

## Checklista för datahanteringsplaner

### Pass 2: Datahantering och datahanteringsplaner

*BAS Online 2021-01-20*

Att utveckla en datahanteringsplan för ett projekt är inte alltid en enkel uppgift. Det finns många detaljer som man åtminstone behöver överväga och det är lätt hänt att man glömmer bort en eller annan detalj. Som stöd i arbetet med datahanteringsplaner så har SND tagit fram en checklista för datahanteringsplaner. Checklistan uppdateras kontinuerligt och följer samma rekommendationer som Horizon 2020 och Science Europe's Core requirement och som även Sveriges universitets- och högskoleförbund (SUHF) och Vetenskapsrådet utgår från i sina rekommendationer. SND:s checklista är avsedd för svenska forskare och syftar till att vara så bred och heltäckande som möjligt samtidigt som den är anpassad efter svensk lagstiftning och etik.

Den senaste versionen av SND:s checklista kan hämtas från SND:s webbplats, antingen genom länken nedan eller genom att du går in under "Hantera data" på webbplatsen och klickar dig vidare.

Checklistan är ett verktyg som med fördel kan användas av forskare eller i mötet med forskare. Den inleds med en kort sammanfattning av datahantering och datahanteringsplaner, och försöker förklara goda anledningar till att använda datahanteringsplaner. Inledningen tar också upp att datahanteringsplanen är ett levande dokument som utvecklas under hela forskningsprojektet, som vi pratade om i en tidigare presentation.

Checklistan är indelad i sju övergripande områden med ett antal olika element inom varje område. Beroende på projekt och disciplin kan man behöva lägga fokus på olika områden, och

tanken med checklistan är att man ska kunna skaffa sig en överblick över vilka element som är relevanta för ett visst projekt.

För varje element finns det en beskrivning av vilken sorts information som det är tänkt att elementet ska innehålla. För de allra flesta element finns det även en motivering till varför det kan vara viktigt att inkludera just den här informationen i datahanteringsplanen. Motiveringen kan vara till hjälp för både en DAU-medarbetare i mötet med forskare och för en forskare i sitt aktiva arbete med datahanteringsplaner.

Forskaren tar ställning till om elementet är relevant för det egna projektet eller inte. Somliga element kan man direkt avgöra om de är aktuella, medan andra kanske går att ta ställning till först längre fram i forskningsprojektet. Anser man att informationen inte är relevant kryssar man för Nej i rutan till höger och går vidare, annars kan man kryssa för Ja, eller, om man är osäker, kryssa för Vet ej och fråga någon eller titta på det igen längre fram. Oavsett vad man väljer så har man i varje fall för varje element funderat igenom om just det planeringsbehovet finns.

Många forskare har påpekat att checklistan är väldigt omfattande. Det är den. Anledningen är att den är avsedd att omfatta ett brett fält av olika typer av data och forskningsområden. Tanken är att forskaren börjar med att snabbt gå igenom listan, markerar vad som är relevant och ignorerar resten. På så sätt avgränsar man vilken information som ska ingå i datahanteringsplanen.

Innehållet beror alltså på vilket forskningsämne och vilken typ av datamaterial det handlar om, och var i forskningsprocessen man befinner sig. Om finansören har krav på en datahanteringsplan i samband med ansökan kommer de kraven också att prägla innehållet.

Även om det är en fördel att datahanteringsplanen påbörjas i ett tidigt skede av forskningsprocessen så kan man givetvis börja ta

fram en plan även om projektet redan är igång. Det är i princip aldrig för sent att påbörja en datahanteringsplan. Även om man befinner sig mitt i ett projekt så finns det fördelar med att få ned saker på prânt så att man inte glömmer bort dem. På det viset blir datahanteringen enklare framöver.

Du kan ladda ned checklistan från SND:s webbplats och titta på den själv så vi ska inte gå in på den i detalj, men jag ska ta upp några saker om vart och ett av de övergripande områdena.

## Översikt

Det första området är *Översikt* och innehåller grundläggande projektinformation. Den här informationen finns ofta redan någon annanstans, till exempel i medelsansökan, så det är uppgifter som man ganska lätt kan kopiera. Det rör sig till stor del om information som man behöver ha på plats redan när projektet startar.

Vem som är *forskningshuvudman* är en uppgift som många forskare inte riktigt har klart för sig, trots att det är en högst väsentlig uppgift. Forskningshuvudmannen är den fysiska eller juridiska person i vars verksamhet forskningen utförs och kan till exempel vara ett lärosäte, en kommun, region, myndighet eller ett privat företag. Forskningshuvudmannen har bland annat det yttersta ansvaret för forskningen och är den som, om det behövs, ska ansöka om etikprövning. Det kan finnas flera forskningshuvudmän för ett projekt men då ska en utses som ansvarig för att ansökan om etikprövning görs.

Det kan också vara så att en forskare arbetar i ett projekt som är förlagt vid ett annat lärosäte, och då är det lärosätet där forskningen bedrivs som har ansvar för data. I en del samarbeten kan det finnas flera organisationer som ansvarar för olika delar av materialet och då behöver man vara lite försiktig med vem som får tillgång till vilka data. Sådana situationer kan bli lite komplicerade

när någon ska göra data tillgängliga, och det kan då vara värt att ta hjälp av en jurist för att reda ut saken i förväg.

Elementet *Medverkande personer och/eller organisationer samt deras roller* är relevant framför allt om det rör sig om ett större projekt. Om det är många människor inblandade så är det väldigt praktiskt om man i sin datahanteringsplan anger vem som har ansvar för olika delar i projektet. Då vet man vem man ska vända sig till och vem som ska följa upp att något skett korrekt, och inte minst att det finns någon som är ansvarig för datahanteringsplanen eller olika delar i den. Beskrivningen av roller är ingenting som måste ingå i en finansieringsansökan, men om det är utrett tidigt i projektet vet man om det behöver sökas finansiering för de här rollerna.

### Skydda forskningsdata

I avsnittet om *Skydda forskningsdata* finns flera delar som är viktiga att vara medveten om. För att säkerställa att datamaterialet hanteras i enlighet med gällande lagar och regler är det väsentligt att planera de här bitarna i god tid. Det mesta av det här togs upp i presentationen om datahanteringsplaner och det handlar alltså om t.ex. etikprövning, samtycke, skyddande av respondent och avtal.

I det här avsnittet kommer även informationsklassning och data-säkerhet in. Det här är viktiga delar och många tänker inte på att data ska hanteras säkert under tiden som projektet pågår. Data ska ju vara tillgängliga för dem som faktiskt ska arbeta med materialet samtidigt som obehöriga ska förhindras åtkomst.

Det handlar om olika säkra *tekniska* miljöer, säkra *fysiska* miljöer – exempelvis låsta rum och kassaskåp – och säkra *administrativa* miljöer, till exempel vem som har nycklar, lösenord, och rättighet att läsa och skriva i filerna. Vi går igenom mer om informationsklassning i nästa presentation om *Säkerhet och personuppgifter*.

Man behöver tänka på vad som är viktigt att ha på plats redan innan projektet startar, vad som är viktigt medan det pågår och vad som är viktigt för att det ska avslutas korrekt. Här kan det vara praktiskt om man har en jurist vid lärosätet att stämna av med så att man är på banan med alla de här delarna. Det finns även möjlighet att vända sig till SND med sådana frågor.

### Samla in/producera forskningsdata

Nästa del handlar om att beskriva de data som ska användas i forskningen.

*Befintliga data* är data som redan finns insamlade av någon och som skulle kunna återanvändas. Inte minst finansörer kan vara intresserade av att få veta om det redan finns data från ett annat projekt som kan (åter)användas och om man kan tänka sig att göra det istället för att en ny datainsamling ska finansieras.

Genom att specificera vilken eller vilka typer av data man ska samla in eller generera, och även när, var och hur insamlingen ska ske, så kan man förbereda sig för vilka resurser som kommer att behövas. Dels för själva insamlingen men också för bearbetningen och hanteringen efteråt.

### Dokumentera forskningsdata

Dokumentation återkommer vi till hela tiden: det är väsentligt att man dokumenterar vad som sker under processen. Dels för att man själv senare ska kunna förstå materialet när projektet är slut, och dels för att andra ska förstå vad man har gjort – så att forskningsresultat kan verifieras och replikeras men också för att möjliggöra sekundäranvändning. Här behöver man skapa rutiner för hur man ska dokumentera, vilket dokumentationsverktyg ska man använda till vad och vilka metadata som är relevanta för forskningen.

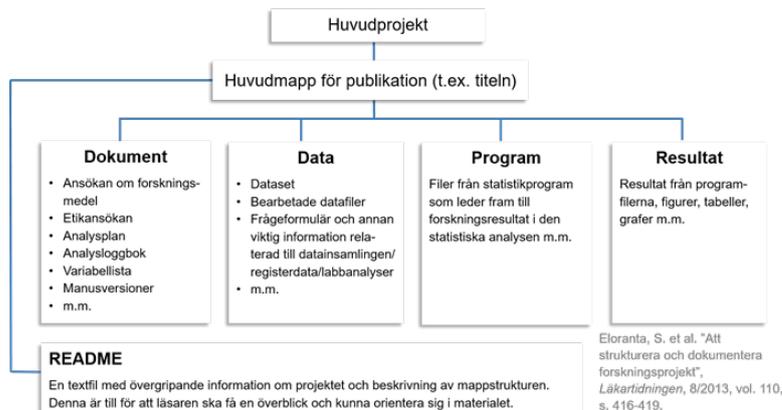
Om det finns någon lämplig metadatastandard som man kan följa är det bra att redan från början utgå från den. Då kan man också försäkra sig om att man dokumenterar sådant som är relevant under tiden som projektet pågår.

Det kanske finns terminologier som också är värda att fundera på tidigt. Inom medicin brukar man till exempel använda så kallade MeSH-termer. Vi kommer att prata mer om både metadata-standarder och användningen av terminologier längre fram i utbildningen, så jag nämner det bara här.

### Organisera forskningsdata

Beskrivningar om hur data organiseras under projektets gång blir ju en ganska stor del i datahanteringsplanen. Om man från början har en tydlig mappstruktur, har bestämt hur man ska arbeta med versionering och hur man ska namnge olika filer, så blir det mycket enklare att hitta de filer och versioner man letar efter och arbetet blir effektivare. Filnamn och versionering var vi inne på i en tidigare presentation, men jag tänkte säga något om mappstruktur.

Det här (se nästa sida) är ett exempel på en mappstruktur i en artikel som publicerats i *Läkartidningen*<sup>1</sup> av författare från Karolinska Institutet. Artikeln är tydlig och enkel, och sammanfattar många viktiga delar vad gäller dokumentation av ett forskningsprojekt. Även om den utgår från forskning på kliniska data så är principerna användbara för de flesta ämnesområden och kan lämnas ut till forskare som ett bra exempel.



Ju större ett forskningsprojekt är och ju fler personer som är inblandade, desto viktigare blir en genomtänkt mappstruktur. I det här exemplet finns en huvudmapp per publikation, som i sin tur innehåller mappar för dokument, data, program och resultat. Innehållet i mapparna är nogga avgränsat så att det ska bli enkelt att hitta det man letar efter. Det finns också en readme-fil med övergripande information och en beskrivning av mappstrukturen.

Man kan antingen beskriva mappstrukturen i sin datahanteringsplan eller så kan man hänvisa till ett annat dokument där man har beskrivit den mer utförligt. En datahanteringsplan behöver alltså inte nödvändigtvis innehålla all information utan man kan också referera till andra dokument.

## Kostnader för hantering av forskningsdata

Området *Kostnader* bör man ta hänsyn till redan vid ansökan så att man kan inkludera eventuella kostnader i den budget man söker. Det omfattar de kostnader som krävs för att man ska kunna hantera data på ett bra sätt under projektiden men även eventuella kostnader för det arbete som kan krävas för att data ska kunna bevaras och göras tillgängliga. Det rör sig både om kostnader för personal och för teknik och eventuellt kostnader för

utbildning i att hantera en viss programvara. Den säkerhet eller det lagringsutrymme som krävs kan vara kostsam, eller så kan man behöva införskaffa särskild hårdvara eller programvara för säkerhetskopiering eller dokumentation.

## Bevara och göra forskningsdata tillgängliga

Den sista delen handlar om att göra data tillgängliga. Trots att det står sist i checklisten så ska man inte vänta med att fundera kring de här frågorna. Tvärtom brukar det vara skönt att för egen del ha ordning på de bitarna tidigt, men även finansiären vill veta redan i ansökan hur man tänker kring det här med att göra data tillgängliga: hur och var tänker man göra data tillgängliga? Tänker man göra hela materialet tillgängligt eller bara delar av det? Finns det restriktioner eller embargon? Och så vidare. Datarepositorier blir särskilt glada över om man har tänkt till kring de här delarna, för det underlättar och snabbar på arbetet med att tillgängliggöra data.

Inte minst är det viktigt att tänka på om det behövs några åtgärder innan man kan göra materialet tillgängligt, som till exempel anonymisering – eller ska man redan från början designa datainsamlingen så att behovet av anonymisering blir så litet som möjligt? Vi kommer ta upp mer om personuppgifter och anonymisering i nästa presentation.

## Sammanfattning

Vi har i den här presentationen tittat närmare på SND:s checklista för datahanteringsplaner. Jag har kort gått igenom de olika delarna av checklisten och nämnt några av de aspekter som kan vara värda att fundera lite extra på, t.ex. huvudmannaskap, roller, dokumentation, mappstruktur och datasäkerhet.

Innehållet i en datahanteringsplan varierar beroende på vilket forskningsämne det rör sig om, var i forskningsprocessen man befinner sig, vilken typ av datamaterial det handlar om och om det

finns särskilda krav från finansiärer eller från policyer på lärosätet. Jag berättade att SND:s checklista är tänkt att fungera som stöd i arbetet med datahanteringsplaner, den är lite som ett smörgåsbord och man plockar de delar som är relevanta för det specifika projektet.

## Referenser

<sup>1</sup>Eloranta, S. et al. "Att strukturera och dokumentera forskningsprojekt", Läkartidningen, 8/2013, vol. 110, s. 416-419.

<https://lakartidningen.se/klinik-och-vetenskap-1/2013/02/att-strukturera-och-dokumentera-forskningsprojekt/>

(Hämtad 2021-01-20)