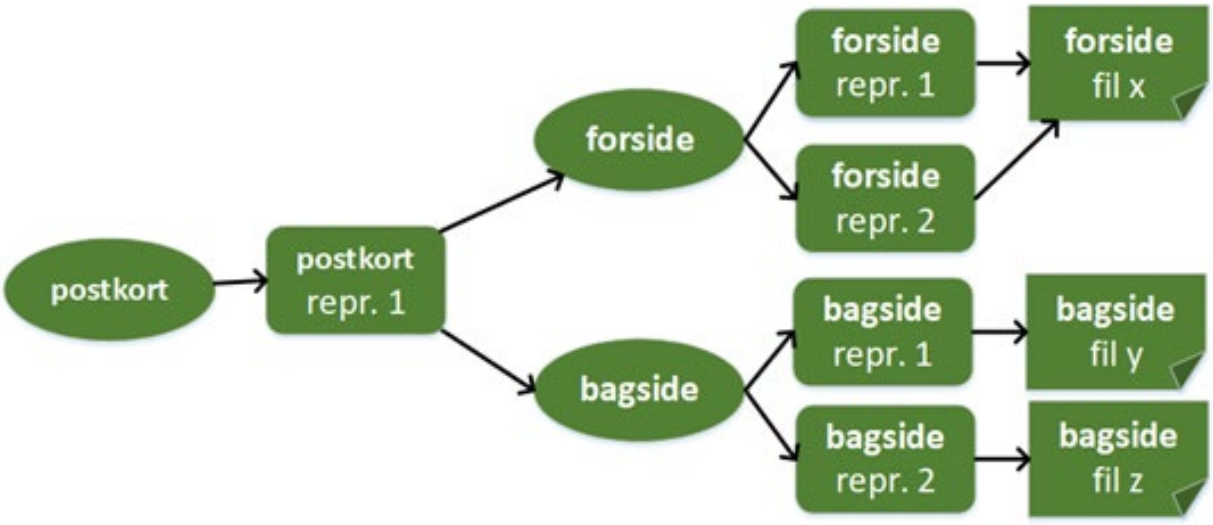
- **Fil/Bitsekvens**
  Denne er blot en fil/bitstrøm, som bevares i et bitmagasin (dataobjektet).



Beslutningen om, hvor mange niveauer der benyttes i en modellering, fastlægges ved ingest til bevaring

Eksempel på modellering af et digitaliseret postkort

Et eksempel på modellering 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 digitalisering. 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 gendigitalisering). 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 på id.kb.dk.

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.

Alle ændringer af digitale objekter og/eller metadata logges i et audittrail. Disse audittrails er at betragte som metadata og skal derfor også bitbevares.

Overordnet set er Det Kgl. Biblioteks strategi at bevare følgende typer metadata:

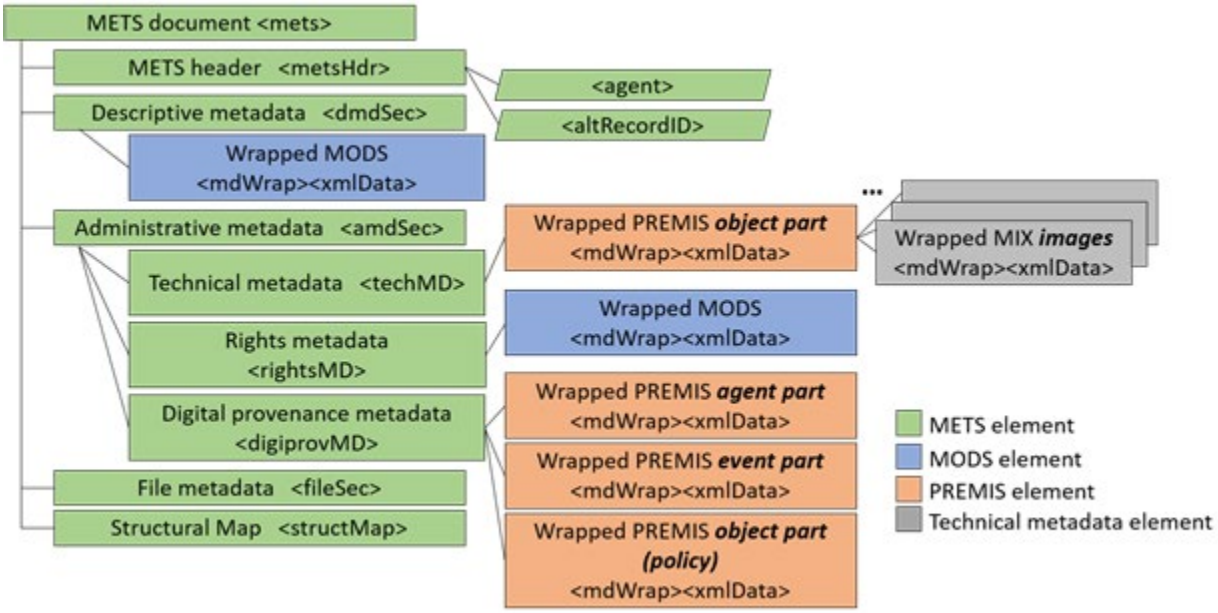
- *Beskrivende metadata*
  til beskrivelse af indholdet
- *Administrative metadata*
  til forvaltning af det digitale objekt, hvilket omfatter:
  - *Tekniske metadata*, fx filformat og informationer fra digitalisering
  - *Bevaringsmetadata*, der er nødvendige for bevaringsaktioner, fx bevaringsniveau og bevaringsstrategi
  - *Digital proveniens*, som indeholder audittrails
  - *Rettighedsmetadata*, fx lovgrundlag for indsamlingen eller ophavsret
- *Strukturelle metadata*
  strukturer der måtte være for det digitale objekt (fx sider i en bog)

Det er et mål, at man på tværs af biblioteket anvender de samme metadatastandarder på en så ensartet måde som muligt. Dette medvirker både til mere effektiv bevaring og gør det nemmere at give adgang til materialet. Funktionsgruppen arbejder derfor på at få etableret en tværgående metadatagruppe for hele biblioteket.

Bitbevaring af metadata

Bitbevaring af metadata sker gennem jævnlig eksport af metadata, inklusiv audittrails, fra repository-systemerne 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.

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



For at sikre, at bibliotekets metadata er brugbare på langt sigt, skal metadataprofil og de metadatastandarder, der anvendes i vores digitale samlinger, være tilgængelige. Dette gælder også, selvom den oprindelige placering for fx et metadataskema ikke længere er tilgængelig. Ansvaret herfor ligger hos Funktionsgruppen for Digital Bevaring, og profiler og standarder opdateres 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. Derfor skal alt materiale, der bevares, i videst muligt omfang gennemgå en karakterisering og validering så tidligt som muligt i materialets livscyklus. Output fra disse processer gemmes som tekniske metadata for materialet.

Derudover bør et objekts tilblivelse og historik beskrives i metadata. 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 metadata, da dette er meget afhængigt af leverandørens tekniske setup. Det skal derfor være et fast punkt i ethvert eksternt digitaliseringsprojekt at afklare, hvilke metadata og -formater en leverandør kan levere.

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

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, kan findes på id.kb.dk.

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 og/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 situationer hvor materialer ikke stammer fra filer, fx Netarkivet, hvor HttpWebResponses bliver pakket som warc-filer til arkivering.

Funktionsgruppen for Digital Bevaring opretholder et samlet overblik over, hvilke dataformater der produceres ved interne processer, og gruppen rådgiver i forhold til eksterne digitaliseringsprojekter samt ved overdragelse/levering af samlinger til biblioteket.

Teknisk infrastruktur

Vision:

- Det Kgl. Bibliotek tilstræber at ensarte bevaringsprocesser på en så optimal måde som muligt med hensyntagen til krav om fortrolighed, tilgængelighed og kompleksitet i de digitale materialer

Mål:

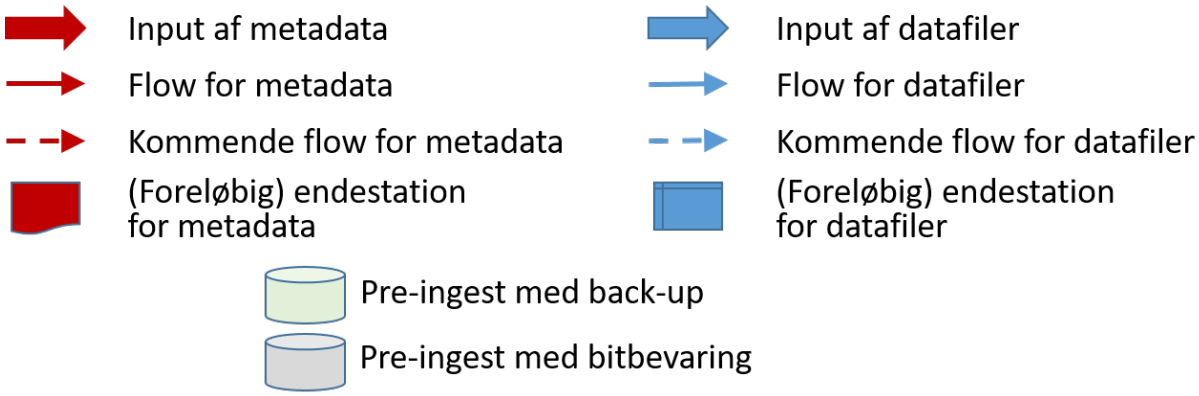
- I 2024 er der idriftsat bevaring af metadata fra KUANA
- I 2025 er der etableret bitbevaring af alle metadata fra Netarkivet
- I 2026 er der etableret et metadata-warehouse, som på sigt kan repræsentere og tilgå bitbevarede metadata for alle bevarede materialer

- I 2025 er alle instanser af SB-DOMS udfaset

Nedenfor er arkitekturen gennemgået i tre trin:

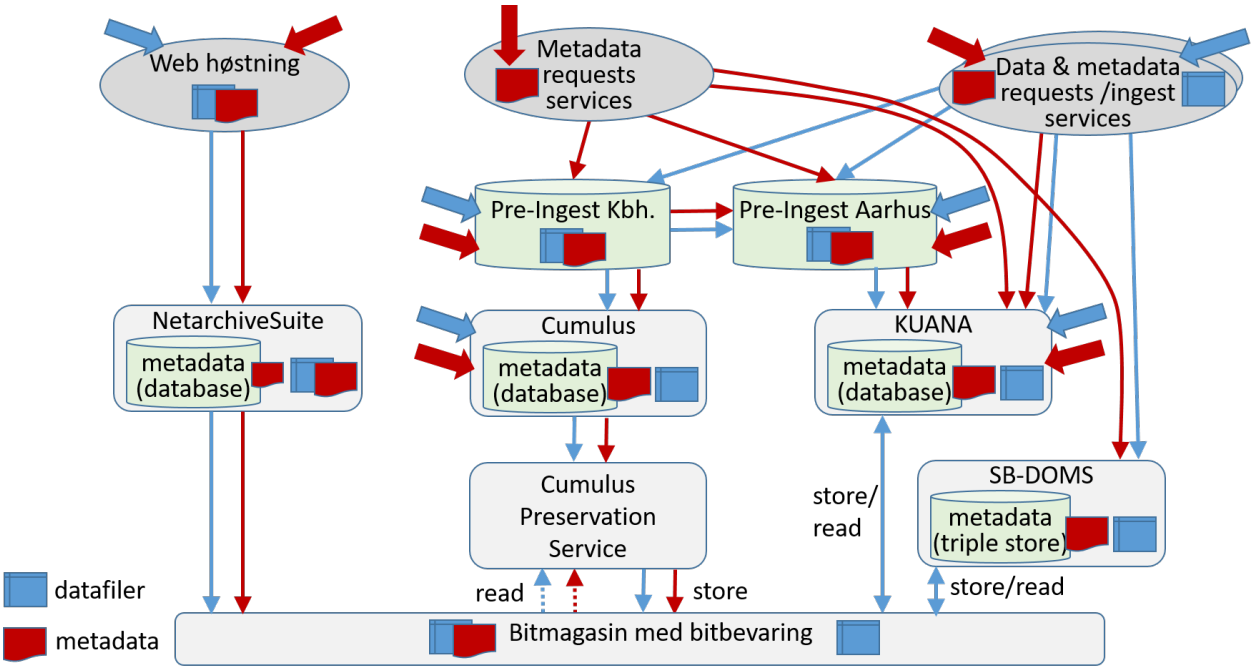
- den nuværende arkitektur
- den fremtidige arkitektur, som Det Kgl. Bibliotek vil stræbe efter at have på plads inden for en 3-5 årig periode
- den arkitektur, som det er visionen at nå frem til på sigt

De tre arkitekturtegninger vil benytte sig af følgende symboler:



Den nuværende arkitektur

Nedenstående figur illustrerer den nuværende arkitektur. Arkitekturen, der herunder omtales, omfatter udelukkende systemer til bevaring, hvilket vil sige, at formidlings- og adgangs-platforme ikke er medtaget. Af samme grund er materialer, som endnu ikke er kommet til pre-ingest, heller ikke taget med.



I den nuværende arkitektur er der forskellige metadata og data ingest-services. Eksempler på sådanne services er, at radio-/tv-data og -metadata modtages fra forskellige aggregatorer, ligesom der modtages både data og metadata via et ingest workflow-system. 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, der lægges direkte i pre-ingest.

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

- Cumulus laver strukturering af data og metadata og giver mulighed for editering af metadata.
  Via Cumulus BevaringsService bitbevares både metadata og filer med det bevaringsniveau, som er specificeret i metadata.
- Netarkivet er med NetarchiveSuite-softwaren integreret med ingest-systemerne til overførsel af data til bitbevaring. Derudover er der et system 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, men metadata bitbevares ikke.
  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, men metadata bitbevares ikke.
  SB-DOMS bygger på en forældet teknologi og udfases over de kommende år.

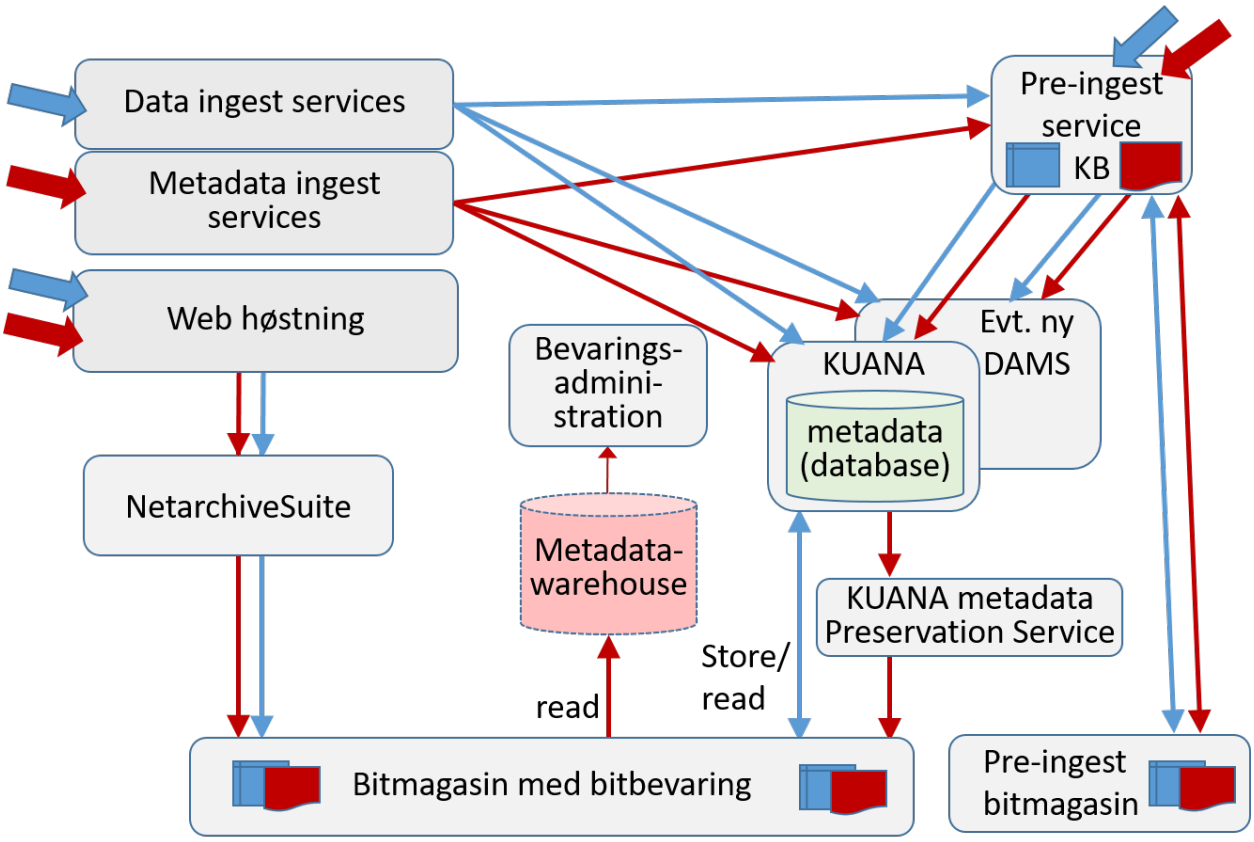
Bevaringsadministration, som blandt andet omfatter bevaringsplaner på samlingsniveau og filformatniveau, vedligeholdes i en separat Samlingsdatabase, som er uafhængig af resten af bevaringsapplikationerne.

Bitmagasinet er i illustrationen angivet som ét bitmagasin, men er i praksis flere forskellige bitmagasiner, alle baseret på bitrepository.org. Dette er beskrevet i afsnittet om bitbevaring.

Arkitektur inden for en 3-5-årig tidshorisont

Der arbejdes på 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 bitbevaret eller er i pre-ingest
- at antallet af repository-systemer, der understøtter bevaring, er reduceret ved, at SB-DOMS er nedlagt
- at der er indført et metadata-warehouse, som kan give overblik over samtlige metadata på ensartet og systemuafhængig vis.



Repository "Evt. ny DAMS" dækker over en eventuel efterfølger til Cumulus.

I denne arkitektur vil ingest af data og metadata til bevaringssystemer 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 være materialer, som skal kurateres før ingest, eller som der endnu ikke findes workflows til, hvorfor materialet i første omgang placeres på pre-ingest. Derfor viser illustrationen, at der også vil være input af data og metadata til pre-ingest. Målet er at minimere data i pre-ingest så meget som muligt.

For at sikre data på pre-ingest bedst muligt etableres et særskilt bitmagasin. Det vil betyde, at data, som endnu ikke er forberedt til at blive indlemmet i systemer til langtidsbevaring, alligevel vil være bitbevaret og sikret bedst muligt. Når data er klargjort og indlemmet i systemer til langtidsbevaring, vil de blive slettet fra pre-ingest-bitmagasinet. På dette punkt adskiller pre-ingest-bitmagasinet sig fra bitmagasinerne til langtidsbevaring, idet sletning ikke er tilladt i disse pga. risikoen for skade på data.

Antallet af repository-systemer, som understøtter bevaring, er indsnævret til 2-3 systemer, idet SB-DOMS udfases, og Cumulus enten udskiftes eller erstattes af KUANA. Netarkivet bibeholdes som særskilt repository-system, da det afviger væsentligt fra de andre systemer. Selvom det er muligt, at KUANA på sigt kan rumme alle informationer om bevarede data, ønsker biblioteket at etablere et metadata-warehouse, da det vil kunne skabe bedre overblik over bevaringen og gøre det muligt at tilgå alle bevarede metadata samlet, ekskl. Netarkivet pga. web-materialers struktur. Mere specifikt vil et metadata-warehouse kunne understøtte:

- fastholdelse af systemuafhængig understøttelse af statistik og overbliksfunktionalitet, som ikke vil kunne understøttes andre steder
- tilgang til bevarede metadata i normaliseret form
- bedre mulighed for at lave exit-strategier for proprietære systemer med bevaringsdata
- at fungere som en aggregator for alle bevarede metadata (ekskl. Netarkivet)
    - med et minimum af kompleksitet, selvom der er tilknyttet flere repository-systemer, der understøtter bevaring
    - med al historik gemt for metadata fra både tidligere og nuværende systemer.

Generelt er der forskellige, lokalt udviklede metadata-formater for eksisterende bevaringsløsninger. Som beskrevet i metadata-afsnittet, er der behov for at få så standardiserede og systemuafhængige metadata som muligt. Den standardiserede metadata-struktur 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 på sigt kunne blive suppleret med data fra metadata-warehouse, hvor det er relevant.

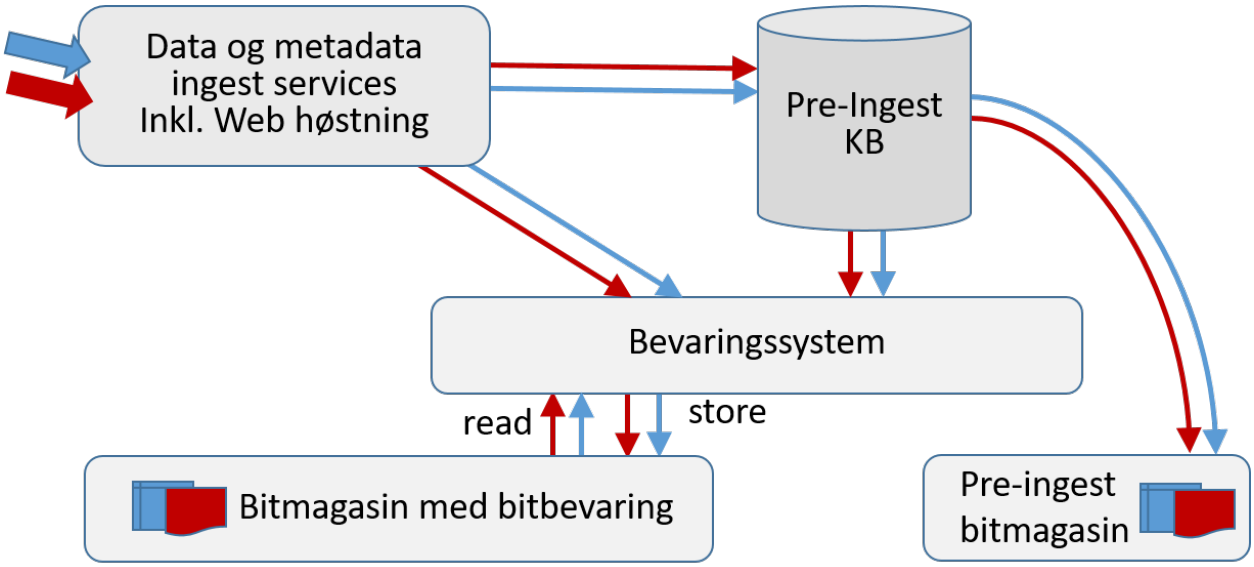
For at ensarte bitbevaring og interface til bitbevaring skal:

- data-filer fremadrettet ikke pakkes i WARC
- metadata, som ønskes bitbevaret, pakkes i WARC (bortset fra Netarkivet)
- identifiers, som bruges i bitbevaringsoperationer og i interface, enten være UUID'er eller identifiers konstrueret af fastfrossen metadata eller høstningsinformationer.

Bitmagasinet baseres på bitrepository.org frameworket uden afkoblede backups.

Den ønskede arkitektur (visionen)

Visionen for Det Kgl. Bibliotek er at have ét 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 fortsat et pre-ingest. Dette skyldes, at der realistisk set altid vil være behov for pre-ingest som forklaret i foregående afsnit. Netarkivet såvel som metadata-warehouse og bevaringsadministration er i ovenstående diagram en del af bevaringssystemet.

Forskning, videndeling og kompetence-udvikling

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:

- Biblioteket fastholder sin aktive tilstedeværelse i relevante faglige organisationer som fx OPF (Open Preservation Foundation) og IIPC (International Internet Preservation Consortium)
- Biblioteket deltager i relevante konferencer og workshops og med mindst ét bidrag på en anerkendt konference om digital bevaring hvert år
- Bibliotekets erfaringer og viden inden for digital bevaring formidles løbende til det faglige, internationale netværk via webinarer, artikler og lignende
- Bibliotekets erfaringer og viden inden for digital bevaring præsenteres løbende til et bredere, nationalt publikum via medierne og kb.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. Biblioteket overvejer løbende andre fora, som kan være relevante at være med i, som fx DPC (Digital Preservation Coalition). Desuden deltages i relevante konferencer, workshops og lignende, hvor medarbejderne dels kan være medarrangører, dels fremlægge egne erfaringer med digital bevaring, 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 mindst et indlæg hvert år.

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

Som en af de førende danske institutioner inden for digital bevaring tilstræber biblioteket også at formidle om digital bevaring på dansk, hvilket bl.a. sker gennem danske medier, kb.dk og networking.

Samarbejde om bevaringsaktiviteter

Vision:

- Det Kgl. Bibliotek samarbejder nationalt og internationalt om bevaringsaktiviteter.

Mål:

- Det Kgl. Bibliotek arbejder aktivt på et udvidet samarbejde med en udenlandsk partner om bitbevaring.
- Det Kgl. Bibliotek deltager aktivt i udviklingen af software via OPF og andre relevante communities.

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 bibliotekets internationale omdømme inden for digital bevaring. Sådanne samarbejder kan blandt andet omfatte driftsfællesskaber, softwareudvikling, software user communities og kompetenceudvikling og -udveksling.

Administration af dokumentet

Nærværende dokument er 2. version af Det Kgl. Biblioteks Strategi for digital bevaring, opdateret i foråret 2024.

Opdatering af dokumentet

Opdatering af dokumentet sikres af Funktionsgruppen for Digital Bevaring og godkendes af styregruppen for denne samt af Det Kgl. Biblioteks direktion. Dokumentet opdateres hvert femte 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 med det formål at understøtte nationalt og internationalt samarbejde om digital bevaring.

Referencer

Det Kgl. Biblioteks Politik for digital bevaring, 2023, https://www.kb.dk/politikker-og-strategier/det-kgl-biblioteks-politik-digital-bevaring

Det Kgl. Biblioteks sikringspolitik, 2019 - ikke publiceret

id.kb.dk, http://id.kb.dk/

ISO/IEC 27001:2022 *Information security, cybersecurity and privacy protection*, https://www.iso.org/standard/27001

ISO 16363:2012 (2012): *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories* http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510