

Kvalitet och genomslag i forskningsarbete med hjälp av datahantering

Hur dina data blir FAIR



Öppen
vetenskap



Innehåll

Inledning



Datahantering som en del av forskningen



Utvärdering av datavärde och -kvalitet



Dela och lagra data

Vad är forskningsdata?

Med forskningsdata avser vi här datamaterial som uppstår och/eller som utnyttjas inom forskning.

Med livscykel för data avses alla faser för förekomst av digitala data, från skapande till långtidslagring och förstöring.

Hantering av digitala material kräver planering och kompetens. Icke-hanterat material kan bli korrupt av sig själv eller så kan det komma i ordning eller förstöras.

Ansvarsfull forskning förutsätter transparens och möjlighet att replikera eller kontrollera det utförda arbetet. Data och program är en viktig del av forskning.

Vad innebär FAIR-principer?

Hittbarhet (*Findable*), tillgänglighet (*Accessible*), interoperabilitet (*Interoperable*) och återanvändningsbarhet (*Reusable*) är mål som när de beaktas stöder god datahantering samt forskningens kvalitet och verkningsfullhet

- Mer om Fair-principerna: <https://www.go-fair.org/fair-principles/> ja <https://www.fairdata.fi/tietoa-fairdatasta/fair-periaatteet/>

FAIR-principerna är avsedda för alla data, såväl kvantitativa som kvalitativa, och syftar till maskinläsbarhet och interoperabilitet för data.

- Samma principer gäller även metadata, med vilka data beskrivs och görs sökbara.

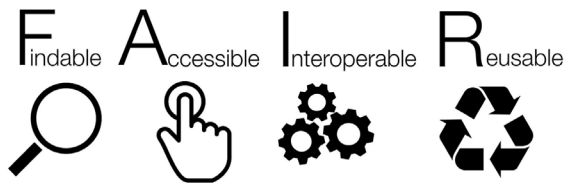
FAIR-data är förståeliga för människan och kan även hanteras maskinellt. Till exempel är en läsarvänlig tabell i en pdf-fil inte enkel att redigera, medan en csv-fil är det. Väl strukturerade data kan till exempel kombineras med andra data av samma format eller så kan sökningar riktas till dem.

FAIR-principerna och forskaren

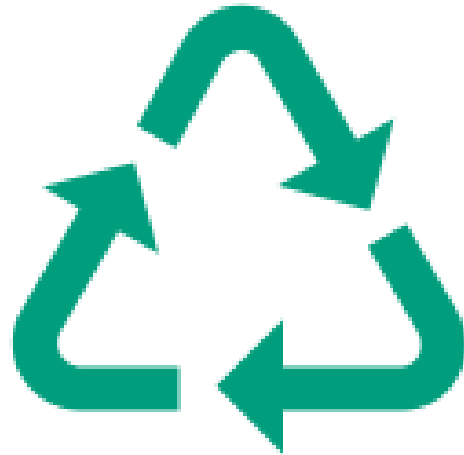
Forskaren kan närma sig ämnet genom att fundera över hur väl en annan forskare skulle kunna replikera det utförda arbetet: Går det att hitta alla data och metoder, är de ömsesidigt förståeliga och användbara?

Forskningsfinansiärer och vetenskapliga förläggare kan förutsätta att FAIR-principerna efterlevs och ifrågavarande data hanteras.

Syftet med anvisningarna i detta dokument är att stödja forskaren i planeringen av sitt arbete, så att det blir så enkelt som möjligt att efterleva FAIR-principerna. I slutet av varje kapitel finns forskarens checklista gällande kapitlets ämnen.



Datahantering som en del av forskningen



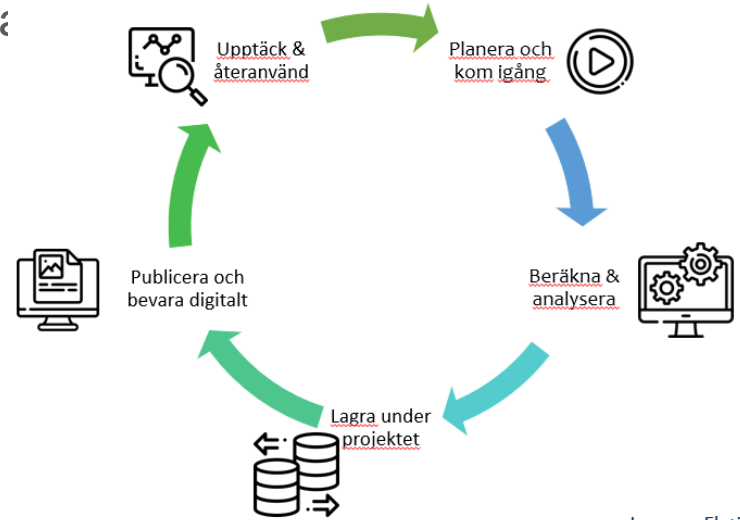
Gör hela forskningen FAIR



FAIR-principerna får inte vara separata från forskningsprocessen. De ska efterlevas under hela forskningen och tillämpade kan de användas i forskningens alla produkter.

Reproducerbarhet kan vara ett kriterium, enligt vilket man beslutar vad som ska lagras. Då måste man absolut även komma ihåg dokumentering och lagring.

I FAIR by design-tänket strävar man efter att planera hela forskningsprocessen så att till exempel metadata ackumuleras automatiskt och de nödvändiga identifierarna är enkla att skapa.



Datahanteringen börjar redan i forskningens planeringsskede (DMP – Data Management Plan)



Genomförandet av FAIR-principerna görs under hela forskningen. Planera med andra ord:

- hela livscykeln för data och observera den i planeringen av forskningen (t.ex. versioner som ska lagras, långtidslagras och förstöras).
- tekniska data gällande ifrågavarande data samt frågor gällande rättigheter.
- de olika formerna av dokumentation (t.ex. källkod, koduppsättning, anvisningar) som tillsammans förbättrar transparensen.

Till ansvarsfull forskning hör transparens i processen, det vill säga öppenhet gällande vad man har gjort med data och hur man har kommit fram till slutresultatet. Detta måste beaktas i materialhanteringsplanen, som ska upprättas redan i projektets planeringsskede och hållas uppdaterad under forskningens framskridande.

Om livscykeln för data (1): Mognadsgraden för data



Med rådata avses data i dess ursprungliga form, så som de har fått direkt från källan, innan de hanterades på något som helst sätt.

Med primärdata avses data som samlats in eller skapats och underhållits för ett visst syfte, till exempel en masterkopia av data som upphovsmannen har ansvar för.

Publicerade forskningsdata är en datahelhet som man kan hänvisa till, som kan växa via en kontrollerad kvalitets-, dokumentations- och versioneringsprocess. Som minimum är dess metadata offentligt tillgängliga och begränsningarna för tillgängligheten har definierats maskinläsbart. Publicerade forskningsdata kan bestå av data som importerats från flera primärkällor.

Om livscykeln för data (2): Skapa och förbereda



Uppkomsten av data kan vara förknippad med olika instrument eller insamlingsmetoder, som är viktiga att dokumentera.

Från rådata till data som kan analyseras

- Data som samlas in för forskning är i det första skedet så kallade rådata. Rådata är startpunkten för hanteringen av data och det vore bra att alltid kunna återkomma till dessa data.
- För analyser ska rådata kvalitetsgranskas (dvs. **valideras**), **beskrivas** och ibland berikas och klassificeras de innan de kan analyseras.

I beredningen av data används ofta olika kommandon eller annan kod

- **Källkod** (*program code*) är ett redskap med vilket data produceras. Till exempel finns det skäl att lagra, dokumentera och publicera använda modeller, analys- och datahanteringskoden (t.ex. R-skript), såvida det inte finns grund för att inte publicera.

Om livscykeln för data (3): Förädling och hantering



Koduppsättning innebär angivelse av variabler, entiteter eller kategorier i data. Dessa materialspecifika koduppsättningar måste dokumenteras och öppnas tillsammans med ifrågavarande data. Sträva efter att tillämpa befintliga standarder och/eller koduppsättningar för att förbättra begripligheten och reproducerbarheten.

Arbetsflöde är en serie funktioner, med hjälp av vilka man kan standardisera de olika skedena av arbetet och säkerställa följdriktiga resultat. Arbetsflöden strukturerar processer och förbättrar effektiviteten. Arbetsflödet möjliggör lagring och publicering av utvecklingshistorik (*lineage*) och proveniens (*provenance*) för data.

När informationsobjekt skapas eller en förändrad version skapas av det, märks det med ett versionsnummer. **Versionshanteringen** för bok över lagrade versioner och möjliggör återställning till dem, vilket är viktigt med tanke på spårbarhet för ifrågavarande data, uppföljning av redigeringar och korrigeringsfel.

Anvisningar gällande livscykeln för data och dokumentering av den



- Spara rådata, såvida lagring av dem är möjligt och ändamålsenligt.
- Dokumentera datakällorna. Använd permanenta identifierare om det är möjligt (t.ex. för rådatauppsättningar, infrastrukturer, sensorer, protokoll).
- Dokumentera datahanteringen och dess principer noggrant, såsom valideringsprocessen, huruvida skrivfel har korrigerats samt konversioner osv.
- Planera versioneringen, långtidslagringen och förstöringen också med tanke på rådata
- Säkerställ att du vid behov kan återgå bakåt i versionerna genom att se till att du skapar en tillräcklig mängd versioner under arbetets gång. Säkerställ att versionerna är entydigt identifierbara.
- Se till att filerna är oskadade genom att räkna kontrollsummor för dem och kom ihåg säkerhetskopieringen. Registrera även dessa åtgärder.
- Fundera över vad som är nödvändigt med tanke på forskningens reproducerbarhet.
- I synnerhet med tanke på data som långtidslagras är noggrann dokumentation viktig, till exempel beträffande apparatur och program som har använts.

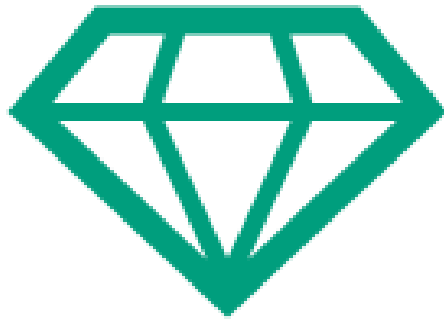
CHECKLISTA: Är din forskning reproducerbar?



1. Styr materialhanteringsplanen (DMP) arbetet under hela livscykeln för data så att hela datahanteringsprocessen är transparent och tillräckligt dokumenterad?
2. Hur har transparens och användningsbegränsningar för data beaktats under hela processen?
3. Utnyttjar man i metadata och själva data gemensam praxis, till exempel standarder och ordlistor?
4. Har forskningsmaterialets livscykel dokumenterats systematiskt och motsvarar beskrivningen verkligheten? Har så många skeden av datahanteringen som möjligt automatiserats och har koden sparats? Finns dokumentationen (den tekniska dokumentationen) av de program och inställningar som har använts tillgänglig?
5. Har data och annan output versionerats?
6. Har data och dokumentationen av dem sparats i ett hänvisningsbart format (permanenta identifierare och metadata)?

Genom att ombesörja reproducerbarheten genomförs FAIR-principerna naturligt som en del av forskningsprocessen och varken data eller dokumentation behöver skapas separat i publiceringsskedet av artikeln eller ifrågavarande data

Utvärdering av datas värde och kvalitet



FAIR-principerna och datavärde



Vidareutnyttjande av data ökar deras värde för vetenskapssamfundet och samhället. Forskaren har ansvar för att för sin egen del definiera värdet på data och se till att det bevaras.

Materialhantering i enlighet med FAIR-principerna möjliggör ansvarsfullt vidareutnyttjande av data och ökar forskningens genomslag.

Ju noggrannare ursprunget för ifrågavarande data dokumenteras, desto enklare är de att använda på nytt. Med anledning av detta måste arbetsflödet, med vilket man får forskningsdata av rådata som kan publiceras, sparas.

Forskaren kan samla in data för att lösa en fråga som hen själv har definierat inom sitt vetenskapliga område, men samma data kan även ha ett värde som en del av en större helhet, till exempel i organisationens databas, som jämförelsedata, som en del av metaanalysen eller inom något annat vetenskapligt område.

Uniciteten hos data och mängden resurser som används för att producera dem ökar vanligen datavärdet. Data har ofta även ett värde ur perspektivet för forskningens transparens och (referent)granskning.

Förändring av värdet på data



Värdet på data kan minska eller öka med tiden: det kan hända att data som har samlats in för lösning av ett aktuellt problem endast kan användas en gång, men de kan även få en ny betydelse till exempel som en del av en tidsserie, som referenspunkt för förändring eller som en del av ett kulturarv. Med tiden kan forskningsmaterialet få ett kulturarvsvärde.

- [Läs mer om kulturarvsvärde och data](#)

Det är svårt att förutse vilka data som är värdefulla i framtiden, vilket är anledningen till att ett noggrant urval av data och säkerställande av begriplighet för data är en väsentlig del av forskning.

- Till exempel har uppföljningsmaterialet gällande islossningen i Aura å i Åbo samlats in från såväl arkivdata som gamla tidningsdata ([hela forskningsartikeln på engelska](#)). Med hjälp av det kan man nu 100–200 år senare analysera effekterna av klimatförändringen.
- Lagringen kan även öka värdet på rådata. Till exempel kan till synes avvikande observationer vara de första tecknen på en begynnande förändring. Därför kan det vara bra att spara rådata, även om de vid insamlingstidpunkten kan tyckas innehålla oförklarliga fel.



Publicerade anvisningar för utvärdering av datavärde och -kvalitet

Det finns många olika sätt och kriterier att definiera värdet på data.

Kvalitetsgranskningen kan till exempel baseras på hur välbevarade data man har, hur täckande, tillförlitliga och enhetliga de är. När alla data inte kan lagras, stöder definiering av värde och kvalitet i synnerhet urvalet av material för långtidslagring.

Härnäst presenterar vi huvudpunkterna från två publicerade anvisningar, som forskare kan utnyttja och tillämpa vid utvärdering av värdet för och kvaliteten på sina data:

- Helsingfors universitets anvisning för definition av datavärde med tanke på långtidförvaring: Krister Talvinen. (2019). Digital Preservation (Fairdata-PAS): Guidelines for UH Evaluators. [Zenodo](#).
- [Statistikcentralens ramverk för kvalitetssäkring av datamaterial \(på finska\)](#).

Även [Riksarkivets anvisningar \(på finska\)](#) innehåller ett kapitel, som behandlar forskningsmaterial och definitionen av värdet i synnerhet ur ett arkivperspektiv.

Synvinklar på bedömningen av förväntat vidareutnyttjande av data



- Är dina data som helhet tillräckligt heltäckande, så att de kan användas på olika sätt i framtiden? Täckning kan avse ett visst fenomen, en viss tidsperiod osv.
- Är dina data tekniskt så bra, att de kan användas på olika sätt i framtiden? Eller måste de hanteras eller förberedas på ett sätt som orsakar orimliga kostnader?
- Kan dina data utöka befintligt material?
- Kan dina data fungera som jämförelsepunkt för något annat material?
- Har dina data analyserats endast delvis?
- Är det rimligt att förutsätta att dina data ännu kan utnyttjas på nya sätt i framtiden?

Synvinklar på bedömning av det framtida värdet på data



- Har dina data i framtiden någon betydelse för utvecklingen av ett visst vetenskapligt område eller ett forskningsämne?
- Kan dina data leda till betydande vetenskapliga rön eller publikationer?
- Är dina data vetenskapligt eller kulturellt unika?
- Kan vidareutnyttjandet av dina data leda till kommersiella tillämpningar, företagssamarbete eller patent?
- Har dina data ett betydande värde för undervisningen, till exempel inom forskarutbildningen?

Synvinklar på bedömning av bevisad betydelse



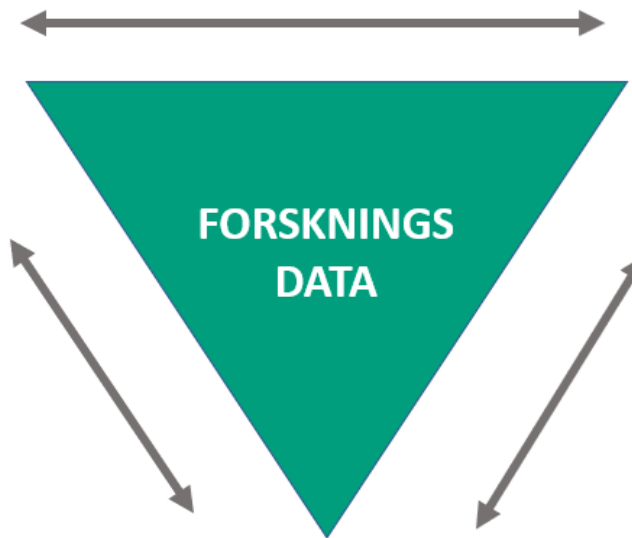
- Har dina data använts för att utarbeta någon särskilt viktig publikation eller ett särskilt viktigt fynd?
- Är dina data avgörande för den nationella eller globala infrastrukturen?
- Har dina data utnyttjats i betydande forskningssamarbete olika organisationer emellan?
- Har produktionen av dina data krävt betydande investeringar och resurser?
- Har till exempel en etisk kommitté bedömt dina data tidigare? Kommitténs bedömningsrapporter kan vara nyttiga i beslutsfattandet, ibland kan det vara av etiska orsaker grundat att lagra data.
 - [Anvisningar för etikprövning inom humanvetenskaperna](#)

Omfattningen av genomslag



Samhälleligt genomslag

- Livskvalitet
- Hälsa
- Miljö
- Offentliga tjänster
- Policyer
- Kreativa funktioner
- Delaktiggörande
- Förståelse



Ekonomiskt genomslag

- Innovationer
- Konkurrensförmåga
- Tillväxt
- Arbetsplatser
- Budgetbesparingar

Akademiskt genomslag

- Teori
- Metod
- Vetenskap
- Teknisk utveckling
- Forskningsutbildning
- Undervisning och utbildning
- Anpassningsbarhet

Data av hög kvalitet beskriver verkligheten



Felfrihet

beskriver hur data motsvarar verkligheten. Genom att granska felfriheten hos data kan man även hitta systematiska förvrängningar i datamaterialet.

Noggrannhet

beskriver hur väl uppgifterna i data motsvarar vad som eftersträvas. Noggrannhet beskriver hur väl data träffar rätt.

(Aktualitet)
ifall relevant

Följdriktighet

berättar om att data är enhetliga och konsistenta. Med följdriktighet kan man även beskriva den ömsesidiga följdriktigheten olika data emellan.

Omfattning

beskriver den eftersträvade tidsmässiga och områdesvisa omfattningen för data samt eftersträvade målenheter och egenskapsuppgifter. Å andra sidan anger omfattningen till vilka delar data innehåller eftersträvade uppgifter.



Data av hög kvalitet har dokumenterats väl

Ursprung

anger att ändringar gjorda i data och informationen i dem kan spåras. Man känner till ursprunget för data.

Begriplighet av metadata

innebär hur täckande metadata beskriver ifrågavarande data och underlättar innehållets begriplighet.

Efterlevnad av rekommendationer

anger att data och dess egenskapsuppgifter efterlever erkända standarder, praxis och författningar, och dessa anges i samband med data.

Data av hög kvalitet kan användas på nytt



Maskinläsbarhet

beskriver om data har strukturerats så att de kan hanteras maskinellt och hanteringen är möjlig att genomföra i olika datasystem.

Användarrättigheter

beskriver hur användarrättigheterna för data har definierats och vad man kan göra med data, det vill säga i vilket användningssyfte datamaterialet kan utnyttjas.

Exakthet

innebär att data är tillgängliga vid den angivna tidpunkten och kompletterade tillräckligt ofta.

Så öppet som möjligt, så begränsat som nödvändigt

Hur värdefulla är dina data?



Alla forskningsdata är unika i sina respektive kontexter. Beskrivningen av dina data och forskningen som bedrivits utgående från och öppnande av data ökar dess värde för vetenskapen. Alla data kan dock inte lagras i all evighet och någon måste ha ett långsiktigt ansvar för de data som ska lagras.

Vid värderingen kan forskaren och organisationen som ansvarar för forskningsdata beakta:

- **Möjligheterna att utnyttja data på nytt,**
 - betydelsen för vetenskapen nu och inom den närmaste framtiden eller på lång sikt,
 - användningen inom det egna vetenskapliga området vs. vidare användning som kompletterande data/data som möjliggör jämförelse,
 - mängden icke-analyserade uppgifter som data innehåller,
- **kvaliteten på data och dokumentation,**
 - hur de stöder forskningens reproducerbarhet eller kvalitetssäkring,
- **det monetära värdet på ifrågavarande data,**
 - satsade medel och arbetsresurser, kommersiellt värde,
 - forskningsfinansiärens intresse,
- **betydelsen av data genom forskningens genomslag i samhället,**
 - till exempel dess betydelse för lösningen av mänsklighetens besvärliga problem,
- **den historiska och kulturella betydelsen för data eller deras vetenskapliga eller kulturella unicitet**
 - som en del av långa tidsserier för såväl forskare som samhället,

Att dela och lagra data



Om skeden i datahanteringen



Datadelning (*data sharing*) kan göras redan under forskningsprojektets gång. Då kan datasetet vara aktivt. Datahantering och kuratering sker efter överenskommelse utgående från planeringen av datahanteringen, så att alla inblandade känner till rättigheter och ansvar med anknytning till ifrågavarande data, hur versionering görs och hur data överlag får och kan användas. ***Delning kan till exempel göras från webbaserad datalagring, en datapool eller med hjälp av e-post, eller i en tjänst avsedd för konfidentiell datadelning.***

När data lagras (*data storing*), dokumenteras de noggrant och metadata som är väsentliga med tanke på användning och hantering ansluts till dem. ***Man kan använda en lämplig tjänst för datalagring*** (*data repository eller data service*).

Publicering av data (*publishing data*) gör data till ett hänvisningsbart material som måste ha en permanent identifierare och hänvisningsanvisningar. Publicerade data har dokumenterats noggrant och åtminstone metadata som gör det möjligt att hitta dem har publicerats. I metadata anges hur man får tillgång till materialet och vilken kvalitet det har för att det ska kunna användas i reproduktionen av forskningen eller inom ny forskning. ***En lämplig tjänst används för publiceringen av data*** (*data repository eller data service*).

Långtidslagring av data (*digital preservation*) är bevarande av materialets begriplighet och ofelbarhet på lång sikt, över årtionden och till med århundraden. I sådant fall beaktas till exempel förändringen av olika tekniker, medier och format. Även konfidentiellt material kan lagras långsiktigt i certifierade tjänster.

Obs! Åtkomsten till materialet kan i alla skeden vara begränsat. Man måste ombesörja ett tillräckligt dataskydd och tillräcklig datasäkerhet!

Sammanställning av dataset med tanke på fortsatt användning



Dataset är en helhet som sammanställts av ifrågavarande data. När data publiceras måste man ta ställning förutom till deras olika förädlingsgrader (level) och dokumentation, även till exempel till

- granularitet och resolution (kornighet) och nödvändig hänvisningsnoggrannhet (identifierare)
- filformat, hur data har fördelats i filer och hur de har dokumenterats (struktur, identifierare)
- hur det publicerade materialet eventuellt versioneras eller ackumuleras i framtiden
- angivelse av tidsserier och olika variabler i metadata och att göra dem sökbara mellan olika dataset.

När du beslutar om lösningar, observera FAIR-principerna – tänk på materialets hänvisningsbarhet, hittbarhet, interoperabilitet och fortsatta användning.

- Det är ofta enklare att dela in data i delar än att sammanföra dem på nytt från olika filer
- Stora filer kan å andra sidan vara tunga att överföra och hantera
- Försök tänka på framtida användare och nya användningssätt
- Kom ihåg att man av dina data även kan ge ut sekundära publikationer (som inte är nya versioner av den primära publikationen)
- Följ god praxis inom ditt område

Val av lagrings- och publiceringstjänster



- FAIR-principerna förutsätter att forskningsresultaten har en permanent identifierare och att deras metadata är sökbara.
 - I praktiken förutsätter detta utnyttjande av gemensamma tjänster avsedda för publicering av forskningsdata
- Definiera minimikraven som publiceringstjänsten ska uppfylla.
 - Dataskydd, användbarhet, tillgänglighet, verktygsstöd, datamodeller, permanenta identifierare
- Utnyttja, om möjligt, tjänster som är specifika för olika vetenskapsområden.
 - Beakta vem som kan utnyttja dina data vid ett senare tillfälle – var kan data enklare hittas?
 - maskinläsbarhet, sökbarhet för metadata
 - Vilka är de tillgängliga resurserna?
 - Kontrollera olika format, innehållets interoperabilitet och innehållets struktur
 - Kom ihåg att man från samma material kan skapa olika kopior för olika ändamål
- Definiera användarrättigheter för ifrågavarande data.
 - Avtalsmässiga kriterier
 - Finansiärens och organisationens riktlinjer

Identifiering av tillförlitliga tjänster



Certifikat som CoreTrustSeal hjälper till att bedöma tjänsternas tillförlitlighet.



Lämpliga tjänster kan du söka till exempel i tjänsten re3data.org. Välj en områdesspecifik tjänst, om möjligt.

Ta hjälp av din egen organisations anvisningar och stödtjänster.

I Finland erbjuds tillförlitliga tjänster som lämpar sig för forskare förutom av organisationer som bedriver forskning, även av kulturarvsorganisationer och CSC.

Överföring av rättigheter eller delning av dem med organisationen eller tjänsten eliminerar inte upphovsrätten, men möjliggör långsiktig hantering av materialet.

Publicering av statistiska data



Statiska data är en oföränderlig helhet, till exempel forskningens bakgrundsmaterial eller data som den publicerade artikeln baseras på. De kan även vara ett utdrag från andra datakällor. Till artikeln kopplas en hänvisning och information om tillgängligheten för ifrågavarande data (*Data & Code availability statement*). **Datasetet och versioner av det måste kunna identifieras entydigt.**

Som forskningsevidens räcker det ofta med en väl dokumenterad fil eller en samling av filer eller material, såvida det i samband med forskningen uppstår nytt material. Observera i sådana fall även program och koder samt delning av dessa.

Välj ett transparent dokumenterat, allmänt och öppet filformat, eftersom detta skapar förutsättningar för fortsatt användning av data.

God dokumentation förutsätter oftast både beskrivande metadata som skapats av en människa och beskriver data, och automatiskt genererade metadata gällande de tekniska egenskaperna för data.

Ofta krävs även annan dokumentation och hänvisningar till olika källor, såsom standarder, ordlistor, metoder, koduppsättningar osv.

Publicering av variabeldata



Det finns olika typer av variabeldata. I **kumulativa data** ackumuleras materialet, och redan insamlade data ändras inte. I **dynamiska data** är retroaktiva ändringar också möjliga.

I kumulativa data kan det nya materialet läggas till antingen som nya filer eller som uppdateringar av gamla filer. Kumulativa data kan publiceras i vissa tjänster.

Databaser som vetenskapliga material kräver särskild dokumentation, planering och underhåll. För beskrivningarna gäller dock samma rekommendationer som för publicering av statiska data.

För att förbli funktionella kräver databaser underhåll, som på lång sikt kan bli dyrt. Därför måste deras livscykel och underhållsarrangemang efter att finansieringen av forskningsprojekten upphör planeras på förhand.

Interoperabilitet och utnyttjande av standarder som tillämpas inom långtidslagring förenklar vid behov överföringen av data till nya miljöer.

Med anledning av forskningens reproducerbarhet finns det orsak att beakta behovet av exakta och entydiga hänvisningar. **En permanent identifierare ska alltid styra användaren entydigt till rätta data.**

Exempel på hänvisning till material som ändras



1. Hänvisa till en delgrupp inom specifika data genom att nämna källan, exakta avgränsningar samt den permanenta identifieraren
 - a. Fiktivt exempel: Data Request T.Jansen; SAHFOS; Work published 2014 via SAHFOS; Area Def: 54–65 ° N, 0–45 ° W. Temporal Def: 1980–2012 (april–augusti) Taxonomic Def: All zooplankton; (dataset). <https://doi.org/10.7487/2014.15.1.1>
2. Hänvisar till datakopier vid en viss tidpunkt.
 - a. Fiktivt exempel: König-Langlo, G., och Sieger, R. (2010). BSRN snapshot 2010-01 as ISO image file (3,75 GB) [Data set]. PANGAEA - Data Publisher for Earth & Environmental Science. (dataset). <https://doi.org/10.1594/pangaea.833424>
3. Hänvisa till data som uppdateras kontinuerligt genom att ange en exakt tidpunkt för hänvisningen. *OBS! Denna metod möjliggör inte nödvändigtvis reproducerbarhet, men är ibland den enda möjliga metoden.*
 - a. Fiktivt exempel: Doe, J. and R. Roe. 2001. The FOO Data Set. Version 2.3. The FOO Data Center. (dataset). <https://doi.org/10.xxxx/notfoo.547983>. Accessed 1 May 2011.
4. Hänvisa till en tidsstämplad databasförfrågan i den versionerade databasen.
 - Fiktivt exempel: R. Roe. 2017. "The Moo Data Query" created at 2017-07-21 10:25:30 PID <https://doi.org/10.xxxx/notmoo.857988> Subset of Moo Database (dataset) PID <https://doi.org/10.xxxx/bigmoo.360873>

Material med begränsad användning



Användningen och lagringen av material kan vara förknippat med etiska, juridiska eller avtalsmässiga begränsningar.

Avtalsmässiga åtgärder som krävs för användningen samt åtgärder med anknytning till processerna och dataskyddet beror på de materialspecifika begränsningarna.

Särskild noggrannhet krävs av material som innehåller personuppgifter eller material som ägs av andra aktörer som är förknippade med avtal och villkor.

Kontakta datastödet inom din organisation om du är osäker på frågor med anknytning till användarrättigheter. Även din organisations datapolicy kan innehålla anvisningar, som kan hjälpa dig att utföra ditt arbete med lugn i sinnet.

För att kunna hänvisa till ett användarbegränsat material, måste det finnas tillräckligt med beskrivande metadata om det, en identifierare och en målsida. I metadata anges vad begränsningarna baseras på och hur man får åtkomst till materialet.

- Till exempel CLARIN ACA-licensen, i vilken materialets användning är begränsat till forskningsbruk (innehåller personuppgifter), till exempel [Språkbanken](#)

Långsiktigt bevarande av data i Finland



Långsiktigt bevarande av forskningsdata (*digital preservation*) innebär att materialets begriplighet bevaras i princip permanent. Byte av generationer av apparatur, program och forskare och förändringen av forskningsparadigm beaktas genom att vid behov redigera data.

Data som bedömts vara värdefulla måste bevaras så länge som möjligt. Bestämningen av värdet på data har många dimensioner, som behandlades i föregående kapitel.

- Datas värde, frågor med anknytning till juridik och ägarskap ska beaktas redan i planeringen av forskningen, så att processen för överföringen till långsiktigt bevarande blir så smärtfri som möjligt.

I Finland erbjuds digital långsiktigt bevarande för organisationer till exempel av CSC:s PAS-tjänst. [Fairdata PAS-tjänsten](#) har utvecklats för långsiktigt bevarande av forskningsdata. Överföringen av materialet till tjänsten förutsätter alltid ett avtal mellan organisationen och undervisnings- och kulturministeriet, det vill säga att vägen till att utnyttja tjänsten för forskarens del alltid börjar med förhandlingar med den egna organisationen.

CHECKLISTA: Har du koll på dina datafrågor?



1. Planera noggrant: Vilka är de värdefulla resultaten som lämpar sig för lagring som är viktiga med tanke på reproducerbarhet och ansvarsfullhet?
2. Är frågorna gällande avtal, underhåll och användarrättigheter under kontroll också med tanke på framtiden?
3. Hur länge lönar det sig att lagra olika resultat och när ska materialet eventuellt förstöras, till exempel onödiga versioner? Kraven från finansiären och organisationer i bakgrunden påverkar.
4. Är de publicerings- och lagringstjänster som du har valt tillförlitliga, lämpliga för ditt material, stöder de FAIR-principerna och är de specifika för olika vetenskapliga områden?
5. Har du tänkt på framtida användare när du sammanställer och dokumenterar ditt material?
6. Har du tagit hänsyn till att dina data eventuellt ska långtidslagras för att möjliggöra användning på nytt i enlighet med deras kulturarvsvärde eller finansiärernas krav?
7. Har du berättat för din organisation om data som är värda att bevaras långsiktigt?
8. Har du informerat forskningsdeltagarna hur materialet delas, lagras och publiceras?
9. Har du kommit ihåg att hänvisa till data i publikationer som du har använt dem för?

Författare

Anneli Lehtisalo

Ari Asmi

Hanna Koivula

Heidi Troberg

Jessica Parland-von Essen

Juha Hakala

Katja Laine

Maria Söderholm

Marjut Vuorinen

Mika Virtanen

Nina-Mari Salminen

Pekka Nygren

Saila Huuskonen

Sonja Sipponen

Tanja Lindholm

Tarja Mäkinen

Timo Taskinen

Tomi Rosti

Tuomas Alaterä

Tuula Pääkkönen



Öppen
vetenskap

