

## Datahanteringsplaner

### Pass 2: Datahantering och datahanteringsplaner

*BAS Online 2021-01-20*

Ämnet för den här presentationen är datahanteringsplaner, ett praktiskt verktyg som hjälper forskaren att få kontroll över hur data ska hanteras under forskningsprocessen och säkerställa att det finns tillräckliga resurser för att data hanteras väl. Datahanteringsplanen fungerar också som en hjälp för forskaren att komma ihåg hur t.ex. oplanerade händelser hanterades och vem som var ansvarig för detta i forskningsprojektet. En datahanteringsplan fylls på allteftersom och alltså ser en plan som skrivs i början av projektet annorlunda ut än mitt under projektet, eller när den färdigställs vid projektslutet. Syftet är att det ska vara ett dokument som forskaren använder som stöd för en ordnad hantering av data, men även för att lättare kunna planera för vad som ska hända efter projektet.

#### Ansökan

Det finns allt fler finansiärer, både nationellt och internationellt, som har krav på att en datahanteringsplan ska ingå när man skriver en ansökan om forskningsmedel, men det finns även de som enbart begär in en datahanteringsplan av de som får finansiering. De datahanteringsplaner som i dagsläget krävs in som en del av ansökan ska antingen vara en bilaga till ansökan eller så är de skrivna som en del i själva ansökan. I båda fallen rör det sig om planer med begränsat innehåll. Syftet med att lämna in en datahanteringsplan i ansökan är främst att visa finansiären att man är medveten om eventuella behov och problem som kan dyka upp.

Oavsett om en datahanteringsplan behövs för ansökan eller inte så fyller en plan tidigt i projektet en annan, högst väsentlig, funktion:

den ger forskarna en anledning till att tänka igenom de behov och kostnader som kan uppstå på grund av datahantering. Vissa delar av datahanteringen kostar pengar, och genom att skriva en datahanteringsplan i ett tidigt skede så kan man lägga in resursbehoven i projektbudgeten. Kanske behöver man köpa in hårdvarukrypterade hårddiskar och låsbara skåp, eller betala för en datahanterare i projektet.

*Under forskningsprojektet* är datahanteringsplanen tänkt att fungera som en arbetsplan och är ett levande dokument som kompletteras och utvecklas under projektets gång. Samtidigt bör en ny version av datahanteringsplanen skapas varje gång det sker en förändring så att man lätt kan gå tillbaka till tidigare versioner om man undrar något, till exempel se över specifika detaljer som rörde ett visst beslut. Det är viktigt att understryka att det i princip aldrig är för sent att påbörja en datahanteringsplan. Hellre att en datahanteringsplan skrivs halvvägs in i projektet än inte alls.

Planen ska vara så detaljerad som den behöver vara för att man ska kunna hålla ordning på datahanteringen i projektet. Den bör omfatta exempelvis rutiner för säkerhetskopiering, namnsättning av dokument och filer, principer för versionering och dokumentation av olika beslut och varför man fattat dessa. Helst ska man också veta vem som ansvarar för vad i projektet, inklusive vem som uppdaterar datahanteringsplanen. Beslut från början av projektet kan komma att förändras, till exempel att man vill inkludera en ny sorts analys som kräver andra datatyper. Behov kan tillkomma, situationer förändras, eller så inser man helt enkelt att allt inte fungerade på det sätt man ursprungligen tänkt sig.

I slutet av forskningsprocessen behöver man se till att datahanteringsplanen innehåller all väsentlig information om hur data ska bevaras långsiktigt, hur de ska göras tillgängliga och vad som ska förändras med data inför arkivering och långtidsbevarande. Här bör noteras att det är bra att redan tidigt i projektet planera inför det

som ska hända med data efter projektslut så att det inte uppstår problem som kunde ha lösts enklare redan i ett tidigt skede av projektet.

## Horizon 2020 och Horizon Europe

Sedan 2016 ställer EU:s ramprogram Horizon 2020 höga krav både på planen som bifogas vid ansökningstillfället och senare versioner där den utvecklas tillsammans med forskningsprojektet. En Horizon 2020-ansökan ska innehålla information om vilka standarder man tänker använda och hur man tänker dela eller göra data tillgängliga. Om man inte kan dela eller göra data tillgängliga så ska man ange varför. Man ska även beskriva hur data är tänkta att kureras och långtidsbevaras så att de kan utnyttjas i framtiden. Inom det nya ramverket Horizon Europe, som börjar gälla från 2021, går man ännu längre. Datahanteringsplanen blir nu obligatorisk och all datahantering ska följa riktlinjerna för FAIR. Man poängterar hur viktigt det är att forskningsresultat och data ska komma samhället tillgodo. Forskningsdata ska göras tillgängliga enligt principen att de är "så fria som möjligt och så begränsade som nödvändigt".

Öppna data är grundinställningen om det inte finns juridiska hinder som förhindrar detta, och en "aktiv" datahanteringsplan är en obligatorisk leverabel, det vill säga ett dokument som ska lämnas in och godkännas som en rapport eller ett resultat från projektet. En första version av datahanteringsplanen ska lämnas in under projektets första sex månader.

Datahanteringsplanen för Horizon-projekt ska omfatta hanteringen av data, vad det är som samlas in och i vilken omfattning, vilka metoder och standarder som tillämpas för datahanteringen och hur data ska göras tillgängliga och långtidsbevaras. En första version av planen ska alltså lämnas in under det första halvåret, så man behöver med andra ord sätta igång med att ta fram planen tämligen omgående. Tanken är förstås att man vill uppmuntra till

att datahanteringsplanen används i projektet och att planen hålls levande och uppdaterad. Den ska också uppdateras åtminstone inför halvtidsrapportering och slutrapportering. Inom Horizon 2020 och Horizon Europe försöker man med andra ord att genom krav få datahanteringsplaner att skrivas, användas och uppdateras inom projekten. Utifrån de H2020-datahanteringsplaner som man kan hitta online finns det emellertid anledning att tro att inte alla projekt utnyttjar planerna på bästa sätt. Kanske höjs nivån med de nya kraven som följer med Horizon Europe. Du hittar en länk till mer information om datahantering i Horizon 2020 under presentationen.

### Innehåll i datahanteringsplaner

Vad är det då som kan ingå i en datahanteringsplan? Jag säger *kan*, för exakt innehåll beror på vilken disciplin det rör sig om, hur projektet är utformat och vilka data som samlas in i det. SND har tagit fram en checklista som ett stöd för att skriva datahanteringsplaner, men den kommer vi till i en senare presentation. Här får du en kort översikt över vad som kan tas upp i en datahanteringsplan.

- Någon form av databeskrivning bör finnas med: vad det är för data man samlar in/genererar, hur mycket man planerar att samla in/generera och så vidare. Ytterst påverkar det här vilka tekniska behov man har för att kunna sköta insamling och lagring av data – behöver man stora hårddiskar eller ett enormt serverutrymme för att kunna hantera datamängderna, eller räcker det med vad man normalt har tillgång till på sitt lärosäte?
- Man bör även nämna något om huruvida det finns andra dataset som redan har samlats in av en själv eller någon annan och som kan (åter)användas i projektet. Det här är i princip motsvarigheten till att ta upp tidigare forskning i en vetenskaplig publikation, det vill säga en fråga om god forskning lika mycket som om kostnader.

- Lagring och datahantering på kort sikt – det vill säga under själva projekttiden – är också något som behöver övervägas, och så är det förstås de etiska och juridiska övervägandena. Finns det över huvud taget några sådana? Datahanteringsplanen är ett bra ställe att börja fundera kring detta:
  - Man ska åtminstone ställa sig frågan om forskningshuvudmannaskap: det är en enkel fråga om man är en forskare på ett lärosäte, för då är det lärosätet som är huvudman. Men så fort projektet inkluderar flera lärosäten eller parter så behöver man fundera över vem som är huvudman för vilka data, och vem som skickar in etikansökan. Om man använder data från andra myndigheter, från vården, privata företag, kanske från främmande länder, vem är då huvudman? Behöver man upprätta avtal om det?
  - Kommer man att handskas med personuppgifter eller kanske till och med känsliga personuppgifter? Behöver man använda informerade samtycken? Tänker man utföra djurförsök? Då behöver det framgå vem som är ansvarig för ansökan om etikprövning.
  - Undersök hur det förhåller sig med informationsklassning för projektets data: hur känsliga data rör det sig om, och hur behöver de hanteras på grund av det? De flesta lärosäten har riktlinjer på det här området men inte alla.
  - Finns det frågor om upphovsrätt som man behöver förhålla sig till?
  - Riskerar det att förekomma Non-Disclosure Agreements, avtal om att man inte får sprida information i data hur som helst för att det kan finnas företagshemligheter eller mot-svarande? I så fall,

vem kan underteckna sådana å projektets vägnar?  
Det finns ett fall där en doktorand skrev under ett Non-Disclosure Agreement som sedan visade sig vara så strängt formulerat att intervjuerna som utfördes inte kunde användas, eftersom risken helt enkelt var för stor att universitetet skulle kunna bli ersättningskyldigt.

- Hur tänker man kring att göra data tillgängliga? Vilket är det bästa sättet för det material man ska samla in, och behöver man planera insamlingen så att det faktiskt är möjligt? Kan delar av en databas publiceras? Det kan vara en god idé att fundera på det här i förväg eftersom det eventuellt kan finnas krav på publicering av data i vissa tidskrifter, och då är det bra om man tagit hänsyn till kraven.
- Arkivering och bevarande bör planeras redan i början av projektet så att man gör det lätt för sig att bevara på lång sikt – och kom ihåg att för svenska arkiv är den långa sikten i princip obegränsad, för all framtid.
- Det är också en bra idé att i en datahanteringsplan för ett större projekt skriva in vem som ansvarar för datahantering i projektet.

Kom ihåg att datahanteringsplaner ska fylla två syften: det ena syftet är att de ska fungera som stöd för forskare i deras datahantering och ska göra forskningen effektivare och enklare. Det andra är att de ska förse arkiv, datarepositorier och sekundär-användare av data, dvs. andra forskare, med relevant information. Att glömma det ena eller andra syftet riskerar att sluta i en plan som inte är användbar för någon. En forskare som betraktar datahanteringsplanen som en administrativ pålaga har inte stor nytta av den!

## Sammanfattning

I den här presentationen har jag tagit upp att en datahanteringsplan är ett verktyg som kan användas före, under och efter ett projekt för att planera och beskriva hur data ska hanteras. Den ser olika ut beroende på var i forskningsprocessen som den kommer in, men under projektet är den ett levande dokument som ska uppdateras efter behov. Den kan innehålla diverse information beroende på projektets disciplin, storlek, datatyper med mera, men det är viktigt att komma ihåg att datahanteringsplanen ska vara till nytta såväl under projektet som efteråt när data lagras och återanvänds.

Nästa presentation kommer att ta upp några centrala områden inom datahantering.

## Referenser

DMP Online:

<https://dmponline.dcc.ac.uk/>

European Commission. Research & Innovation.

*Participant Portal H2020 Online Manual:*

[https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)