

# Allmänna anvisningar för datahanteringsplaner 2020

Hur man skriver en datahanteringsplan (data management plan, DMP) - LÄS DETTA FÖRST!

- Läs igenom alla frågor först!
- Datahanteringsplanen ska komplettera din forskningsplan – besvara de frågor som är tillämpbara på ditt projekt, men undvik överlappningar med forskningsplanen.
- I forskningsplanen beskrivs de vetenskapliga metoderna och analysen av datamaterialet. I datahanteringsplanen beskrivs teknisk och administrativ behandling av datamaterialet.
- För att undvika upprepningar kan du hänvisa till din forskningsplan i datahanteringsplanen och vice versa.
- Använd datahanteringsplanen som ett riskbedömningsdokument: visa att du kan identifiera, förutse och hantera eventuella risker i anknytning till behandlingen av data inom ditt forskningsprojekt.
- Skriv datahanteringsplanen utgående från ditt forskningsprojekt. Kopiera inte in exempel från andra texter.
- Skriv bara sådana meningar som du själv förstår.
- Svara på frågorna i det fall att de är relevanta för ditt forskningsprojekt. Om en fråga inte är relevant, motivera varför.
- Svara åtminstone på frågorna i huvudpunkterna. Du behöver inte svara på alla underpunkter separat.
- Lägg till bakgrundsinformation såsom exempelvis den sökandes och projektets namn, projektnummer, finansieringsprogram och datahanteringsplanens version.
- Visa din förmåga till datahantering och dokumentation exempelvis genom att namnge datahanteringsplanens olika versioner på ett tydligt sätt.
- Följ din organisations eller finansiärens anvisningar.

Varför är datahantering och att skriva en datahanteringsplan (DMP) viktigt?

- Det är god forskningspraxis!
- Risken att förlora forskningsdata minskar.
- I planen kan du redan i förväg ta ställning till eventuella komplicerade frågor som gäller ägande- och nyttjanderätt.
- Planen är ett stöd för öppen tillgång och främjar produktivt framtida samarbete.
- Du bemöter krav från finansiärerna.
- Du sparar tid och pengar.
- Datahanteringsplanen visar dina projektledarfärdigheter.

Data ett brett begrepp inom DMP-kontexten. Data inbegriper all information och allt material som dina forskningsresultat grundar sig på. Du kan koncentrera dig på sådana data som är ditt ansvarsområde.

Datahanteringsplanen ska beskriva hur forskningsdata kommer att hanteras genom forskningens hela livscykel. Datahanteringsplanen är ett levande dokument som bör uppdateras allteftersom forskningsprojektet framskrider.

Rutinerna för hantering av forskningsdata ska följa FAIR-principerna, vilket innebär att forskningsdata ska vara sökbara (Findable), tillgängliga (Accessible), interoperabla (Interoperable) och återanvändbara (Reusable).

Lycka till med datahanteringsplanen!

Allmänna anvisningar för datahanteringsplaner 2020

Tuuli projekt: <https://wiki.helsinki.fi/x/b3eZCQ>

Uppdaterad 24.1.2020



## 1. Allmän beskrivning av data

1.1 Vilka slags data baserar sig forskningen på? Hurdana data samlas in, produceras eller återanvänds? Vilka filformat används? Ge även en grov uppskattning av hur mycket data som kommer att produceras eller samlas in.

Beskriv kort vilka slags data som samlas in eller produceras. Ange också vilka slags befintliga data som kommer att (åter)användas. Till exempel texter, bilder, fotografier, mätningar, statistik, fysiska prover eller koder.

Kategorisera dina data genom att dela in dem på ett åskådligt sätt, till exempel:

- A) data som samlas in för detta projekt
- B) data som är ett resultat av processen
- C) befintliga data som insamlats tidigare och återanvänds i detta projekt
- D) projektledningens dokument samt projektrapporter, osv.

Kategoriseringen ska följa licensriktlinjerna för dina dataset. Beskriv till exempel kortfattat i enlighet med vilken licens du har rätt att (åter)använda data. Kategoriseringen kan utgöra en generell struktur för resten av datahanteringsplanen.

Ange filformat för alla data. I vissa fall kan formatet som används under forskningsprojektet vara ett annat än det format som används vid arkiveringen av datamaterialet. Ange båda formaten. Filformatet är av central betydelse för framtida tillgänglighet och återanvändbarhet av data.

Beskriv i datahanteringsplanen hur mycket lagringsutrymme som behövs för dina data – inte hur många informanter som deltagit i projektet. Det räcker med en grov uppskattning av storleken på dina data, till exempel mindre än 100 gigabyte, ungefär 1 terabyte, eller flera petabyte.

Rekommendationer för bästa praxis

- Använd tabeller eller punktuppställningar för att koncist presentera typer av data, filformat, programvara, och storlek på data.
- Exempel på filformat: .csv, .txt, .docx, .xlsx, .tif.
- När du listar vilka filformat som kommer att användas ska du komma ihåg att nämna all slags särskild eller ovanlig programvara som krävs för att läsa eller använda data, i synnerhet om programvaran kodas inom ramen för projektet.
- Du kan även uppskatta ökningen av data under projektets gång, exempelvis per vecka: "Projektet producerar/samlar in cirka 100 gigabyte data per vecka."
- **UNDVIK ÖVERLAPPNING MED FORSKNINGSPLANEN!** Dataanalysen och metodologiska frågor som gäller projektets data ska beskrivas i forskningsplanen.

## 1.2 Hur kontrolleras datamaterialets enhetlighet och kvalitet?

Beskriv hur metoderna som används för att samla in, analysera och bearbeta data kan påverka kvaliteten på projektets data och hur du ämnar minimera riskerna.

Kvalitetskontroll av data säkerställer att data inte ändras av misstag och att kvaliteten upprätthålls under hela livscykeln. Kvalitetsproblem kan uppstå i samband med den tekniska hanteringen, konverteringen eller överföringen eller i samband med kontextuell hantering och analys.

Rekommendationer för bästa praxis

- Transkriptioner av ljudupptagningar eller videointervjuer ska kontrolleras av en annan person än den som transkriberat dem.
- Analogt material ska digitaliseras med så hög resolution som möjligt för precision.
- Vid alla konverteringar ska det ursprungliga informationsinnehållet säkerställas.
- Programvara som använder sig av kontrollsummor ska användas.
- Ordna övningsessioner och skriv anvisningar för att säkerställa att alla i din forskningsgrupp är insatta i kvalitetskontroll och kan förutse risker i anknytning till kvaliteten på data.
- UNDVIK ÖVERLAPPNING MED FORSKNINGSPLANEN! Detaljer i anknytning till dataanalys, metoder och verktyg ska beskrivas i forskningsplanen. Du ska till exempel inte inkludera beskrivningar av instrumentens kalibrering i datahanteringsplanen.

## 2. Iakttagande av etiska principer och lagstiftning

2.1 Vilka juridiska frågor är relevanta för datahanteringen? (Till exempel GDPR och annan lagstiftning som påverkar behandling av data.)

Frågor i anknytning till rättigheter samt juridiska och etiska aspekter omfattar all forskningsdata. Visa att du känner till relevant lagstiftning i anknytning till behandlingen av dina forskningsdata. Om du hanterar personuppgifter eller känsliga uppgifter ska du beskriva hur deltagarnas identitet kommer att skyddas och hur du kommer att anonymisera eller pseudonymisera data.

Rekommendationer för bästa praxis

- Bekanta dig med din organisations etiska riktlinjer och datasäkerhetspolicy och var beredd att följa bestämmelserna i dem.
- Om en etikprövningsnämnd ger ett utlåtande om din forskning ska du i datahanteringsplanen ange hur du ämnar följa direktiven (t.ex. hur deltagarnas identitetsskydd tryggas genom att man avlägsnar känsliga eller personliga uppgifter ur data innan de görs tillgängliga).
- Inbegriper din forskning försökspersoner?
- Ingår behandling av personuppgifter? Om du kommer att behandla personuppgifter, ange vilken typ av data du ämnar samla in.

- Med personuppgifter avses alla uppgifter som anknyter till en identifierad eller identifierbar person. Personuppgifter är till exempel namn, telefonnummer, positionsuppgifter och uppgifter om far- och morföräldrars genetiska sjukdomar.
- Dataombudsmannens byrå (<https://tietosuoja.fi/sv/behandling-av-personuppgifter>)
- UNDVIK ÖVERLAPPNING MED FORSKNINGSPLANEN! Detaljer gällande etiska frågor, liksom utlåtanden från etiska kommittéer och användning av försöksdjur ska beskrivas i forskningsplanen.

## 2.2 Hur hanteras rättigheterna till det datamaterial du använder, producerar och delar?

Ange hur ni kommer överens om nyttjanderätten för projektets data, såväl insamlat, producerat som (åter)använt datamaterial? Här kan du använda kategoriseringen du gjort upp under punkt 1. Varje kategori inbegriper olika rättigheter och licenser. Beskriv de processer i anknytning till överföring av nyttjanderätter som är relevanta för ditt projekt. Beskriv frågor i anknytning till konfidentialitet om det är relevant för ditt projekt.

### Rekommendationer för bästa praxis

- Kontrollera din organisations policy i fråga om äganderätt, nyttjanderätt och rätten att sprida data.
- Har du fått samtycke för förvaring och tillgängliggörande av data?
- Avtal om äganderätten bör ingås i ett så tidigt skede av projektets livscykel som möjligt.
- Beakta också finansiärens principer för upphovsrätt och immaterialrätt.
- Det rekommenderas att all data, alla koder och all programvara görs tillgängliga för återanvändning t.ex. genom Creative Commons (<https://creativecommons.org/choose/>), GNU (<https://www.gnu.org/licenses/gpl-3.0.en.html>), MIT (<https://opensource.org/licenses/MIT>) eller en annan ändamålsenlig licens.

## 3. Dokumentation och metadata

3.1 På vilket sätt dokumenterar du ditt datamaterial så att det är sökbart, tillgängligt, interoperabelt och återanvändbart? Vilka metadatastandarder, README-filer eller annan dokumentation kommer att användas för att andra ska kunna förstå och återanvända datamaterialet?

Dokumentationen av data möjliggör att andra (både människor och datorer) kan hitta, använda och citera dataset och filer. Metadata är väsentlig information som t.ex. berättar var, när, varför och hur datamaterialet samlades in, bearbetades och tolkades. Utan ändamålsenlig dokumentation är ditt datamaterial värdelöst. Beskriv verktyget, exempelvis Qvain, som du ämnar använda för att dokumentera dina dataset. Du behöver inte nämna metadatastandarder om du inte använder sådana. Förutse öppnandet av ditt projekts metadata eller dokumentation redan här. Du ska dock ge en detaljerad beskrivning av öppnandet i punkt 5.

UNDVIK ÖVERLAPPNING MED FORSKNINGSPLANEN! Dokumentation på datanivå (data-level) (<https://www.ukdataservice.ac.uk/manage-data/document/data-level.aspx>) och detaljer i fråga

om experiment, analysmetoder och den vetenskapliga kontexten ska beskrivas i forskningsplanen. I datahanteringsplanen ska du beskriva dokumentationen på projektnivå (study-level) (<https://www.ukdataservice.ac.uk/manage-data/document/study-level.aspx>).

Rekommendationer för bästa praxis

- Beskriv alla dokumentationstyper som kommer att användas (t.ex. README-filer, metadata o.d.) för att andra ska kunna hitta, förstå och återanvända ditt datamaterial.
- När du följer FAIR-principerna (<https://www.force11.org/group/fairgroup/fairprinciples>) säkerställer du att ditt datamaterial är sökbart, tillgängligt, interoperabelt och återanvändbart.
- Minimikrav för dokumentation av datamaterial, se till exempel Qvain Light (<https://www.fairdata.fi/en/qvain/qvain-light-user-guide/>).
- Använd sådana forskningsverktyg som automatiskt skapar metadata i standardiserade format.
- Identifiera vilken slags information som ska sparas för att andra forskare ska kunna hitta, få tillgång till, tolka, använda och citera dina forskningsdata.

#### 4. Lagring och säkerhetskopiering under forskningsprojektet

##### 4.1 Var kommer datamaterialet att lagras och hur sker säkerhetskopieringen?

Ange var datamaterialet lagras och säkerhetskopieras under forskningsprojektet. Beskriv metoderna för bevaring och tillgängliggörande av datamaterialet efter att forskningsprojektet har avslutats utförligare i avsnitt 5.

Fundera över vem som ska ansvara för säkerhetskopiering och återställande. Om flera forskare är involverade ska du göra upp en plan med dina samarbetspartner och säkerställa en trygg överföring mellan de medverkande.

Visa att du känner till vilka lagringsmöjligheter din organisation erbjuder. Det räcker inte med att hänvisa till IT-tjänster. Slutligen är det du – inte IT-avdelningen eller din organisation – som ansvarar för ditt datamaterial.

Rekommendationer för bästa praxis

- Det rekommenderas att du använder dig av en säker förvaringsplats som erbjuds och förvaltas av din egen organisations IT-avdelning.
- ANVÄND INTE externa hårddiskar som primär lagringsplats.

##### 4.2 Vem reglerar åtkomsten till datamaterialet och hur övervakas skyddad åtkomst?

Det är ytterst viktigt att trygga datasäkerheten, i synnerhet om datamaterialet innehåller känsliga data, personuppgifter, politiskt känslig information eller affärshemligheter. Ange vilka som har

tillgång till ditt datamaterial, vad de har rätt att göra med det och hur du säkerställer att data tryggt överförs till dina samarbetspartner.

Rekommendationer för bästa praxis

- Åtkomsten till datamaterialet ska regleras och övervakas enligt graden av konfidentialitet.

## 5. Öppna, publicera och arkivera data efter avslutat forskningsprojekt

5.1 Vilken del av datamaterialet kan göras öppet tillgängligt eller publiceras? Var och när kommer datamaterialet eller dess metadata att göras tillgängliga?

Beskriv huruvida du har för avsikt att publicera datamaterialet eller på annat sätt göra det öppet tillgängligt delvis eller i sin helhet. Om datamaterialet delvis eller i sin helhet inte kan göras öppet ska du motivera varför.

Om ditt projekt innehåller känsliga data som inte kan öppnas, beskriv hur projektets metadata kommer att öppnas. Beskriv hur känsliga uppgifter lagras på ett säkert sätt under punkt 5.2.

Öppen tillgänglighet främjar återanvändningen av forskningsdata.

Rekommendationer för bästa praxis

- Du kan publicera en beskrivning av datamaterialet (dvs. metadata) utan att göra själva datamaterialet öppet tillgängligt. På så sätt kan du begränsa tillgängligheten.
- Publicera datamaterialet i ett datarepositorium eller i en datapublikation.
- Sök lämpliga repositorer på [re3data.org](https://re3data.org).
- Kom ihåg att kontrollera finansärens och vetenskapsområdets rekommendationer för datarepositorier och också de nationella rekommendationerna.
- Det rekommenderas att all data, alla koder och all programvara görs tillgängliga för återanvändning t.ex. genom Creative Commons (<https://creativecommons.org/choose/>), GNU (<https://www.gnu.org/licenses/gpl-3.0.en.html>), MIT (<https://opensource.org/licenses/MIT>) eller en annan ändamålsenlig licens.
- Sträva till att använda repositorer eller förläggare som ger permanenta identifikationer (t.ex. DOI, URN) till ditt datamaterial.
- **BLANDA INTE IHOP MED PUBLIKATIONSPLANEN!** Publikation av en vetenskaplig artikel är inte samma sak som publikation av datamaterial. Datapublikationer är specialiserade forum för publikation av forskningsdata.

5.2 Var bevaras datamaterial av långsiktigt värde och hur länge?

Beskriv kort vilka data som ska arkiveras och hur länge. Kategorisera dina dataset i enlighet med den planerade förvaringsperioden:

- A) Data som förstörs när projektet har avslutats.
- B) Data som arkiveras under en verifieringsperiod, som kan variera enligt vetenskapsområde; 5–15 år.
- C) Data som arkiveras för eventuell återanvändning; i t.ex. 25 år.
- D) Data av långvarigt värde som arkiveras i en kuraterad tjänst för framtida generationer i tiotals eller hundratals år.

Beskriv vilka data som förstörs efter att projektet har avslutats och hur dessa kommer att förstöras. Beskriv principerna för det arkiverade datamaterialets tillgänglighet. Överväg att anlita arkiv med riktlinjer för kuratering.

Rekommendationer för bästa praxis

- Kom ihåg att kontrollera både nationella och finansiärens och vetenskapsområdets rekommendationer för datarepositorier.

## 6. Datahanteringsansvarsområden och resurser

6.1 Vem (t.ex. roll, ställning och enhet) ansvarar för specifika uppgifter i datahanteringen under projektets livscykel (den dataansvariga)?

Här kan du sammanfatta alla ansvarsområden som beskrivits i föregående frågor.

Rekommendationer för bästa praxis

- Beskriv datahanteringsroller och ansvarsområden i fråga om exempelvis lagring av data, produktion av metadata, datakvalitet, lagring och säkerhetskopiering, arkivering och tillgängliggörande av data. Namnge ansvariga personer när det är möjligt.
- I fråga om samarbetsprojekt, förklara hur datahanteringen koordineras och vilka som är ansvariga inom de olika samarbetsorganisationerna.
- Ange vem som ansvarar för genomförandet av datahanteringsplanen, och för att den granskas och uppdateras vid behov.
- Planera in regelbunden granskning av datahanteringsplanen.
- Fastställ också vem som är ansvarig för datamaterial från ditt projekt efter att det är avslutat.

6.2 Vilka resurser behövs för hanteringen av ditt datamaterial för att data ska kunna öppnas och förvaras i enlighet med FAIR-principerna (sökbart, tillgängligt, interoperabelt och återanvändbart)?

Uppskatta vilka resurser som behövs (till exempel tid och kostnader) för hantering, förvaring och tillgängliggörande av data. Beakta vilken typ av extra IT-tjänster och resurser som behövs, och vilka extra kostnader som dessa innebär.

## Rekommendationer för bästa praxis

- Kom ihåg att specificera datahanteringskostnaderna i budgeten, enligt finansiärens krav.
- Förklara hur nödvändiga resurser (till exempel tid) som behövs för att bereda datamaterial för tillgängliggörande/förvaring (kuratering) har beaktats i kostnadskalkylen. Beakta noggrant och motivera alla eventuella resurser som behövs för att leverera data. Dessa kan inkludera kostnader för förvaring, hårdvara, personalresurser (tid), kostnader för beredning av data för förvaring, och kostnader för användning av repositorer.