

# Forskarens handbok för öppen forskning

Handboken baserar sig på *Avoimen tieteen ja tutkimuksen käsikirja* som producerats inom det finska undervisnings- och kulturministeriets Initiativ för öppen forskning och vetenskap. Den ursprungliga handboken, liksom denna översatta och reviderade version, publiceras med Creative Commons [CC BY 4.0](#) licens. Översättning av Jessica Parland-von Essen. Den finska förlagan finns att läsa på adressen <http://avointiede.fi/www-kasikirja>.

Version 1.0 / 9.2.2016

[digihist.se/handbok](http://digihist.se/handbok)

## Forskarens handbok för öppen forskning

### Inledning

Vad är öppen forskning?

Varför öppenhet är en god idé

### Informationsförvaltning inom forskning

Forskarens checklista för informationsförvaltning

### Forskningsdata

Forskarens checklista för forskningsdata

### Datahantering

#### Att öppna data

Teknisk öppenhet

Öppenhet gällande användning

Embargo

### Fördelar med att dela forskningsdata

Vad innebär öppen tillgång i praktiken?

Nyttan för forskaren

Nyttan för forskarsamfundet

Nyttan för finansiärerna

Nyttan för samhället

### Publiceringsprocessen för forskningsdata

1. Välj ut och preparera data

2. Ge en så öppen licens som möjligt

3. Spara materialet hos en trovärdig tjänst

4. Dokumentera dina data och se till att de får en beständig identifikator

5. Att göra en applikation

### Datasäkerhet

Risk

Åtgärder

### Metadata för forskningsdata

Forskarens checklista för metadata

### Identifikatorer och referenser

Forskarens checklista för identifikatorer och referenser

### Den öppna forskningsprocessen

Forskarens checklista för en öppen forskningsprocess

### Att publicera sina resultat

Forskarens perspektiv på Open Access

Forskarens checklista för publicering

### Sociala medier och webbpublicering

Forskarens checklista för sociala medier

## Inledning

För att öppen forskning och vetenskap ska kunna förverkligas krävs kunskap. Forskare och forskningsorganisationer måste försäkra ett tillräckligt kunnande i informationsförvaltning, för att man ska kunna säkra att forskningsmaterial, -metoder och publikationer finns öppet tillgängliga för alla. Inom det finska initiativet för öppen forskning och vetenskap vill man stöda forskarna och organisationerna genom att skapa direktiv, incitament och tjänster för ändamålet. En grundläggande princip är att användningsvillkoren för forskningsmaterial och publikationer definieras tydligt och inte begränsas utan orsak.

Öppenheten kan för forskare te sig som både ett hot och en möjlighet. Omvärldsförändringarna erbjuder forskare många möjligheter att främja det egna arbetet och dess genomslag, men utvecklingen kräver ofta att man sätter sig in i nya saker, såsom informationsförvaltning och frågor om upphovsrätt. Avsikten med denna handbok är att fungera som ett stöd för forskare och forskningsadministrationen.

Det finska initiativet för öppen vetenskap och forskning, som startats av undervisnings- och kulturministeriet, rekommenderar att alla resultat publiceras med öppen licens eller som öppen källkod. Den allmänna principen är att återanvändning inte ska begränsas och alla användningsvillkor presenteras klart. För öppna material rekommenderas den standardiserade, maskinläsbara CC BY 4.0-licensen. För metadata om material rekommenderas CC0, medan den så kallade MIT-licensen lämpar sig för programkod, eftersom den också möjliggör kommersiell användning, något som eftersträvas både inom EU och nationellt i Finland. Forskare, forskningsgrupper och forskningsorganisationer uppmanas att utnyttja existerande arkiv och tjänster, som till exempel PubMed Central, arXiv, OpenAIRE, Zenodo och i Finland IDA, Etsin, Aila och Språkbanken och högskolornas och forskningsinstitutens egna publikationsarkiv och informationssystem för forskningsinformation (CRIS).

## Vad är öppen forskning?

Öppenhet är en central princip inom forskning och vetenskap. Genom öppenhet kan vi tillsammans skapa nya möjligheter för deltagande i forskningsarbetet både inom forskarsamfundet, och för beslutsfattare och medborgare. Öppen forskning har blivit ett internationellt sett viktigt sätt att främja forskningen och dess inflytande i samhället. Ökningen beror på förbättrade möjligheter för spridning, verifiering, granskning och replikering. En förutsättning är att de publikationer, forskningsdata, metoder och kunnande och de tillhörande tjänsterna är allmänt tillgängliga. Öppenheten måste förverkligas både vad gäller skaffandet och skapandet av data, och angående hur de använts och behandlats. Det intresse de väcker berättar för sin del om deras relevans. När forskningens resultat och produkter finns allmänt tillgängliga ökar vetenskapens och forskningens betydelse och inverkan.

***Öppen forskning och vetenskap är inte bara en samling åtgärder och rekommendationer, utan också en förändringsprocess inom hela sättet att bedriva forskning.***

Vetenskapen tillhör alla. Genom de nya tillvägagångssätten kan vi demokratisera vetenskapen och öka dess inflytande. Då forskningsprocessen digitaliseras och öppnas upp, ger det nya möjligheter för forskarna och alla som är intresserade att samarbeta och kommunicera. Öppenheten ger möjligheter till förnyelse. Det är också viktigt att ge unga och oetablerade forskare som verkar utanför etablissemangen möjlighet att vara delaktiga på lika villkor. Genom öppenhet kan alla delta utgående från sina egna förutsättningar att studera, forska, kritisera, övertygas och främja vetenskapen. Det här ger också forskningen större legitimitet och bidrar till att främja företagsamhet, men lika viktigt är att det sprider kompetenser som kritiskt tänkande och informationskunnighet i samhället.

Öppenhet har alltid varit en målsättning inom vetenskapen, eftersom man på det sättet har kunnat garantera en högre kvalitet och motarbeta förfalskningar. Forskningens självkorrigering karaktär kräver transparens och en möjlighet att replikera forskning. Då material, metoder och resultat är redovisade och tillgängliga kan de kontrolleras och verifieras och kvaliteten blir bättre. Öppen forskning gör att vetenskapen kan utvecklas snabbare, felaktigheter korrigeras och man kan undvika onödigt dubbelarbete. Med öppen forskning och vetenskap eftersträvas beständighet, användbarhet, tillgänglighet och legitimitet. Öppenheten är också en självskriven del av forskningsetiken. Alla material måste vara möjliga att hitta och utnyttja också på längre sikt.

Goda tillvägagångssätt inom öppen forskning och vetenskap härstammar från generella forskningsetiska direktiv, god informationsförvaltning och mer specifika rekommendationer. Det finska forskningsetiska rådet (TENK) har tillsammans med forskarsamfundet förnyat rekommendationerna<sup>1</sup> år 2012, samt gett förhållningsdirektiv i fall av misstankar av brott mot dessa direktiv.

---

<sup>1</sup> Forskningsetiska delegationen: God vetenskaplig praxis, <http://www.tenk.fi/sv/god-vetenskaplig-praxis-anvisningar/god-vetenskaplig-praxis>.

Öppen forskning och vetenskap är ett kontinuum, som spänner över hela forskningsprocessen. Information och kunnskap finns utspridd hos olika aktörer i systemet, bland olika nätverk och samfund. I detta läge kan vi öka öppenheten internt och tillsammans utveckla informationshanteringsprocesser och arbetsmetoder, och även utåt genom att öppna upp hela processen för övriga forskare och det övriga samhället genom att ge dem möjligheter att utnyttja material, metoder och resultat.

I ett öppet och välfungerande arbetssätt ingår en strävan efter hög kvalitet: ärlighet och noggrannhet i alla skeden. Det samma gäller förstås forskningsmaterialen som en del av denna process, men för deras del måste man också se till frågor om trovärdighet, förtroende och användbarhet. Att ombesörja datas kvalitet är också viktigt. Det innebär att information inte förändras utan orsak eller dokumentation, att den inte korrumpas, skadas eller försvinner. Att man kan lita på forskningsdata är ett självändamål, som man inte ska pruta på utan rimliga orsaker. De åtgärder man vidtar måste ändå alltid stå i proportion till hur viktiga och värdefulla data är. Ju viktigare information, desto mer måste man satsa på dess kvalitet och bevarande. Användningsändamålet är avgörande vid bedömningen och man ska alltid fästa uppmärksamhet vid god informationsförvaltning och dokumentation.

## Varför öppenhet är en god idé

Då informationen har digital form öppnas nya möjligheter för vetenskaplig forskning. Digital information kan enkelt och snabbt delas och kommenteras. Att dela forskningsdata och -resultat öppet är ett sätt att främja öppenhet. Men öppen forskning innebär också att man delar andra skeden av processen. Ett känt exempel är Human Genome Project, där ett nytt och öppet arbetssätt skapade en tidigare oskådad effektiv och snabb metod att kartlägga människans arvsanlag. Också inom paleogenetik och astronomi finns kända exempel på öppna forskningsmetoder.

Öppenheten har visat sig vara ett bra sätt att främja transparens och verifierbarhet inom forskning. Det har som helhet ökat validiteten och legitimiteten i resultaten. Öppenheten ger också både forskare och forskningsorganisationer möjligheter till bättre synlighet, men kräver en del nya kompetenser inom informationsförvaltning och hantering av forskningsprocessen.

Öppenheten gör att forskningens resultat är lättare att hitta och få tillgång till, både inom forskningssamfundet och samhället i stort. Det skapar möjligheter, men kräver också planering och eftertanke av forskarna. Öppenheten kan nämligen begränsas av etiska eller juridiska skäl.

En stor del av forskningen är i dag rätt svår att hitta och få tag i för allmänheten. Målsättningen med denna handbok är därför att stöda en bättre informationsförvaltning. Den är skriven som ett öppet samarbete av många sakkunniga i Finland inom Initiativet för öppen forskning och vetenskap.

Både nationellt och internationellt sett finns det redan många forskningsorganisationer och -finansiärer som verkar för öppen forskning. EU-kommissionen har publicerat ett dokument, "Recommendation on Access and Preservation of Scientific Information", där vissa målsättningar och principer slås fast. OECD publicerade år 2007 rapporten "Principles and Guidelines for Access to Research Data from Public Funding", som lett till bättre tillgång till forskningsdata. OECD Council gav följande år en rekommendation, som hade stor inverkan på EU:s PSI-direktiv för att öppna den offentliga sektorns dataresurser. Som en följd av detta står många parter, såsom finansiärer, högskolor och universitet, bibliotek, dataarkiv, akademiska sällskap och vetenskapsförläggare, inför stora möjligheter och utmaningar på den öppna forskningens område.

Då utvecklingen framskrider kommer mängden av tillgängliga forskningsmaterial och -publikationer ständigt att öka. Som drivkrafter fungerar finansiärernas krav på öppenhet, växande mängder av Open Access-publikationer, som *PLoS One* som blivit världens största vetenskapliga tidskrift, och de allt fler och växande mängderna data- och publikationsarkiv. European Research Council kräver öppen publicering av forskning de finansierar både gällande artiklar och monografier och rekommenderar det för forskningsdata. Inom pilotprojektet inom Horizon2020 gäller öppenhetskravet även data. Enligt en utredning av Knowledge Exchange-nätverket uppmuntras öppenhet allt mera och flera stora europeiska forskningsfinansiärer har det redan som ett krav. Också forskningsorganisationer i till exempel Tyskland har börjat skapa nya kutymer.

Europakommissionen har finansierat två stora projekt för öppen vetenskap: RECODE och FOSTER. Det föregående har fokuserat på frågor om forskningsdata och har gett rekommendationer och direktiv gällande arkivering och dissemination. FOSTER har för sin del fokuserat på att ta fram material om öppen forskning och hantering av forskningsdata. Open Access-publicering främjas också av SPARC Europe, som är ett nätverk av forskningsorganisationer och en systerorganisation till det nordamerikanska Scholarly Publishing and Academic Resources Coalition, som arbetar med samma frågor. Även den europeiska sammanslutningen av forskningsuniversitet League of European Research Universities (LERU) har producerat direktiv för hantering av forskningsdata och Open Access-publicering.

Det allmänna intresset för öppen forskning har ökat och betydande vetenskapliga tidskriftsförläggare såsom *Nature* och *Science* har börjat publicera om ämnet. Man försöker också utveckla koncept och praktiker för hur forskningsdata och -metoder ska kunna publiceras i samband med traditionell utgivning. Olika projekt, som DataCite, publikationsarkivet arXiv, EU-projektens OpenAIRE med datarkivet Zenodo, lika väl som växande dataarkiv som PANGAEA och Dryad, skvallrar om den öppna forskningens frammarsch i världen. Också många stora forskningsinfrastrukturer (t ex ESFRI-projekten) delar sina resultat genom öppna tjänster. Att dela forskningsdata är inte kutym inom alla forskningsområden beroende på olika typer av utmaningar, men intresset är klart växande.

Det finns flera olika typer av utmaningar då det gäller att förverkliga öppen forskning och man försöker lösa dem i olika nationella och internationella nätverk och samarbeten. Det internationella Research Data Alliance-nätverket (RDA) har som motto "Research Data Sharing without Barriers" och dess arbetsgrupper sysslar bland annat med att utarbeta modeller för att beskriva forskningsdata och skapa förutsättningar för återanvändning. Open Knowledge Foundation har skapat internationella nätverk, där också finska forskare är aktiva. Den finska facebookgruppen Open Science Finland har över 300 medlemmar. Open Knowledge Foundation fokuserar ofta på att öka kunskapen bland sakkunniga, i detta fall särskilt forskare och beslutsfattare. År 2015 var föreningens tema inom vetenskapen öppenhet inom den högre utbildningen och utbildning inom öppen forskning.

## Informationsförvaltning inom forskning

I forskningsarbetet ingår mycket hantering av information. Ofta tänker man på informationssökning eller referenshantering, som fortfarande också är de saker som forskaren ofta främst förknippar med bibliotek. Men dagens forskare kämpar med mycket annan information, som dessutom allt oftare är digital: anteckningar, korrespondens, forskningsdata, bilder, ljudinspelningar, datorapplikationer, artikelversioner, rapporteringssystem, publikationssystem och -arkiv, kodböcker, loggar, planer, ansökningar, budgetar och rapporter för att nämna en del. Ofta delar man dessutom material och dokument med kolleger, vilket kan leda till att det finns flera olika parallella versioner i omlopp.

Att planera hur man ska hantera all denna information är en god idé. Bibliotekets personal är utbildad i informationsförvaltning och det är bra att ha dem med som stöd både för den enskilda forskaren och för organisationen som helhet. Det lönar sig till exempel att utreda hurdana verktyg och plattformar det finns för samarbete och projekthantering, hantering av data och dokument och hurdana tjänster det finns för att publicera olika produkter som uppstår, redan då man planerar sitt forskningsprojekt för att kunna beakta de kostnader som uppstår. Man måste också utreda hurdana direktiv forskningsorganisationen och finansiären har, för att inte tala om vad lag och etik ställer för ramvillkor. Allt detta gör man klokt i att skriva ner i en plan för informationsförvaltning. Ibland kan det hända att man måste tänka om under arbetets gång. Planen måste då revideras, så att det kan fungera som dokumentation och handledning för alla som har med forskningen att göra.

En central del av informationsförvaltningen och dess planering är hanteringen av forskningens källmaterial, som allt oftare inkluderar också data. Digitala material är känsliga och misslyckas hanteringen av data, kan hela forskningens trovärdighet bli lidande. En välgjord planering är också ofta en förutsättning för att data ska kunna publiceras. Därför kräver också allt fler finansiärer en plan för hanteringen av forskningsmaterialet som en del av projektplanen. Avsikten är att uppmuntra forskarna till att göra god planering, så att man i slutändan har användbar data av god kvalitet. Det Samhällsvetenskapliga dataarkivet i Finland har producerat en bra guide i ämnet som finns tillgänglig på webben.<sup>2</sup> Guiden kan hjälpa också forskare inom andra områden att göra upp sina planer. Förutom att planera för data, lönar det sig också att planera redan i starten för hur och var man vill publicera sina övriga produkter, t ex eventuell kod och artiklar. Man måste då beakta vilka lagar och avtal som gäller och hur man kan genomföra sitt projekt så att man kan publicera så mycket och så öppet som möjligt. Det lönar sig också att med öppet sinne fundera på när man publicerar olika saker: det kan kanske vara en idé att publicera till exempel rådata redan i ett tidigt skede? Sådan öppenhet kan ge överraskande positiva konsekvenser i form av samarbetserbjudanden eller motsvarande.

För att stöda informationsförvaltningen inom forskningen har man utvecklat många olika typer av tjänster och system (forskningsinfrastruktur). Målsättningen är en palett som erbjuder beständiga lösningar för att bevara, söka och tillgängliggöra data, metoder och publikationer. Sådana tjänster är bland andra PubMed Central, arXiv, OpenAire+, Zenodo och i Finland IDA, Etsin, Aila, AVAA och Språkbanken, publikationsarkiv och CRIS-system samt det kommande systemet för långsiktigt digitalt bevarande. Hittbarhet och användbarhet är särskilt viktiga för öppen forskning och det är därför viktigt att alla tjänster utvecklas i samförstånd och koordinerat.

---

<sup>2</sup> Guiden finns tillgänglig även på engelska <http://www.fsd.uta.fi/aineistonhallinta/en/>.



## Forskarens checklista för informationsförvaltning

- Planera forskningens informationsförvaltning noga redan då du skriver forsknings- och projektplanen
  - Vilka material är relevanta?
  - Kom ihåg både existerande material och sådant som uppstår under arbetets gång
  - Utred frågor om rättigheter och ägande för dessa material. Hurdana avtal kan behövas?
  - Hurdan dokumentation kommer att behövas om materialen, deras hantering och forskningsmetoderna?
  - Hur kan du säkra att din undersökning kan valideras?
- Utred finansiärernas, arbetsgivarens, datakällornas och eventuella informanternas eller subjekts rättigheter, krav och rekommendationer. Sök reda på modellavtal och be om hjälp av forskartjänster, bibliotek, dataarkiv mm.
- Välj så öppna licenser som möjligt för dina produkter. På så sätt har du också själv tillgång till dem i framtiden.
- För programkod är MIT-licensen ett bra alternativ, för databaser CC0 och för publikationer CC BY 4.0
- Ta reda på var data inom ditt eget område finns och utred om du kan utnyttja dataarkiv och kopiera, dela och flytta materialen
  - Ta reda på vilka direktiv och standarder som används för liknande material
  - I vilket format och hur arkiveras och publiceras de?
  - Hur ska man beskriva och dokumentera dem så att andra kan använda dem?
  - Hur är det med metoder?
- Ta reda på var resultat publiceras bäst och fatta preliminärt beslut om kanaler
- Planera och beskriv hur du säkrar kvaliteten på data. Beskriv processerna, granskas data separat? Publiceras en dataartikel? Ta reda på kutymen om du inte känner den. Saker som påverkar kvaliteten av forskningsdata är bland annat:
  - Felfrihet
  - Standardenlighet
  - Koherens
  - Entydighet
  - Begriplighet
  - Aktualitet
- Beräkna och beakta även kostnader för att öppna material och publikationer
- Fundera på vad som händer med dina data när projektet är slut. Vilka material kan vara värda att arkivera och publicera? Hur kan du garantera att resultaten går att validera? Kriterier för arkivering är bland annat:
  - Materialet är unikt och kan inte reproduceras
  - Historisk eller vetenskaplig betydelse (kompletterar t ex existerande samlingar)
  - Stor sannolikhet för återanvändning
  - Materialets ekonomiska värde

## Forskningsdata

När man publicerar forskningens resultat, måste också de material man använt vara i skick. Med material avses här de resurser som forskaren använder eller producerar under forskningsprocessen. De är nödvändiga för att man ska kunna validera forskningen och de kan också ge forskaren merit om de publiceras. Det är de omgivande forskningsorganisationerna och beslutsfattarna som måste tillgodose behovet av tillförlitliga tjänster och resurser för ändamålet. Planering, dokumentation och delandet av materialen är på forskarens ansvar. Det ska förstås göras så att man följer etik, lagstiftning och existerande avtal.

För att forskningsdata ska kunna användas måste det finnas information åtminstone om dess ursprung. För digitala material betyder det beskrivande och teknisk information om vad som ingår, alltså metadata. Det behövs information om hur filen eller filerna är strukturerade, när de har tillkommit och hur de eventuellt behandlats. Denna information kan sparas till exempel i dataarkivets metadatakatalog, i kodböcker eller annan dokumentation. Tillsammans med datafilerna utgör denna information då forskningsdata.

Det är god vetenskaplig praxis att publicera också forskningsdata, eftersom det gör att forskning kan valideras lättare. Det är också att rekommendera att man publicerar också outnyttjade data. Detta är en detalj som också finansiärerna med fördel kan fästa uppmärksamhet vid då man evaluerar informationsförvaltningsplanerna. Publicering av forskningsdata är också en sak, som lönar sig att nämna i meritförteckningen. Till exempel det finska modelldokumentet för akademiska meritförteckningar innehåller en separat rubrik ”Meriter inom produktion och tillgängliggörande av forskningsdata”.

Det är viktigt att forskningsorganisationerna satsar på tjänster inom informationsförvaltningen, såsom dataarkiv och bibliotekstjänster, så att datas kvalitet blir så bra som möjligt. Det krävs också samarbete mellan sakkunniga inom olika områden. Forskarna besitter avgörande kunskaper om sina egna material, men det behövs professionellt stöd för att säkra bevarande, användbarhet och tillgänglighet. Därför måste man satsa på och värna om tjänster, samarbete och samverkan. Det är också mycket viktigt att man integrerar planering av informationshantering i utbildningen, eftersom det ofta redan under studierna uppstår forskningsdata.

Öppenhet kräver också rättighetshantering. Tydliga data policy-dokument och avtal gör situationen lättare att överblicka. Det behövs till exempel rättigheter att kopiera och modifiera data och även att ge dylika rättigheter till tredje part, om det inte finns skäl att begränsa detta. Parter då det gäller rättigheter är förutom forskaren åtminstone organisationerna där de arbetar, de som bevarar och distribuerar material. Det finns bra material om övriga rättighetsfrågor på [Samhällsvetenskapliga dataarkivets webbplats](#).

Det är viktigt att tänka på också eventuella externa dataarkivs perspektiv. Forskningsdata kommer in och kan distribueras vidare med varierande villkor, som bör kunna hanteras. Arkivet kan till exempel behöva kunna skilja på forskare, studenter och andra användare. Det är också viktigt att beakta frågor om sekretess och integritet och hur länge olika eventuella begränsningar är i kraft.

Arkivet bör också ha beredskap att reagera om en tredje parts rättigheter blivit kränkta. Om material inte blivit försett med licens, bör man göra separata avtal med eventuella upphovsrättsinnehavare.

## **Forskarens checklista för forskningsdata**

- Skriv redan i din forskningsplan hur du tänker bevara och behandla dina forskningsmaterial under och efter arbetet
- Red ut alla etiska, juridiska och avtalsfrågor redan innan du påbörjar arbetet.
- Slut de avtal som behöver slutas med:
  - dem som producerat materialet
  - informanter och subjekt
  - dem som äger materialet
  - dem som bevarar materialet
  - alla forskarna i projekt
  - förlag eller andra utgivare
  - övriga parter
- Beakta både forskningsperioden och tiden efteråt
- Red ut hur ditt arbetsavtal påverkar situationen
- Ju öppnare lösningar du kan uppnå, desto enklare brukar det bli att publicera
- Ibland kan det krävas mycket planering och ibland extra arbete, men det brukar förbättra kvaliteten och synligheten för dig och din forskning
- Ta hand om rättighetsfrågor, licenser och att tillräcklig dokumentation blir gjord
- Se till att dina forskningsdata kan hittas i kataloger såsom Etsin i Finland
- Kom ihåg att det lönar sig att publicera också sådant material som du inte själv använt eller som gett nollresultat

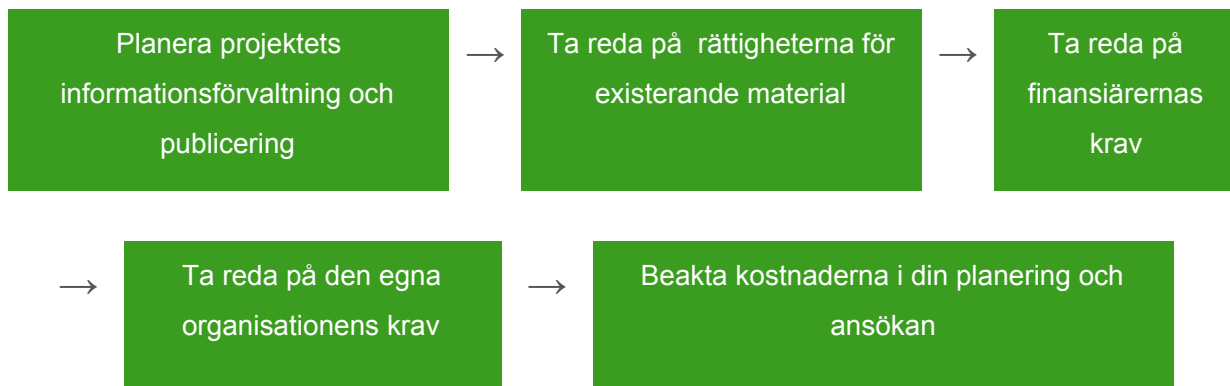
## Datahantering

Planeringen av datahanteringen är en viktig del av att planera forskningen. Genom att göra den omsorgsfullt kan man säkra att forskningsdata blir användbart. Då man planerar insamling, hantering, bevarande och återanvändningsmöjligheter blir också forskarens och forskningsgruppens arbete redigare. En plan är dessutom ofta en förutsättning för finansiering. I det finska Tuuli-projektet har man samlat material från olika organisationer gällande informationsförvaltningen. Läget utvecklas ändå hela tiden och det lönar sig alltid att söka information på den egna organisationens webbsidor eller hos forskartjänsterna.

Redan i starten lönar det sig alltså att ha en egen plan där man dokumenterar hanteringen av data. Det är bra att hålla dokumentet uppdaterat under arbetets gång. Viktiga frågor att svara på är enligt Samhällsvetenskapliga dataarkivet följande:

- Hurdant material samlar man in eller skapar man?
- Hur sker detta?
- Vilka rättigheter (ägarande, administration, upphovsrätt) hör till materialet?
- Vem bestämmer om hur materialet används?
- Hur informeras informanter eller andra subjekt?
- Vilken mjukvara används för att spara och behandla data?
- Hur säkras den tekniska kvaliteten på data?
- Vilka filformat och lagringsmedier används?
- Hurdana rättigheter ska olika användargrupper ha?
- Hur sköts säkerhetskopiering?
- Hur görs dokumentationen vid behandling av data?
- Var sparas metadata?
- Hur sköts datasäkerheten?
- Vad händer med materialet efter projektets avslutande?

På dataarkivets webbsida finns mera information om ämnet.



## Att öppna data

Då det gäller vetenskapliga artiklar har vi vant oss vid att de antingen är öppet tillgängliga eller inte. För forskningsdata finns det många flera aspekter och grader av öppenhet.

### Teknisk öppenhet

Ett öppet format betyder i allmänhet något icke-kommersiellt, oproprietärt format, som man inte behöver någon kommersiell mjukvara för att använda. Till exempel textfiler i ASCII-format går att öppna med vilket som helst program av notepad- eller texteditortyp oberoende av operativsystem, medan Microsoft Word-dokument inte nödvändigtvis går att öppna så att innehållet återges korrekt utan ett Word-program. På samma sätt lönar det sig att spara och dela information i tabellformat som kommaseparerade filer (csv), hellre än som en Excel-fil. Det lönar sig att spara information som ska delas i ett sådant format, att den kan utnyttjas i så många operativsystem som möjligt utan mjukvara som man måste betala för.

Maskinläsbarhet betyder att informationen har en sådan form att den lätt kan bearbetas maskinellt. Det vanliga PDF-formatet är till exempel inte tekniskt sett ett särskilt dynamiskt format och data i sådana filer kan inte lätt återanvändas. Maskinläsbara format bygger ofta på olika typer av XML. På många områden finns XML-baserade format för att flytta på information mellan olika system eller program. Till exempel då det gäller geodata är GML ett sådant format. I mjukvaror för olika specialområden finns ofta en möjlighet att konvertera data till sådana mer generiska format.

Tim Berners-Lees femstjärnsmodell ger en konkret bild av nivåerna på teknisk öppenhet för data. Linked Data Finland har dessutom lagt till två stjärnor.

- 1\* Publicera dina data i vilket som helst format med öppen licens, t ex Creative Commons
- 2\* Publicera dina data i en strukturerad form, t ex en XML- eller Excel-fil
- 3\* Använd ett icke-proprietärt, öppet filformat, t ex CSV i stället för Excel
- 4\* Använd dig av URI, unika identifikatorer för dina data
- 5\* Länka dina filer till andras URI för att skapa kontext
- 6\* Dokumentera data och dess schema (struktur och format)
- 7\* Validera din data (kontrollera att data följer schemat, kan göras automatiskt)

Det handlar alltså om hur dynamiska och användbara dina data är. Om man använder sig av öppna länkade data enligt modellen ovan, är resultatet oftast maskinläsbart och entydigt och går att använda för många olika ändamål. Ju mer identifikatorer du använder, desto bättre är kvaliteten generellt. Dessa frågor är rätt tekniska och det lönar sig ofta för forskare att konsultera sakkunniga, till exempel det egna områdets dataarkiv.

Mera information om maskinläsbarhet kan man hitta till exempel hos Helsinki Region Infoshare eller i Open Data Handbook.

## Öppenhet gällande användning

Ibland är det endast informationen om ett dataset som är öppet tillgänglig och det kan krävas att användaren separat godkänner användningsvillkor eller någon licens för att få tillgång till själva data. Att ett dataset finns på nätet betyder inte automatiskt att man kan använda det hur som helst. Det kan finnas både tekniska och juridiska begränsningar för vad man kan göra med materialet. Man kanske bara kan studera data, men inte nödvändigtvis röra dem. Det kan ibland löna sig eller vara möjligt för forskaren att publicera eller öppna bara delar av sina material. Det är alltid upphovsmannen som har rätt att besluta om användningen, om man inte gett ifrån sig denna rätt genom något avtal (Lagen om upphovsrätt 404/1961). Material kan antingen vara helt öppna eller också kan användningen begränsas till exempel till forskningsändamål. Detta är aktuellt främst om det ingår känsliga data (i Finland personuppgiftslagen 523/1999 11§) eller sekretessbelagd information (i Finland lagen om myndigheters offentlighet 621/1999 24§) eller affärs- eller yrkeshemligheter, som måste utelämnas då man publicerar data. Det finska Initiativet för öppen forskning och vetenskap rekommenderar användning av CC BY 4.0-licens om inte innehållet kräver annat. Creative Commons-verksamheten upprätthålls i Finland av Helsinki Institute of Information Technology (HIIT) och Medielaboratoriet vid Aalto-universitetet.

I Etsin, söktjänsten för forskningsdata i Finland, anges på vilka villkor materialet kan användas. Upphovsmannen kan ge fria rättigheter till återanvändning, men både god vetenskaplig praxis och lagen om upphovsrätt kräver ändå att man nämner upphovsmannen om man använder ett dataset. Detta gäller också om materialet försetts med CC0. Då ett dataset försetts med en Creative Commons-licens innebär det att man inte behöver ingå några separata förhandlingar eller avtal för att utnyttja det. För den offentliga sektorn i Finland rekommenderas enligt JHS 189-direktivet CC BY-licensen, som är i enlighet med EU:s PSI-direktiv. Även Avoindata.fi-tjänsten stöder användningen av CC-licenser.

## Embargo

Till exempel att forska i omfattande globala företeelser kräver ofta kostsam och arbetsdryg datainsamling och hantering men också god tillgänglighet. I sådana fall avtals ibland om ett embargo, dvs en tidsperiod under vilken producenterna har ensamrätt att utnyttja data, så att de har möjlighet att ta fram och publicera sina resultat innan de släpper data fria för övriga forskare. Längden på dessa embargoperioder kan variera beroende på förläggare, finansiärer och forskningsorganisation. Ofta handlar det om en period på 1–3 år från det att insamlingen inletts.

För publikationer som artiklar avses med embargo ofta en förläggares monopolperiod för en text, innan forskaren fritt får arkivera eller publicera texten på webben. Embargotiden räknas från publikationsdatum hos förläggaren. Om både en digital- och en pappersversion finns, räknas tiden från den tidigaste utgivningen. Embargotiderna varierar mellan olika förläggare men är i normala fall 6–12 månader. Alla förläggare kräver inte embargo.

## Fördelar med att dela forskningsdata

Genom öppen tillgång till forskningsdata kan man främja återanvändningen av dem. Detta är lönsamt förutom för den som producerat data, också för finansörerna, forskarsamfundet och samhället i stort, både nationellt och internationellt sett.

### Vad innebär öppen tillgång i praktiken?

Forskningsdata kan ha många olika grader av öppenhet beroende på licenser eller separata avtal. I princip är det upphovsmannen som avgör graden av öppenhet, men hen måste givetvis beakta lagstiftning och etik. Användning av lämplig licens är att rekommendera.

### Nyttan för forskaren

Att publicera data över webben kan ge forskare synlighet också internationellt. Publikationer som finns öppet tillgängliga får mera hänvisningar och citat än material som endast finns bakom betalväggar. Om också underliggande data finns tillgängliga växer synligheten ytterligare (se t ex "Data reuse and the open data citation advantage"<sup>3</sup> och "Scientists who share data publicly receive more citations"<sup>4</sup>). Bättre synlighet kan i sin tur bidra till bättre förutsättningar för fortsatt finansiering.

Öppenheten ökar också kommunikationen mellan forskare och kan leda till nya kontakter och viktig respons, som kan leda till att data kan förbättras. Dessutom kan det leda till nya forskningsfrågor och nya svar. Genom att publicera sina data kan en forskare också uppmuntra andra till att göra samma sak. Det lönar sig alltid att använda licenser, som klargör och säkrar forskarens egna rättigheter.

Det är redan i dag möjligt för forskare att inkludera publicerade dataset i sin meritförteckning, enligt de rekommendationer Finlands Akademi gett.

### Nyttan för forskarsamfundet

Forskersamfundet som helhet kan gynnas av öppna forskningsdata bland annat på följande sätt:

- Arbetet blir effektivare och snabbare. Återanvändning av data sparar betydande resurser, eftersom det ofta är dyrt och arbetsdrygt att samla in och putsa data. En del dataset är så unika att det kan vara omöjligt, mycket dyrt eller oetiskt att samla in dem på nytt (FSD 2013)

---

<sup>3</sup> Piwowar HA, Vision TJ. , "Data reuse and the open data citation advantage." PeerJ. 2013 Oct 1;1:e175. doi: 10.7717/peerj.175. eCollection 2013, <http://www.ncbi.nlm.nih.gov/pubmed/24109559> [29.11.2015].

<sup>4</sup> PeerJ. "Scientists who share data publicly receive more citations." *ScienceDaily*, 1 October 2013. <http://www.sciencedaily.com/releases/2013/10/131001091451.htm> [29.11.2015].

- Forskarsamfundet kan använda data för att validera resultat eller för att tolka dem på nytt eller använda dem på helt nya sätt t ex inom andra discipliner. På så sätt kan också nya forskningsmetoder tas fram. (FSD 2013)
- Stora tillgängliga dataset möjliggör komparativa studier över tid och rum. (FSD 2013)
- Att dela data ger dem större beständighet över tid.
- Tillgång till forskningsdata effektiviserar undervisning och utbildning av nya forskargenerationer.
- God tillgång till forskningsdata ökar kommunikation och samarbete också mellan forskare inom olika discipliner, vilket leder till framsteg. (FSD 2013)
- Enligt OECD:s rekommendationer<sup>5</sup> är öppen tillgång till forskningsdata en central förutsättning för ett fungerande internationellt forskningsamfund.

## Nyttan för finansiärerna

Den öppenhet och transparens där även öppna forskningsdata ingår underlättar också finansiärernas uppgift:

- Det gör det enklare att rikta finansiering rätt då data finns tillgängliga. Vid misstankar om fusk är validering lättare.
- Också produktionen av forskningsdata och dess kvalitet kan beaktas då man bedömer en forskares meriter.

## Nyttan för samhället

Då forskningsdata som producerats med gemensamma eller allmännyttiga medel finns tillgängliga ökar forskningens transparens, trovärdighet och nytta också ur skattebetalarens perspektiv. Öppenheten förbättrar demokratin och ger också underlag för faktabaserat beslutsfattande. Den offentliga förvaltningen och företagen kan dra nytta av öppna forskningsdata och det skapas bättre förutsättningar för nya innovationer och t ex start up-företag. Det är av avgörande betydelse att data är tillräckligt väl dokumenterade och till sitt innehåll begripliga, så att de kan återanvändas.

Att öppna forskningsdata främjar också internationellt sett jämlikhet och kan stöda fattigare länder i deras utveckling.

---

<sup>5</sup> OECD Principles and Guidelines for Access to Research Data from Public Funding. 2007. <http://www.oecd.org/sti/sci-tech/38500813.pdf> [29.11.2015].



# Publiceringsprocessen för forskningsdata

## 1. Välj ut och preparera data

Du måste beakta lagar, etik, avtal och finansiärernas och din arbetsgivares direktiv. För att ett dataset ska bli komplett forskningsdata, måste det dokumenteras och beskrivas tillräckligt väl. När du väljer ut data, börja t ex med sådana data som har ett bra format eller sådant som du vet kan vara efterfrågat. Du kan också i förväg ta kontakt med parter som du tror kan vara intresserade av dina data. Det kan hjälpa dig att fatta beslut. Du kan också bekanta dig med Julkinen data-webbsidorna eller Open Knowledge Open Data Handbook.<sup>6</sup> Det finns också en MOOC-kurs<sup>7</sup> om att öppna data.

- Du behöver inte publicera alla data med en gång
- Kontrollera datas kvalitet
- Anonymisera vid behov (se närmare Samhällsvetenskapliga dataarkivets webbsidor)
- Ta reda på vilken metadata som behövs
- Förse data med identifikatorer och referenser
- Konvertera data till ett öppet maskinläsbart format om möjligt

## 2. Ge en så öppen licens som möjligt

En licens definierar upphovsmannens och användarens rättigheter och gör återanvändning mycket enklare. Om rättigheterna är oklara kan det oavsiktligt försvåra återanvändning. Då man väljer licens bör man kolla upp vilka direktiv och rekommendationer som finns inom organisationen, hos finansiärerna och eventuella andra parter.

## 3. Spara materialet hos en trovärdig tjänst

Det lönar sig att välja ett dataarkiv som är trovärdigt och stabilt och som också har möjligheter att distribuera materialet effektivt. Metadata måste finnas synliga och sökbara och gärna vara öppna så att de kan exporteras till andra kataloger eller söktjänster. Du kan bekanta dig med olika dataarkiv på re3data-tjänsten eller be om hjälp i den egna organisationens forskartjänster eller bibliotek. Om du arkiverar dina data utomlands bör du se till att informationen också kommer inhemska kataloger till kännedom, t ex genom ett CRIS-system eller i Finland via Etsin, där man manuellt kan lägga till information om dataset.

## 4. Dokumentera dina data och se till att de får en beständig identifikator

I Etsin kan man manuellt föra in metadata och information överförs automatiskt dit också från en del finska dataarkiv. Om du använder något av de arkiven, behöver du inte själv skapa någon information i Etsin. En beständig identifikator, så som en DOI eller URN, gör det möjligt att entydigt hänvisa till ett visst dataset och det motsvarar på sätt och vis böckernas ISBN-nummer.

---

<sup>6</sup> Open Data Handbook, <http://opendatahandbook.org/>.

<sup>7</sup> Datan avajaan MOOC, <http://datakoulu.fi/kurssit/datan-avaajan-mooc/>.

Etsin- och IDA-tjänsterna ger automatiskt URN-nummer för forskningsdataset. Om man arkiverat sina data utomlands, kan man manuellt föra in deras beständiga identifikator i katalogen.

## **5. Att göra en applikation**

Ibland kan det vara värt att skapa en skild applikation, t ex en webbsida där man kan söka eller studera data närmare. I Finland finns AVAA-tjänsten för detta ändamål.

## Datasäkerhet

Datasäkerhet betyder i allmänhet att ett objekt, i detta fall forskningsdata, skyddas tillräckligt mot hot mot dess trovärdighet, integritet och tillgänglighet. I vissa fall kan det finnas skäl att fästa uppmärksamhet vid sekretess och autenticitet. För forskaren innebär det i praktiken att man hanterar informationen så att man ser till tillräckliga begränsningar vad gäller åtkomst och användning av data när det är påkallat. Stöd och råd för detta kan man få av den egna organisationens tjänster och av Samhällsvetenskapliga dataarkivet i Finland.

Det lönar sig alltid att fundera över var och hur man bevarar sina data: är förvaringsplatserna tillräckligt pålitliga och säkra? Tjänster för forskning är ofta tryggare och mer oproblematiske vad gäller lagstiftning än allmänna kommersiella tjänster. Om man har känsliga data måste man vara särskilt försiktig.

Ibland är man tvungen att begränsa åtkomsten till information av sekretessskäl. Då måste någon ha ansvaret för vilka rättigheter (läsa, skriva, dela) som delas ut och till vem. Ofta är det den som äger systemen och data som ansvarar för detta, men ibland kan det vara forskaren själv som har åtminstone en del av ansvaret. Då det gäller teknisk hantering av åtkomst och rättigheter behövs tekniker för autentisering (t ex användarnamn och lösenord) och auktorisering (utdelande av olika rättigheter). I Finland används inom forskningssektorn allmänt Haka-systemet för autentisering och för hanteringen av auktorisering finns REMS-tjänsten. Det är också viktigt att systemen man använder sparar tillräckligt noggranna loggfiler.

Till åtkomsten hör lika så den fysiska aspekten: vem kommer åt dataterminaler och media och behöver detta begränsas och kontrolleras för att t ex trygga datas integritet eller sekretess?

Med risker avser man skador som kan uppstå antingen av en extern händelse, av en aktiv åtgärd eller av passivitet. Man kan tänka sig att en risk kan värderas utgående från dess sannolikhet multiplicerat med effekterna av den skada som sker. Då man bedömer sannolikheten för olika skador är tidigare erfarenheter ofta av värde, medan effekterna ofta har mera att göra med hurdan själva objektet är. Risksituationer kan uppstå antingen av aktivt ofog eller av tekniska orsaker. Det senare är vanligare då det gäller datahantering.

Dataintrång är mycket vanligt i olika webbtjänster, medan det är mycket ovanligt att ett helt dataarkiv skulle förstöras. Om data kan återskapas, är effekten av deras försvinnande eller korruption inte nödvändigtvis så stor.

Att data är sekretessbelagt betyder att endast vissa parter har rätt att hantera dem. Om information är offentlig behöver man inte begränsa åtkomsten, men i andra fall måste man utnyttja autentisering, auktorisering eller andra metoder för att begränsa åtkomsten. Den finska staten har en särskild förordning (VnA 682/2010) om klassificering, men utanför statsförvaltningen kan klassificeringen göras friare, t ex i kategorierna offentlig, begränsad, intern, konfidentiell och sekretessbelagd. Den som äger data bestämmer hur materialet klassificeras och vem som har rätt att hantera dem.

Datas integritet betyder att endast de som har tillbörliga rättigheter kan göra förändringar. Det betyder också att de är koherenta och inte korrumpas av felaktigheter.

Tillgänglighet innebär att personer som har rättighet kan komma åt och utnyttja data enligt överenskommelse. För snäv åtkomst kan till exempel leda till att materialet inte är tillgängligt. I vissa fall är autenticitet och proveniens särskilt viktiga. Då måste man kunna påvisa att data är oförändrade.

Syftet med dataskydd är att skydda människors integritet på ett sådant sätt att inte personuppgifter behandlas eller avslöjas utan tillstånd. I den finska lagen om personuppgifter (523/1999) regleras hur sådana data ska hanteras. Säkerhetskontroller för att säkra dataskydd kan innebära tekniska kontroller eller annan kontroll som baserar sig på lagstiftning eller avtal. Ibland är dylika dyra att göra eller de kan försvåra åtkomsten, utan att alltid helt eliminera alla risker. Det är inte alltid ändamålsenligt eller möjligt att eliminera samtliga riskfaktorer.

I tabellen framkommer vanliga risker vid datahantering och de åtgärder och kontroller som kan vidtas.

<b>Risk</b>	<b>Åtgärder</b>
Förlust av data pga fel eller misstag	Säkerhetskopiering, versionshantering
Känsliga data läcker	Åtkomstkontroll, rättighetshantering, kryptering, sekretessavtal
Korruption av data	Kontrollsummor, säkerhetskopiering, versionshantering, integritetscheck
Otillgänglighet	Avtal om service, säkerhetskopiering, val av trovärdiga leverantörer
Forskaren blir åtalad pga felaktigt bruk av personuppgifter	Avtal, åtkomstkontroll, kryptering
Systemägaren missbrukar data	Avtal, val av certifierade system och leverantörer

Dataintrång	Skydd av arbetsterminaler och servers, goda lösenord och åtkomstkontroll
Malware	Tekniskt skydd och trovärdiga system och leverantörer
Spionage	Tekniskt skydd och trovärdiga system och leverantörer
Åtal pga brott mot gällande lagstiftning	Säkrande av avtal och lagenlighet, försäkring
Förlust av data	Planering av bevarande
Data går ej att hitta	Metadata av god kvalitet, trovärdiga och sakkunniga serviceproducenter

När en forskargrupp eller enskild forskare planerar datasäkerheten, lönar det sig att gå till väga systematiskt och förnuftigt. Att vidta förhastade åtgärder utgående från lösa rykten ökar sällan säkerheten, utan det är viktigare att se till att grunden är sund och fungerande.

Då man vill försäkra sig om datasäkerheten lönar det sig att börja med att identifiera de viktigaste faktorerna och sekretessklasserna som gäller. Är informationen offentlig eller bör man begränsa åtkomsten? Vanligen behövs särskilda rättigheter för att kunna ändra information. Data som kan ändras anonymt är sällan trovärdiga, åtminstone om man inte har någon kontroll på förändringarna som görs. Man ska också fundera på integritets- och åtkomstfrågor.

Det lönar sig att dokumentera risker och säkerhetsåtgärder och fundera över hur kontroller och andra åtgärder görs och av vem. Ofta lönar det sig att överlåta tekniska rutiner på tjänsteleverantörerna om det låter sig göras. På så sätt kan ägaren utlokalisera en del av riskhanteringen.

Se även The ISF Standard of Good Practice for Information Security.<sup>8</sup>

---

<sup>8</sup> The ISF Standard of Good Practice for Information Security,  
<https://www.securityforum.org/tool/the-standard-of-good-practice-for-information-security/> [29.11.2015].

## Metadata för forskningsdata

För att en forskare ska kunna använda ett dataset måste hen kunna bedöma om materialet lämpar sig för forskningsuppgiften. Man behöver då ofta information om hur materialet kommit till, hur det ursprungligen använts, tillgängligheten, och vem som skapat materialet och hur man får använda det. Sådan information kan anges i ett strukturerat format i form av metadata.

Metadata kan ingå som en del av data eller i en separat katalog till exempel i en databas. Metadata kan delas upp i tre olika typer beroende på deras syfte: 1. Beskrivande metadata beskriver materialets innehåll och karaktär. Denna information produceras antingen av forskaren själv eller en tredje part, t ex dataarkivets personal. Användningen av existerande ordlistor, tesaurusar, klassificeringssystem och ontologier är att rekommendera. Dessa finns ofta integrerade i olika informationssystem. 2. Administrativ metadata definierar nödvändig information om tekniska egenskaper och rättighetsfrågor. Denna information är viktig också med tanke på långsiktig bevaring av data. 3. Strukturell metadata beskriver strukturen, t ex hur olika delar av ett dataset förhåller sig till varandra. Administrativa och strukturella metadata kan ofta skapas maskinellt.

Metadata skapar förutsättningar för att data kan hittas och att man kan få tillgång till dem och återanvända dem. Genom att skapa metadata kan forskaren få synlighet och merit. Den som vill använda data kan för sin del identifiera, evaluera och får tag i dem genom att få tillgång till dessa uppgifter.

Hur mycket och hurdan metadata som behövs varierar mellan olika forskningsområden. Inom många discipliner har man utvecklat egna metadataformat som är anpassade just för det områdets behov. Dessa format kan man ofta studera i tjänster som hanterar forskningsdata för respektive område. Dessa hjälper också forskare att skapa relevanta metadata.

Idén med metadata är alltså att besvara frågor som är relevanta inom respektive forskning. I tabellen nedan presenteras vanliga typer av beskrivande metadata och de frågor de besvarar. Då man skapar system eller scheman för metadata ska man fästa uppmärksamhet vid de relevanta frågorna, det vill säga de saker som är viktiga att känna till om ett dataset i framtiden. Med hjälp av dessa frågor kan man gestalta de scheman och format för metadata som ligger som grund för olika tjänster.

Med tanke på kvaliteten av metadata är det viktigt att fästa uppmärksamhet vid en tillräckligt täckande, entydig och begriplig beskrivning av forskningsmaterialet, till exempel då det gäller användningsvillkor och rättigheter. God metadata är både forskarens och datas visitkort.

<b>VAD?</b>	
Vad heter materialet?	Namn (title)
Vad innehåller det?	Beskrivning (description)
Till vilket område/ vilken disciplin hör det?	Klassificering (classification)
Vilka saker har materialet att göra med?	Ämne (subject)
På vilket språk är materialet?	Språk (language)
Vilka variabler ingår?	(dimension)
Vilken publikation hör det till?	(publication)
<b>VAR?</b>	
Vid vilken organisation är materialet producerat?	Organisation (organization)
I vilket projekt är materialet producerat?	Projekt (project)
I vilket land är materialet producerat?	Land (country)
Var kan man hitta materialet?	Katalog (catalog)
I vilket format finns materialet tillgängligt?	Filformat (format)
<b>NÄR?</b>	
Vilken tidsperiod täcker innehållet?	Täckning över tid (temporal)
När är materialet publicerat?	Publicering (issued)
När har materialet ändrats?	Revidering (modified)

<b>VEM?</b>	
Vem har skapat materialet?	Upphovsman (author)
Vem äger materialet?	Ägare (owner)
Vem distribuerar materialet?	Distributör (distributor)
<b>HUR?</b>	
Hur har materialet skapats?	Beskrivning av metod (method)
Hur kan man få mer information?	Publikationer som gäller metoden (link)
Hur får man använda materialet?	Användningsvillkor (rights)
Hur får man mera information om materialet?	Kontaktuppgifter (distributor contact)
Hur hänvisar man till materialet?	Identifikator (identifier)
Hur hänvisar man till materialet i en publikation?	Referens (citation)
<b>VARFÖR?</b>	
Varför är materialet producerat och publicerat?	Datahanteringsplan (DMP)
Till vad är materialet avsett?	Ämne (subject) / Beskrivning (description)

### Forskarens checklista för metadata

- Se till att alla dina forskningsdata och publikationer får bra och öppen metadata
- Uppgifter som gäller tid och rum är också viktiga, ange dem enligt kutym. Tjänster har ofta direktiv för detta.
- Använd egen branschterminologi och ämnesord.
- Kom också ihåg att ge en tillräckligt allmän men noggrann beskrivning i form av fri text (description): Vad handlar det om? Länka gärna till artiklar som hänvisar till materialet.
- Utnyttja färdiga listor på ämnesord, ontologier och vanliga termer så att materialet faktiskt kan hittas
- Licens och användarvillkor är en mycket viktig del av metadata. Ge en så öppen licens du kan.



- Kom ihåg att beakta information om olika versioner eller åtgärder som vidtagits, så att det blir klart och entydigt vilken typ av data det handlar om och hur den eventuellt förhåller sig till andra versioner.
- Vid dataarkiv och bibliotek hittar man professionellt stöd för produktion av metadata. Tveka inte att be om hjälp!
- Den som är intresserad av metadata kan bekanta sig med Aila, Det nationella digitala bibliotekets metadatarekommendationer eller Digital Curation Centres metadataregister.

## Identifikatorer och referenser

Hållbar forskning kräver också unika beständiga identifikatorer. En identifikator är till exempel en egen unik kod för en publikation, organisation, ett dokument eller en person. Med hjälp av den kan man försäkra sig om att man hänvisar till just precis den enhet eller person man avser. Man har nytta av till exempel identifikatorn för en forskare, ORCID, särskilt om någon använder flera olika namnformer eller om det finns flera personer som har samma namn. Med hjälp av identifikatorer kan man också säkra att hänvisningar och referenser är bestående över tid och att länkar inte går sönder. I den digitala miljön kan byten av datasystem eller nya versioner kräva att forskare agerar långsiktigt för att deras forskning ska vara trovärdig. Också ämnesord och begrepp kan ha egna identifikatorer, som gör att de kan ha olika parallella språkvarianter och deras inbördes relationer kan beskrivas i maskinläsbart format. En sådan tjänst erbjuds till exempel av den finska Finto-ontologitjänsten.

De system som tjänar den öppna forskningen måste beakta skillnader och olika behov som finns inom olika forskningsområden, men också sträva till så bra sökbarhet och kompatibilitet som möjligt. För att forskningsinformation ska vara så användbar och sökbar som möjligt måste man fästa uppmärksamhet vid informationsstrukturer och semantik, det vill säga hur man uttrycker olika saker med ordval och identifikatorer, som borde vara så enhetliga och kompatibla som möjligt.

Det finns olikheter mellan olika publikationer och dataset som beror på olika praktiker och traditioner inom olika forskningsområden. I allmänhet (t ex inom ERC, Horizon 2020) rekommenderas att man i första hand utnyttjar det egna forskningsområdets publikationer och dataarkiv, då sådana finns. Då får materialen en ändamålsenlig kontext och struktur. Då man väljer dataarkiv bör man också granska kritiskt om tjänsten möjliggör öppenhet och återanvändning i tillräcklig grad. Dessutom är det viktigt att informationen om publikationer och forskningsdata och tillräcklig metadata och identifikatorer också hittar fram till de egna CRIS-systemen och i Finland till Etsin-tjänsten.

Gemensamma system för identifikatorer är speciellt viktiga då informationssystem utbyter data med varandra över öppna gränssnitt. I sådana fall måste innehåll såsom kodböcker och deras versioner också kunna särskiljas och identifieras. Internationella system för unika identifikatorer är till exempel systemen med telefonnummer, e-postadresser, böcker (ISBN-nummer) och varor (EAN-kooder). Andra identifikatorer är till exempel URI-namn (Uniform Resource Identifier).

För publikationer kan man beroende på publikationstyp använda sig av t ex ISBN (monografier) eller andra beständiga identifikatorer (persistent identifier, PID), av vilka Handle används i publikationsarkiv, DOI i kommersiella förläggares system och URN av nationalbiblioteken i deras digitala samlingar. För forskningsdata används i finländska system ofta URN. Det finns flera tjänster som delar ut av de URN-identifikatorer som Nationalbiblioteket tillhandahåller. Utöver dataset behöver också inblandade juridiska personer identifikatorer (organisationer, dataarkiv, forskargrupper osv). Användningen av dessa utreds i Finland ännu separat.

Enligt de finska direktiven för god vetenskaplig praxis bör forskare erkänna andra forskares arbete och insatser genom att hänvisa till dem då de används. Om man i detta sammanhang använder sig av standardiserade identifikatorer kan dessa kontrolleras bättre och man har också möjlighet att trygga att länkar hålls fungerande över tid. På så sätt stöder identifikatorerna informationssökning och synlighet. Ontologier, register och auktoritetstjänster kan erbjuda dylika standardiserade identifikatorer att använda i metadata.

## **Forskarens checklista för identifikatorer och referenser**

- Se till att du har en egen ORCID-identifikator och använd densamma så ofta du kan
- Använd alltid identifikatorer då du ges möjligheter
- Se till att dina egna publikationer och dataset har identifikatorer
- Skaffa alltid en ny identifikator då du skapar en ny version, så att referenserna är entydiga
- Undvik att i onödan skaffa flera identifikatorer för samma objekt
- Använd hänvisningar och identifikatorer flitigt!

## Den öppna forskningsprocessen

I den digitala världen kan man också använda öppenheten som en metod för att effektivisera forskningsprocessen och öka sitt genomslag. Trycksakskulturen styrde forskningsprocessen mot att publicera färdiga resultat som en manifestation av vetenskaplig kunskap och merit. Samverkan och distribution av kunskap före publicerandet av resultaten har varit svårt av praktiska orsaker och ofta begränsat till en mindre intern akademisk krets. I dag är det möjligt att dela och göra forskningsarbete tillsammans i en större krets, till och med globalt.

Då texter och olika versioner och kopior av dem kan spridas vitt och brett och även den stora allmänheten har tillgång till stora mängder information, skapar det också ett tryck på forskningen att kunna öppna upp och sprida vetenskaplig information så mycket som möjligt. Det är också viktigt att visa hur vetenskapen och forskningen fungerar, vad kunskapen baserar sig på och hur den egentligen blir till. Forskare gör ofta sitt arbete just för att det är både spännande och intressant och också denna aspekt av saken kan man med fördel försöka förmedla till en större publik. Eftersom det i dag också finns så mycket falsk och felaktig information i omlopp, har det blivit allt viktigare att göra allmänheten mer delaktig i själva forskningsprocessen närhelst det är etiskt och juridiskt möjligt. Under en öppen process är det också lättare att hitta svagheter och fel redan under arbetets gång på samma sätt som man tidigare gjort på seminarier eller konferenser forskare emellan. Detta är ägnat att förbättra kvaliteten och trovärdigheten.

Man kan öppna upp processen genom att använda gratisjänster som finns på webben, till exempel Flickr, SlideShare, YouTube, Vimeo, bloggar, Twitter, Facebook eller andra liknande tjänster. Dessa kan också fungera som interna kommunikationskanaler mellan forskare i samma projekt eller inom ett forskningsområde. De har då den fördelen att de också kan vara öppna mot en större allmänhet. En forskare som är aktiv inom sociala medier skapar sig samtidigt en bra plattform att ta till om man vill korrigerera felaktiga uppgifter i offentligheten eller i traditionella medier. Det är mycket önskvärt att forskarsamfundet är starkt närvarande på webben. Det är också bra att planera hur de material som föds på webben eventuellt ska hanteras och arkiveras. Vi återkommer i ett senare avsnitt till användningen av sociala medier.

### *Offentliggjord information kan inte stjälas – Forskaren Samuli Ollila*

Så kallade gråa publikationer har traditionellt varit just ”halvfärdiga” resultat eller presentation och testande av olika idéer till exempel vid seminarier eller konferenser. Då man publicerar gråa material lönar det sig att kontrollera att man inte sinkar t ex senare patentansökningar. För övrigt har man tidigare ofta låtit bli att öppna publicera preliminära resultat eftersom de betraktas som osäkra eller på grund av konkurrens. Man är kanske rädd för att någon ska stjäla ens resultat eller idéer eller att man ska förlora sitt eget anseende på grund av felaktigheter. Men att publicera material tidigt, till exempel som en powerpointpresentation i ett publikationsarkiv är samtidigt ett sätt att markera upphovsrätt och underställa det tidig kritik redan innan man producerar en slutgiltig publikation. Det kan därför också ses som ett slags öppen kollegial granskning. Dessutom ger sådant publicerande möjlighet för andra att hänvisa också till pågående arbete, vilket kan vara värdefullt om det handlar om ett långt projekt där de slutliga resultaten och publikationen dröjer länge. Man måste förstås fästa uppmärksamhet vid källkritik och informationskompetens, men man

bör också kunna lita på den offentliga diskussionen och att vetenskaplig information alltid kan underbyggas med argument och verifieras. Forskare bör alltså vara noggranna med att alltid ange kontexten för saker som publiceras offentligt och skriva ut alla reservationer tydligt i varje sammanhang. Eftersom material kan sprida sig mycket snabbt på webben är det därför skäl att till exempel i powerpoint- och andra presentationer alltid tydligt ange datum och i vilket sammanhang de presenterats, så att var och en kan göra en korrekt kritisk utvärdering av innehållet och bedöma hur informationen kan användas.

Man kan också öppna upp forskningsprocessen genom att bjuda in lekmän eller en större allmänhet i att skapa eller behandla forskningsdata. Det kan vara möjligt att öka forskningens samhällsnytta på detta sätt. Kvalitetskontroll kan genomföras på olika sätt, till exempel genom att låta olika personer upprepa samma uppgift, genom att göra stickprov eller analysera data statistiskt. Allt detta är det skäl att planera noggrant redan i början av projektet och alla processer och åtgärder ska förstås dokumenteras. Det är också viktigt att låta dem som medverkat få ta del av hur arbetet framskrider och vilka resultat man fått.

## **Forskarens checklista för en öppen forskningsprocess**

- Planera kommunikationen på förhand
- Kommunikation kan vara både intern inom ett projekt och extern, beakta båda delarna
- Fundera på hur öppen den interna kommunikationen kan vara utåt
- Den externa kommunikationen riktar sig både till andra forskare och till allmänheten
- Kom ihåg bloggar och sociala medier som är dagens öppna seminarier och ett effektivt sätt att nätverka
- Utnyttja mediekontakter, skriv pressmeddelanden – be om hjälp av organisationens informatör
- Fundera över i vilket skede du kan publicera vad (rådata, forskningsdata, metoder, arbetspapper, presentationer, video ...?)
- Hur hanterar ni versionshantering och delande inom projektet?
- Vilka program och plattformar kan ni använda för samarbete och att skriva tillsammans?
- Hur kan man länka mellan olika medier, plattformar och kanaler? Vad vill ni arkivera?
- Hänvisa till dig själv, puffa intressanta material, återanvänd egna material i olika medier och sammanhang
- Kom ihåg att också förse bilder, bloggtexter mm med Creative Commons-licenser för bättre spridning
- Sprid flitigt också av kollegernas material, det är viktigt att vara generös och korrekt på webben
- Delta i diskussioner på bloggar, Twitter osv. Var positiv och konstruktiv. Då kan du också puffa egna material på ett bra sätt
- Programkod kan delas och utvecklas t ex på GitHub
- Man kan skriva öppet på bloggar, wikier eller t ex på Google Drive. Men kom alltid ihåg datasäkerheten!
- Det finns många plattformar för samtal: Eduuni i Finland, BaseCamp, f1000 eller sociala medier och deras grupper. Kom ihåg arkiveringen!
- Kom ihåg, att fast du inte kan publicera allt, kan du ofta dela något eller åtminstone berätta vad du jobbar med!

## Att publicera sina resultat

Vetenskaplig utgivning skiljer sig på många sätt från annat publicerande. De vetenskapliga publikationerna anses ha fyra olika funktioner:

- kollegial granskning (kvalitetskontroll)
- registrering av publikationen
- informationsspridning
- arkivering

Författarna får vanligen ingen separat ersättning för sitt skrivande, utan de är tvärtom ibland tvungna att betala förläggaren för att få sin text publicerad. Nyttan för skribenten är alltså inte en direkt ekonomisk nytta, utan de kommer på omvägar genom att de visar vilka resultat man varit med om att producera och de ger också forskaren egen merit.

Utöver att fungera som ett medium för vetenskaplig kommunikation, fyller de vetenskapliga publikationerna och den kollegiala granskningen viktiga funktioner vad gäller kvalitetsgranskning och meritering. Publicering sker i dag huvudsakligen utanför forskningsorganisationerna. Trots att forskarna som verkar vid universitet och andra forskningsinstitutioner har en avgörande betydelse som skribenter, granskare och redaktörer är det oftast vetenskapliga samfund eller kommersiella förlag som ansvarar för utgivningen och dess utvecklande.

Sedan millennieskiftet har Open Access-publicerandet vunnit terräng som ett alternativ till de tidigare affärsmodellerna, som byggd på prenumerationsavgifter. Utvecklingen började redan i början av 1990-talet då den digitala utgivningen tog fart och bland andra preprint-arkivet inom fysik arXiv grundades. Å ena sidan steg priserna på digitala vetenskapliga tidskrifter mycket snabbt, samtidigt som material började finnas tillgängligt gratis på webben. I arXiv publiceras idag nästan alla relevanta artiklar inom området, ofta redan innan de publicerats i traditionella tidskrifter. Detta har förbättrat tillgängligheten och snabbat på kommunikationen på området.

Politiskt blev man mer medveten om utvecklingen i och med de internationella Budapest- (2001) och Berlin- (2003) deklARATIONERNA. Den senare undertecknades också av de finska universitetens rektorer år 2006. I praktiken var avsikten att parallellt främja två olika modeller för publicering och öppen tillgänglighet: den gyllene vägen (Gold OA) som baserar sig på förläggare som utgivare, och den gröna vägen (Green OA) som innebär parallellpublicering i publikationsarkiv. Diskussionen om för- och nackdelar med de olika alternativen har varit omfattande och pågår fortfarande. Olika mellanformer och andra alternativ finns också. I verkligheten har båda huvudalternativen sina fördelar, och de borde inte ses som varandra uteslutande utan som kompletterande modeller.

Uppgifter om publicering har också i Finland blivit en viktig mätare för verksamhetens kvalitet och resultat i och med den nya finansieringsmodellen för universiteten som antogs 2013. Både CRIS-systemen och publikationsarkiven är centrala delar då det gäller att följa med forskningens resultat.

## Forskarens perspektiv på Open Access

I princip utesluter inte den gyllene vägen, dvs utgivningen på förlag med eventuella därtill hörande avgifter (APC, Article Processing Charge) användningen av publikationsarkiv, som alltid är att rekommendera. Genom att publicera sina texter i publikationsarkiv kan forskaren försäkra sig om att texten finns bestående och öppet tillgänglig i digital form både för en själv och för alla andra. Öppen publicering gör inte att upphovsrätten försvinner, men den gör det möjligt att göra klara och tydliga hänvisningar, det ger tillgänglighet och möjliggör även forskning i data. Också forskningsorganisationerna och andra aktörer har nytta av att de resultat som produceras i deras hägn finns samlade, synliga och hittbara. Därför rekommenderar eller till och med kräver många aktörer open access-publicering. Upphovsmannen kan ändå förse sitt arbete med en lämplig Creative Commons-licens, för att underlätta dess spridning. Då man väljer förläggare finns det orsak att vara mycket försiktig och se upp för dåliga tidskrifter som bara är ute efter ekonomisk vinning.

Då man gör upp förlagsavtal lönar det sig alltså att se till att man förbehåller sig tillräckliga rättigheter för till exempel parallellpublicering. Många förläggare godkänner att man arkiverar sin artikel på webben, men det finns ofta olika villkor om vilken version man får lägga ut eller när man får göra det. Man kan bekanta sig med olika tidskrifters linje i Romeo/Sherpa-tjänsten, men det lönar sig förstås alltid att kontrollera saken direkt hos förläggaren. Det är viktigt att i publikationsarkivet klart uppge vilken version det handlar om. De flesta förläggare godkänner parallellpublicering efter en s.k. embargo-period, t ex ett halvt eller ett år efter att tidskriften publicerats. Man kan också parallellpublicera monografier. Detta kan ofta göras efter två år, om nya upplagor inte tagits. Bildmaterial utgör ofta en utmaning och ibland kan det enda alternativet tyvärr vara att man tar bort bilderna inför parallellpubliceringen om man inte avtalat om saken eller upphovsrätten har gått ut. Det finns oftast handledning att få inom den egna organisationen.

## Forskarens checklista för publicering

- Se till att publicera så öppet som möjligt
- Om du vill skaffa patent, måste det ske före publiceringen
- Parallellpublicering påverkar inte din upphovsrätt, men ökar din synlighet
- Ta reda på vilka publikationsarkiv din finansiär och arbetsgivare rekommenderar
- Genom att använda en licens som möjliggör återanvändning (t ex CC BY 4.0) förbättrar du möjligheterna för att ditt arbete delas och vidareutvecklas
- Välj alltid öppen publicering då du har möjlighet att göra det
- Sträva efter snabbhet och öppenhet i publiceringsprocessen. Undvik förläggare som kräver långa embargo-perioder eller som försvårar parallellpublicering
- Be om hjälp t ex från biblioteket då du gör avtal med förläggare
- Kom ihåg identifikatorer och hänvisningar i länkform. Använd dem så ofta du kan.
- Se till att dina publikationer får bra metadata. Be om hjälp i biblioteket.
- Publicera och länka till metoder och forskningsdata, det ökar både synlighet och validitet.
- För att publicera data kan du i Finland också utnyttja AVAA-tjänsten.
- Se till att alla dina publikationer finns i publikationsarkiv. Vid behov kan du spara dina texter på flera ställen, t ex i ett internationellt domän-specifikt och organisationens eget publikationsarkiv. Det ökar synligheten.
- Också gråa material kan ofta publiceras i publikationsarkiv. Det är ett gott alternativ, för man får samtidigt en beständig länk och bevaringen är tryggad. Till exempel konferensers

webbsidor är som bekant inte alltid eviga. Kom ändå ihåg att se till att du inte publicerar något för tidigt, så att t ex ett patent blir omöjligt att skaffa.

- Välj en så öppen licens på dina material som möjligt (Creative Commons)
- Finns det andra sätt att publicera det du kommit fram till? Infografer, video, visualisering?
- Var noggrann med etik och dataskydd. Utförlig dokumentation och öppenhet är god praxis.



## Sociala medier och webbpublicering

De sociala medierna och andra liknande publicerings- och kommunikationsverktyg har blivit allt viktigare redskap också för forskare. Det är förhållandevis enkelt att kommunicera eller skriva tillsammans på olika webbaserade tjänster. De är ändå förknippade med vissa aspekter som är bra att ta hänsyn till då de används i vetenskapligt arbete. Om man till exempel skriver tillsammans är det bra att från början ha klart utskrivna regler om rättigheterna till arbetet som alla godkänner. Om man vill inkludera andras material, bilder eller dylikt måste man alltid säkra att man har tillräckliga rättigheter att göra det. Det hör förstås till god kutym att alltid ange upphovsmannens namn. Om denna är en pseudonym eller en annars anonym profil, duger denna som uppgift.

För att bearbeta och dela bilder på webben finns det också färdiga tjänster, men kom alltid ihåg att kontrollera vilka rättigheter och licenser respektive material har. Creative Commons har en egen söktjänst, där man kan söka material som försetts med licens. Också Wikimedia Commons har mycket material med tydliga licenser.

Om du har för avsikt att publicera bilder eller inspelningar där enskilda personer kan identifieras, ska du alltid be om lov. Om man till exempel vill streama ett seminarium är det bra att be talarna skriva på ett tillstånd. Fråga också om lov innan du lägger ut andras presentationer på nätet.

Då man parallellpublicerar ska man beakta den egna organisationens direktiv gällande publikationsarkiv. Man kan inte heller lita på att webbtjänster och sociala nätverk alltid kommer att finnas tillgängliga. Att parallellpublicera artiklar eller data i tjänster som ResearchGate eller Academia uppfyller inte finansierarnas kriterier på open access-publicering. Det ena utesluter inte det andra och den synlighet dylika tjänster erbjuder kan vara viktig, men dylika tjänster ersätter inte öppen publicering eller arkivering.

### Forskarens checklista för sociala medier

Då du använder sociala medier i ditt arbete är det bra att beakta följande saker:

- Om du har för avsikt att skydda dina resultat med patent eller annan immateriellrätt som vid registreringen kräver absolut nyhetsstatus, sprid inte information om ämnet i förväg i sociala medier. Sådan information ska hållas hemlig åtminstone till ansökan är inlämnad.
- Du kan inte publicera något sådant i sociala medier eller annars på webben, som i avtal har definierats som konfidentiell information.
- Om du i din forskning arbetat med till exempel intervjuer eller andra material där personuppgifter ingår, måste du vara noggrann med personskydd och etik. Nämn ingen information som kan göra en person möjlig att identifiera.
- En del förläggare vill endast ge ut material som inte publicerats tidigare. Vad detta innebär i praktiken är ofta oklart och kan variera.
- Beakta din egen organisations och lagens krav på datasäkerhet och -skydd. Webbtjänster tar ofta mycket litet ansvar för detta.
- Det är bra att i forskargrupper och projekt diskutera igenom användningen av webb och sociala medier, så att alla är på det klara med till exempel vilken information som är konfidentiell.