Dataopbevaring ved AU

Vejledning til systemejere om valg af dataopbevaring

**Dokument:**

| **Version** | **Dat0** | **Ændret af** | **Ændring (hvad er ændret)** | **Godkendt af** |
| --- | --- | --- | --- | --- |
| 1.0 | 23.11.2016 |  |  |  |

**Indholdsfortegnelse:**

[1 Indledning 2](#_Toc470699324)

[2 Læsevejledning 3](#_Toc470699325)

[3 Vejen frem til implementering af dataopbevaringen 3](#_Toc470699326)

[4 Er datasamlingen bevaringsværdig? 5](#_Toc470699327)

[5 Afleveringspligt til Rigsarkiv og identifikation af datasæt 5](#_Toc470699328)

[6 Forudsætninger som datasættet skal opfylde 7](#_Toc470699329)

[7 Valg af dataopbevaringsløsning 8](#_Toc470699330)

[8 Implementering af dataopbevaring 10](#_Toc470699331)

[9 Definitioner 11](#_Toc470699332)

[9.1 Definition af statiske data 11](#_Toc470699333)

[9.2 Definition af ikke-aktive data 11](#_Toc470699334)

[9.3 Definition af AU-arkivpligt 11](#_Toc470699335)

[9.4 Definition af opbevaringsperiode 11](#_Toc470699336)

[9.5 Definition af udløbstidspunkt 11](#_Toc470699337)

[9.6 Definition af datasæt 11](#_Toc470699338)

[9.7 Definition af arkiveringsversion 11](#_Toc470699339)

[9.8 Definition af bevaringsværdige data 12](#_Toc470699340)

[10 Referencer 13](#_Toc470699341)

# 

# Indledning

Formålet med denne vejledning, som er rettet mod systemejere/datasamlingsejere[[1]](#footnote-1), er at være en støtte i situationer, hvor der er behov for og/eller tanker om at opbevare nogle data under andre betingelser end hidtil. Det kan være fordi de ikke længere skal bruges rutinemæssigt/hyppigt, men alligevel ikke blot må slettes og glemmes. En situation, der typisk opstår ved udfasning af et ældre it-system, men det kan også være i forbindelse med andre behov for at arkivere data, f.eks. ved ændringer i systemer forårsaget af organisatoriske omlægninger.

Betegnelsen ”dataopbevaring” udtrykker at det ikke drejer sig om ”arkivering” i biblioteksmæssig forstand, og at de løsninger der anvendes ikke udelukkende er egentlige arkiveringssystemer, idet nogle af dem er databaseløsninger. Se evt. mere om dette i reference [1], hvor baggrund, definitioner, rationaler m.m. omkring dataopbevaring ved AU er detaljeret forklaret.

# Læsevejledning

Det er hensigten med vejledningen at den kan bruges til at få en overordnet forståelse af ”dataopbevaring ved AU”, og at den derfor vil være en god forberedelse, inden der tages kontakt til AU IT om den konkrete opgave. F.eks. sådan at læseren allerede på forhånd har en fornemmelse af hvilken løsning der vil være relevant. Ved spørgsmål må man selvfølgelig altid kontakte AU IT, og kontaktinformationer findes via hjemmesiden [2].

Det er tilstræbt at stille denne vejledning op så afsnittene kommer i en logisk rækkefølge i forhold til opgaven. Imidlertid er det en fordel at orientere sig i hele vejledningen, samt efter behov i referencerne, inden man går i gang med detailarbejde under de enkelte punkter. Det kan f.eks. være en fordel at have orienteret sig om de enkelte dataopbevaringsløsninger, enten i vejledningerne som findes via hjemmesiden [2], eller i den oversigtlige beskrivelse i ”Fælles ramme for dataopbevaring ved AU” [1].

Oversigt over definitioner findes i afsnit 9, og referencer til baggrundsoplysninger m.m. findes i afsnit [10].

# Vejen frem til implementering af dataopbevaringen

Inden for den fælles ramme for dataopbevaring ved AU er der fire forskellige løsningsmuligheder for et konkret dataopbevaringsbehov, og denne vejledning skal støtte systemejerens arbejde med at komme i gang med at finde den rigtige løsning. Der kan være mange ting at overveje ved valget mellem de fire løsninger, og i hvert enkelt tilfælde vil der sandsynligvis være et eller andet ”specielt” der skal findes ud af. Dette kan der ikke tages højde for på forhånd, men hvis man er gået igennem de overvejelser og forberedelser som skitseres her, så skulle man være godt rustet til at tage fat på implementeringsarbejdet i samarbejde med AU IT, samt BI-enheden hvis man vælger den dataopbevaringsløsning de tilbyder.

Se en grafisk oversigt over beslutningsprocessen for valg af løsning i Figur 1, og evt. vend tilbage til den efterhånden som resten af vejledningen læses.



Figur . Guide til valg af dataopbevaringsløsning. Start ved den grå bolle, og slut ved en grå bolle med ring om.

# Er datasamlingen bevaringsværdig?

Om data er bevaringsværdige er altid en vurderingssag, men konklusionen er ikke altid lige oplagt, og derfor er der udarbejdet et AU-bud på en definition af bevaringsværdighed for datasamlinger, idet de nedenstående betingelser kan bruges som guideline. Mindst én af betingelserne skal være gældende om et datasæt for at det er bevaringsværdigt:

* Der er et dokumenteret behov for fortsat at anvende dem internt i AU, eller hos samarbejdspartnere, som AU har en aftale med.
* Hvis en nærmere undersøgelse/vurdering ikke kan afvise, at der fortsat er behov for dem
* Der er en god begrundelse for, at de vil have værdi i fremtiden
* Hvis det vurderes, at de kan have værdi i kommende forskningsprojekter
* Som dokumentation for forskningsresultater
* Data er unikke, svære at genskabe og har en høj kvalitet
* Har retlig eller administrativ betydning
* Der er AU-arkivpligt for dem.
* Hvis Rigsarkivet har besluttet, at de er omfattet af afleveringspligten. Hvis man har data (forskningsdata, journaliseringsdata eller andre data), skal man henvende sig til Rigsarkivet, og oplyse hvilke data man har. Derefter vurderer Rigsarkivet, om disse data skal afleveres til Rigsarkivet.

Konklusionen om bevaringsværdigheden forudsættes dokumenteret inklusive begrundelsen/begrundelserne.

Hvis det konkluderes at datasamlingen ikke er bevaringsværdig, så er den umiddelbare konklusion at den kan udfases, dvs. slettes inden for en kort tidsfrist.

# Afleveringspligt til Rigsarkiv og identifikation af datasæt

Når en datasamling er fundet bevaringsværdig ifølge definitionen i afsnit 4 skal den vurderes i forhold til nogle overordnede spørgsmål bl.a. omkring afleveringspligt til Rigsarkivet og behov for videre anvendelse ved AU.

Hvis det konkluderes at der er behov for at dataopbevare ved AU, så skal dét datasæt, der skal bruges identificeres præcist, idet det ikke på forhånd er givet, at det er hele den oprindelige datasamling, der skal gemmes.

Bemærk at der også tages højde for dén situation, hvor AU – pga. en forpligtelse – er nødt til at gemme en datasamling i en periode, selvom vi vurderer at vi ikke selv har behov for at tilgå data. Hvis denne periode er så lang, at vi vurderer at det kan give problemer med versionsopdateringer af system og/eller database, så er vi nødt til at udarbejde en ”arkiveringsversion”, hvilket betyder at data lægges om til én af dataopbevaringsløsningerne, med tilhørende metadata, dokumentation og anvendelse af simple og/eller åbne standarder for filformater.

Nedenfor (i Figur 2) gives et overblik over de nødvendige overordnede overvejelser i starten af dataopbevaringsopgaven.

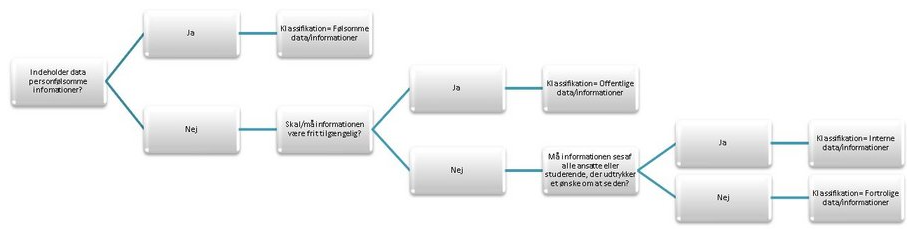
|  | **Spørgsmål/opgave vdr. datasamlingen** | **Svar og konklusion** | **Supplerende forklaring** |
| --- | --- | --- | --- |
|  | Er der afleveringspligt til Rigsarkivet? | Hvis ja, så skal der under alle omstændigheder foretages aflevering til Rigsarkivet.  Fortsæt med spørgsmål 2.  Hvis nej, så fortsæt med spørgsmål 3. | Afleveringspligten til Rigsarkivet indebærer at opdaterede arkiveringsversioner skal afleveres med nogle års mellemrum. |
|  | Er arkiveringen hos Rigsarkivet tilstrækkelig til at opfylde behovet for fremtidig adgang til datasættet? | Hvis ja, så skal endelig aflevering til Rigsarkivet foretages, og datasættet skal ikke dataopbevares hos AU.  Hvis nej, så fortsæt med spørgsmål 3 | AU vil – mod betaling – kunne få adgang til digitale arkivalier som vi selv har afleveret, såfremt de ikke indeholder oplysninger om personers private forhold. |
|  | Er der behov for fremtidig adgang til data? | Hvis nej, så vurdér om backup af den eksisterende datasamling kan løse opgaven, eller om det er nødvendigt at udarbejde en arkiveringsversion.  Hvis dataopbevaring skal foretages, så fortsæt ved punkt 5.  Hvis ja, så skal det afklares hvilken dataopbevaringsløsning, der dækker det aktuelle behov for brugeradgang.  Fortsæt ved spørgsmål 4. | Hvis et datasæt skal opbevares, selvom der ikke aktuelt er behov for at kunne tilgå data, så er opbevaringsperioden afgørende for valg af løsning. Er den tilstrækkelig kort kan man gemme en almindelig backup af data (og evt. system) i ”AU filarkiv”, idet man kan regne med at nuværende version af system vil kunne anvendes nogle år frem.  Er opbevaringsperioden længere er det nødvendigt at udarbejde en arkiveringsversion, som kan gemmes i ”AU filarkiv” |
|  | Hvilke krav er der til adgang til data? | Vurdér om behovet for dataadgang i virkeligheden er så omfangsrigt, at datasættet skal være en del af et system i drift, dvs. at det ikke er egnet til dataopbevaring.  Hvis konklusionen er at der skal dataopbevares, så fortsæt med punkt 5. | Hvis det f.eks. viser sig at behovet for adgang til data kræver en avanceret brugergrænseflade, eller adgang til data, som har performancemæssig betydning, er det ikke nødvendigvis hensigtsmæssigt at dataopbevare.  Således må man vurdere/estimere om ressourcerne til en kommende forvaltningsopgave samt evt. udviklingsopgave til en brugergrænseflade ligger indenfor hvad der er rimeligt. |
|  |  | Identificér datasættet. Dvs. dén delmængde af datasamlingen, der skal dataopbevares, idet det ikke nødvendigvis er hele datasamlingen der er behov for.  Sørg for at forudsætningerne for at foretage en dataopbevaring er på plads, og vælg en dataopbevaringsløsning. | *Se afsnittene 6 og 7.* |

Figur . Oversigt over vurdering af en bevaringsværdig datasamling. Udgangspunktet er, at den pågældende datasamling er vurderet til at være bevaringsværdig (se afsnit 4), idet den ellers skal udfases.

# Forudsætninger som datasættet skal opfylde

Før et datasæt kan dataopbevares med én af AU’s dataopbevaringsløsninger, skal nedenstående forudsætninger være på plads:

* Det skal være klart beskrevet hvilke data datasættet består af, idet det ikke nødvendigvis er hele den oprindelige datasamling.
* Datasættet skal være klassificeret i henhold til AU’s informationssikkerhedspolitik [4], og klassificeringen skal være verificeret/opdateret for højst et år siden. Resultatet skal være registreret i systemlisten eller datasamlingslisten på AU-reolen [3]. Vær især opmærksom på om datasættet indeholder persondata, eller om de er omfattede af kontrakter med andre institutioner eller virksomheder. Spørgsmålene i klassificeringen er oversigtligt vist i Figur 3.
* Data i datasættet er ”ikke-aktive”, dvs. de er statiske OG der er ikke andre systemer, der er afhængige af dem.
* Datasættet er dokumenteret.  
  Se de overordnede krav til metadata [1], samt de detaljerede krav til metadata i de enkelte vejledninger for dataopbevaringsløsningerne [2]
* Datasættet skal have en datasamlingsejer



Figur . Beslutningsdiagram for klassifikation af data (Kilde: au.dk under ”Informationssikkerhed”)

# Valg af dataopbevaringsløsning

Der tilbydes fire dataopbevaringsløsninger hos AU:

1. **AU filarkiv**
   1. Opbevaring i filformater på et filarkiv med læseadgang
   2. Opbevaringsperiode kan være 2, 7 eller 12 år, hvorefter data slettes
2. **BI’s datalake**
   1. En samling af kildesystemets tabeller med en kopi af kildedata
   2. Der kan foretages opslag i kildedata og evt. etableres rapportering til løsningen
3. **Oracle databaseopbevaring**
   1. Opbevaring i Oracle database
   2. Adgang til data via SQL-værktøjer, MS Excel, MS Access o.l.
   3. Evt. adgang via dedikeret web-baseret brugerflade med minimal funktionalitet
4. **MS SQL databaseopbevaring**
   1. Opbevaring i MS SQL database
   2. Adgang til data via SQL-værktøjer, MS Excel, MS Access o.l.
   3. Evt. adgang via dedikeret web-baseret brugerflade med minimal funktionalitet

**Fælles for dataopbevaringsløsningerne gælder:**

* Datasættets indhold kan gøres tilgængeligt for et begrænset antal navngivne brugere, typisk med en begrænset funktionalitet (der er dog en undtagelse ved BI-standardrapporter. Se Figur 4).
* Indholdet i datasættet er read-only, når det er placeret i dataopbevaringsløsningen.
* Datasættet bliver gemt i et aftalt antal år (opbevaringsperioden), dog således at forskellige delmængder af datasættet kan have forskellige opbevaringsperioder (bemærk AU filarkiv i Figur 4).
* Senest ved udløbet af den aftalte opbevaringsperiode adviseres datasamlingsejer om den kommende sletning af data, idet der kan indgås en ny dataopbevaringsaftale, hvis der i mellemtiden er opstået behov for forlængelse af datas levetid.

| **Karakteristika** | **1**  **AU filarkiv** | **2**  **BI’s datalake** | **3**  **ORACLE** | **4**  **MS SQL** |
| --- | --- | --- | --- | --- |
| Muligheder for adgang til data | Fungerer som filservice, dvs. adgang via ”Stifinder” o.l.  Læses kun sjældent | Adgang til standardrapporter for et væsentligt antal brugere.  Adgang via standard SQL-værktøj for et begrænset antal brugere. | Adgang via standard SQL-værktøj.  Evt. adgang via simpel, dedikeret web-baseret brugerflade. | Adgang via  MS Excel/MS Access.  Evt. adgang via simpel, dedikeret web-baseret brugerflade. |
| Mulig læsehastighed | Langsom, idet data skal hentes fra bånd | Der er onlineadgang | Der er onlineadgang | Der er onlineadgang |
| Mulig opbevaringsperiode | 2, 7, 12 år eller ”uendelig”.  Opbevaringsperiode bestemmes ved aftaleindgåelse | Opbevaringsperiode bestemmes ved aftaleindgåelse | Opbevaringsperiode bestemmes ved aftaleindgåelse | Opbevaringsperiode bestemmes ved aftaleindgåelse |
| Opbevaringsform | Filbaseret | Dataudtræk | Database/  Dataudtræk | Database/  Dataudtræk |

Figur . Karakteristika og kendetegn for de fire Dataopbevaringsløsninger.

Ved valget af en databaseopbevaringsløsning skal man først og fremmest overveje behovet for funktionaliteten i adgangen til data. Uanset om det aktuelle datasæt kommer fra en filbaseret eller databasebaseret løsning, så bør man gøre sig de samme overvejelser omkring krav til brugeradgang/bruger­grænse­flade, idet et filbaseret datasæt på dét grundlag kan lande i en databaseløsning og omvendt.

I overvejelserne om behovene for brugergrænseflade bør man bl.a. overveje:

* Skal data kunne tilgås af brugere, eller skal data blot kunne genskabes
* Skal data udtrækkes på forespørgsel, eller indgår de i beskrevne forretningsprocesser

De spørgsmål man skal besvare for umiddelbart at vælge en dataopbevaringsløsning fremgår af Figur 1.

# Implementering af dataopbevaring

Implementeringen af dataopbevaringen foregår i samarbejde med AU IT. Find kontaktinformationer på ”Dataopbevaring ved AU” på au.dk [2]. Samme sted kan der også findes detaljerede vejledninger til de enkelte løsninger.

En oversigt over datasamlingsejers opgaver i forbindelse med implementeringen fremgår af venstre kolonne tabellen i Figur 5. Man indgår aftale om én af de fire løsninger, og afhængig af hvilken løsning det drejer sig om, er det forskelligt hvem man skal have fat i. De allerfleste af datasamlingsejers opgaver er præcis de samme uafhængig af den valgte løsning, men ved hjælp af krydserne i tabellen kan man se hvor der er nogle undtagelser.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Opgaver for datasamlingsejer** | **BI’s**  **datalake** | **AU filarkiv** | **ORACLE** | **MSSQL** |
| Indgå aftale om dataopbevaring med: | BI  AU ØB | Gen App  AU-IT drift | Adm App  AU-IT drift | Adm App  AU-IT drift |
|  |  |  |  |  |
| Identificere data, der skal udtrækkes til opbevaring | x | x | x | x |
| Beskrive hvilke data datasættet består af | x | x | x | x |
| Udarbejde metadata/dokumentation | x | x | x | x |
| Verificér/opdatér klassifikationen af data | x | x | x | x |
| Teste og verificere data op mod kildedatabasen i samarbejde med AU IT | x | x | x | x |
|  |  |  |  |  |
| Fastlæg opbevaringsperioden | x | x | x | x |
| Afklar hvilke brugere, der skal have læseadgang til data via brugergrænseflade | x | x | x\* | x\* |
| Afklar hvilke brugere, der skal kunne foretage opslag i kildedata | x |  | x | x |
| Afklar hvilke brugere, der skal have adgang til løbende at lægge data i arkiv |  | x |  |  |
| Vælg hvilket/hvilke filformat, der skal anvendes for de dataopbevarede data |  | x |  |  |

Figur . Oversigt over datasamlingsejers opgaver.

\*) Kun i tilfælde hvor der er behov for etablering af en simpel, dedikeret web-baseret brugerflade

# Definitioner

Dette afsnit opsummerer definitioner af begreber, som anvendes i forbindelse med dataopbevaring ved AU. Formålet er at kunne beskrive dataopbevaring så præcist som muligt. Definitionerne er nærmere forklaret i [1].

## Definition af statiske data

Data der ikke vil komme ændringer til er ”statiske”.

## Definition af ikke-aktive data

”Ikke-aktive” data er statiske data, hvorom der er sikkerhed for, at der ikke er andre it-systemer som er afhængige af dem.

## Definition af AU-arkivpligt

Der er AU-arkivpligt for en datasamling, dvs. AU skal sørge for at bevare datasamlingen, hvis

* Det er angivet i lovgivning eller andre udefra kommende regler/forpligtelser
* Det er stadfæstet via interne regler, retningslinjer eller ledelsesbeslutning ved AU.

## Definition af opbevaringsperiode

Opbevaringsperioden for en datasamling er det tidsrum, som der maksimalt må gå inden der igen tages stilling til om data fortsat er bevaringsværdige.

## Definition af udløbstidspunkt

Udløbstidspunktet for en datasamling er dét tidspunkt, hvor både datasamlingen og de backup’er der findes af den, må slettes uden yderligere overvejelser/forespørgsler. Dvs. datas levetid slutter ved udløbstidspunktet.

## Definition af datasæt

Ved et datasæt forstår vi den samlede mængde af data, der arkiveres under én aftale om dataopbevaring.

## Definition af arkiveringsversion

En arkiveringsversion af et datasæt er en udgave af datasættet, der er ændret med det formål, at data med stor sikkerhed skal kunne læses, forstås og dermed genbruges efter mange år. Dette indebærer at data gemmes i simple og åbne standarder for filformater, samt med metadata-dokumentation i et simpelt tekstformat, f.eks. efter Rigsarkivets anbefalinger.

## Definition af bevaringsværdige data

Om data er bevaringsværdige er ofte en vurdering, og man kan anvende nedenstående betingelser som guideline, idet mindst én af betingelserne skal være gældende om et datasæt for at det er bevaringsværdigt:

* Der er et dokumenteret behov for fortsat at anvende dem internt i AU, eller hos samarbejdspartnere, som AU har en aftale med.
* Hvis en nærmere undersøgelse/vurdering ikke kan afvise, at der fortsat er behov for dem
* Der er en god begrundelse for, at de vil have værdi i fremtiden
* Hvis det vurderes, at de kan have værdi i kommende forskningsprojekter
* Som dokumentation for forskningsresultater
* Data er unikke, svære at genskabe og har en høj kvalitet
* Har retlig eller administrativ betydning
* Der er AU-arkivpligt for dem.
* Hvis Rigsarkivet har besluttet, at de er omfattet af afleveringspligten. Hvis man har data (forskningsdata, journaliseringsdata eller andre data), skal man henvende sig til Rigsarkivet, og oplyse hvilke data man har. Derefter vurderer Rigsarkivet, om disse data skal afleveres til Rigsarkivet.

# Referencer

1. **[Fælles ramme for dataopbevaring på AU](file:///C:\\confluence\\pages\\viewpage.action?pageId=30705787)** – Beskrivelse af dataopbevaringsløsninger.  
   Udarbejdet af projektet ”Etablering af løsninger til dataopbevaring på AU”, 2016.
2. **Dataopbevaring på AU**Et site under medarbejderservice på au.dk.  
   Vejledninger og andre oplysninger på dette site vedligeholdes af forvaltningsorganisationen for Dataopbevaring.  
   <http://medarbejdere.au.dk/administration/it/dataopbevaring/>
3. **AU-reolen**  
   AU IT’s fælles indgang til dokumentation.   
   Findes pr. oktober 2016 på Sharepoint: <http://projektportal.auit.au.dk/Implementering%20af%20AU%20IT/auitarkitektur/SitePages/AU-reolen.aspx>
4. **AU’s informationssikkerhedspolitik.**  
   Se AU’s hjemmeside under ”Informationssikkerhed. <http://www.au.dk/informationssikkerhed/informationssikkerhedspolitik/>

1. Oftest er det systemejeren, der også har ansvaret for den (eller de) datasamling(er) der hører til it-systemet, og derfor også er ”datasamlingsejer”. Det er imidlertid muligt at uddelegere ansvaret for en datasamling, og derfor er der brug for at kunne skelne mellem ”systemejer” og ”datasamlingsejer”. [↑](#footnote-ref-1)