

# Statsvetenskapliga förbundet

Förbundsredaktör: Katarina Roos

## Ett öppet vetenskapssystem

av Katarina Roos

För snart två år sedan påbörjade jag ett manus till förbundssidorna om open access. Rubriker i dags- och fackpress vid den tiden signalerade att frågan om öppen tillgång väckte starka känslor. Nu, ett par år och en pandemi senare är öppen tillgång en standardpost vid ansökningar till forskningsråden. Förlagsbranschen har anpassat sig och kritikernas värsta farhågor verkar inte ha besannats. Öppen tillgång till vetenskapliga publikationer är emellertid bara en del i det politiska och akademiska projekt som går under benämningen Open Science. Öppen vetenskap är temat för förbundssidorna denna gång.

### OPEN SCIENCE – ÖPPEN TILLGÅNG TILL VETENSKAPLIGA PUBLIKATIONER

EU har varit och är en drivande aktör för processen som ryms under paraplyet Open Science. Redan med Horizon 2020 som antogs 2013 formulerades öppen tillgång till vetenskaplig publicering som ett krav. En fördröjning av publicering på upp till 6 månader godtogs dock (European Commission 2016). Nederländerna är det medlemsland som drivit frågan hårdast. Många avgörande beslut fattades under första halvan av 2016 då landet var ordförande i EU:s ministerråd. Det nederländska ordförandeskapet deklarerade att open access var en deras mest prioriterade frågor. I maj 2016 fattade medlemsländernas ministrar i Konkurrenskraftsrådet ett inriktningsbeslut som markerade den formella starten på en process som i själva verket hade pågått sedan lång tid tillbaka. Kommissionären med ansvar

för forskning, vetenskap och innovation gav ett särskilt sändebud i uppdrag att leda arbetet med en plan för open access och att bygga en koalition till stöd för denna. Tidsramen var mycket kort och det särskilda sändebudet beskrev i en intervju uppdraget som omöjligt (Enserink 2016a/b).

Science Europe och Europeiska forskningsrådet (ERC) är också centrala aktörer i processen. Den förstnämnda är en sammanslutning av nationella forskningsfinansierare och i denna ingår Forte, Formas och Vetenskapsrådet. Representanter för de två institutionerna samt nationella forskningsfinansierare engagerades i arbetet. Resultatet blev cOAlition S och överenskommelsen Plan S. På hemsidan beskrivs cOAlition S som en grupp av nationella forskningsfinansierare, europeiska och internationella organisationer och välgörenhetsorganisationer som genom en internationell överenskommelse förbundit sig att – med stöd av EU-kommissionen och Europeiska forskningsrådet implementera Plan S. Den innebär att, “...all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo.” (cOAlition-S 2021). Nyckelordet i meningen är *immediately*. All vetenskaplig publicering måste göras tillgänglig utan dröjsmål. När Plan S offentliggjordes hösten 2018 inbjöds nyckelaktörer att lämna synpunkter, vilket ledde till att slutdatum för genomförande flyttades från 1 januari 2020 till 1 januari 2021.

## OPEN SCIENCE – ÖPPEN TILLGÅNG TILL FORSKNINGSDATA

Öppen tillgång till vetenskapliga publikationer var dock inte det enda som EU:s medlemsländer enades om vid det tidigare nämnda rådsmötet i maj 2016. Temat för mötet var "The transition towards an Open Science system" (Council of the European Union 2016), på svenska "ett öppet vetenskapssystem" (Vetenskapsrådet 2021). I ett sådant är även öppen tillgång till forskningsdata [min kursivering] en hörnsten. I rådslutsatserna hänvisas till FAIR-principerna, findable, accesible, interoperable och reusable, vilka vi har anledning att återkomma till. I Sverige är det Vetenskapsrådet som av regeringen har fått huvudansvar för att samordna övergången mot öppen vetenskap tillsammans med andra nyckelaktörer.

En kärnfråga är naturligtvis definitionen av forskningsdata. Enligt Vetenskapsrådet är det "...data som har samlats in och/eller skapats under forskningens gång" (Vetenskapsrådet 2020: 14). "Forskningsdata kan vara av många olika typer och format, såsom resultat från mätningar och experiment, observationer från fältarbete, enkätsvar, intervjuer, bilder, film, och texter. Forskningsdata kan också vara av olika analysnivå, grad av bearbetning och sammanställning (Vetenskapsrådet 2018: 12). Min enkla tolkning är att det betyder att det mesta som genereras i en forskningsprocess skulle kunna betraktas som forskningsdata.

I regeringsuppdraget till Vetenskapsrådet och i regeringens forskningsproposition uttrycks målet som "att en omställning till öppen tillgång till forskningsresultat [min kursivering] inklusive relaterade vetenskapliga publikationer, konstnärliga verk och forskningsdata bör vara genomförd fullt ut senast inom tio år" (Utbildningsdepartementet 2016), vilket är 2026. Vän av ordning undrar om resultat innefattar fler komponenter än publikationer, konstnärliga verk och data och i så fall vad? En ledtråd hittar vi i regeringens "Nationella färdplan för det europeiska

forskningsområdet 2019–2020" (Utbildningsdepartementet 2019). Här beskrivs att en grundförutsättning för ett öppet vetenskapssystem är öppen tillgång till vetenskapliga publikationer och forskningsdata, men att det även finns fler komponenter i ett sådant system. Målet är "att fler delar av forskningsprocessen så snart som möjligt kan samverka med det omgivande samhället. De delar som avses är exempelvis öppna forskningsdata, öppet tillgängliga publikationer, öppen källkod, utvärderingsmetoder som tar hänsyn till samhällsnytta, öppna lärresurser och medborgarforskning" (Utbildningsdepartementet 2019: 17). Detta med samhällsnytta i utvärderingsmetoder känns igen från den pågående utvärderingen av statsvetenskaplig forskning.

En annan kärnfråga är *när* forskningsdata bör göras tillgängligt. I regeringsuppdraget till Vetenskapsrådet framgår att forskningsdata som ligger till grund för vetenskaplig publikation bör bli öppna samtidigt som publikationen (Utbildningsdepartementet 2017). Samtidigt betonas att även forskningsdata som *inte* ligger till grund för publikation också bör göras öppet tillgängliga eftersom det finns en outnyttjad potential i den här typen av data som skulle kunna leda kunskapsläget framåt (Vetenskapsrådet 2018: 27). Med den utgångspunkten blir det betydligt svårare att avgöra vilken typ av data som ska göras öppen och när.

## FAIR: AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

I början av 2013 arrangerades en tredagars workshop vid det ansedda Lorentz Center vid universitetet i Leiden. På deltagarlistan fanns ett trettiootal personer från forskningsinstitut, universitet och förlag. Det blev den officiella startpunkten för FAIR, "a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the FAIR data principles" vars vision är "to optimise data sharing and reuse by humans and machines" (FAIR 2021). FAIR står för findable, accesible, interoperable och reusable. Initiativet lyckades

uppenbarligen vinna europeiska makthavares gillande eftersom FAIR-principerna inkorporerades i Horizon 2020. Principerna har fått oerhört stort genomslag. En googling på sökorden "forskningsdata" och "research data" resulterar i sökträffar med nästintill identiska formuleringar och ordalydelser vilka sammanfaller med FAIR.

*Findable* (F) betyder att data ska kunna hittas av andra på internet. En grundförutsättning är att data måste vara i digital form. Det enda som explicit undantas som forskningsdata är fysiska föremål. Som exempel ges arkeologiska objekt. Målsättningen är att data ska digitaliseras så långt det är möjligt. Om det inte går ska åtminstone beskrivande data om data (metadata) vara i digital form. Här kommer *acesible* (A) in i bilden. De som liksom undertecknad brukar bänka sig framför tv-programmet Antikrundan har vid det här laget förstått värdet av ett föremåls proviniens. Denna visar sig vara a och o även i FAIR sammanhang. För att andra ska kunna återanvända ett datamaterial krävs en gedigen och strukturerad beskrivning av detsamma. Denna beskrivning kallas metadata och "kan bestå av beskrivande eller kontextuell information eller information om datas ursprung (provinien), såsom hur variabler är definierade, hur data är strukturerat, hur olika datamängder relaterar till varandra, information om hur och av vem data har tagits fram eller om särskild mjukvara behövs för att läsa eller analysera data" (Vetenskapsrådet 2018, 12). Metadata är en form av dokumentation som måste vara strukturerad på ett specifikt, systematiskt sätt för att kunna läsas av både människor och datorer (snd.gu.se). För att metadata ska kunna läsas maskinellt måste etablerade namnkonventioner och internationella standards användas. Jag kan alltså inte beskriva med ord och meningar hur och varför data genererats, av vem och i vilken kontext. *Interoperabilitet* (I) är ett begrepp som används för att beskriva hur olika system eller datamängder "kommunicerar" med varandra. För att data ska kunna delas och användas av andra måste det finnas en teknisk komparabilitet.

Vetenskapsrådet skriver att "digitala objekt som publiceras på nätet ska vara beständiga och sammanlänkingsbara med andra typer av digitala objekt" (Vetenskapsrådet 2018: 23). Slutmålet är att data ska kunna delas och användas av andra, vara *reusable* (R). En förutsättning för detta är att de andra principerna uppfylls. Det krävs också en infrastruktur som gör delning möjlig och säker och olika molntjänster är därför under uppbyggnad.

Att dessa principer utarbetats i en vetenskaplig miljö som präglas av en positivistisk kunskapssyn är uppenbart. Grundtanken är att data går att kombinera och återanvända oberoende av kontext och ursprungligt syfte. Det beskrivs i termer av att *hitta ny* kunskap och att göra *nya upptäckter* i redan insamlade data, inte om att konstruera kunskap i sammanhang och samspel. *Data harvesting* är ett återkommande uttryck i sammanhanget. Det finns kunskap som bara väntar på att bli *upptäckt* bara forskare delar med sig och bjuder in andra att skörda. I styrdokumenterna finns skrivningar av typen *optimal* återanvändning av data (Council of the European Union 2016, 9) och *optimal* kunskapsdelning (Council of the European Union 2016: 5). *Rätt* kombination av *rätt* data antas ge *optimala* lösningar. Bara begreppet *data* synliggör vilken vetenskaplig tradition som tillåts att bli normerande. Ett annat sätt att beskriva det materiella objektet hade kunnat vara att använda begreppet forskningsmaterial som i likhet med data ligger till grund för analys i en forskningsprocess, men data är det självklara begreppet inom de discipliner som utgör normen för processen. Det kan emellertid också tolkas i ett annat ljus, nämligen att det som avsågs var data i dess strikta bemärkelse. I rådsslutsatserna framgår att det handlar om "political support for setting the right framework conditions for research and innovation focused on big data and for strengthening the whole data value chain Europe (Council of the European Union 2016: 3). Någonstans på vägen har dock den här processen kommit att omfatta, och bli normerande för mer än enbart big data.

### VARFÖR OPEN SCIENCE?

I de styrdokument som jag har tagit del av motiveras öppen vetenskap med olika typer av argument. Ett är att öppen vetenskap leder till "ökad kvalitet i den vetenskapliga produktionen" (Vetenskapsrådet 2018: 11). Ett annat är att det förväntas leda till "användning av data inom nya områden och på nya sätt" (Vetenskapsrådet 2020: 17). Vari den kvalitetsökningen består är oklart liksom vilka dessa nya områden och sätt är. Kanske kan argumentet sammanfattas med att forskningens slutresultat antas bli bättre om processen är transparent och öppen för fler än en liten, sluten grupp av forskare. Transparens och öppenhet i forskningsprocessen är återkommande argument. På sin hemsida skriver Vetenskapsrådet att "Öppen vetenskap syftar till att göra den vetenskapliga processen mer öppen och inkluderande. Med öppen tillgång till vetenskaplig information kan forskningens resultat användas av andra- både inom och utanför forskarsamhället. Det gynnar både forskningen och ett kunskapsintensivt samhälle." (Vetenskapsrådet 2021). Transparens kan betraktas som ett självändamål men också som ett medel för att uppnå andra (ekonomiska) mål. I regeringens nationella färdplan för det europeiska forskningsområdet 2019–2020" beskrivs öppen vetenskap som en "demokratifråga där hela samhället kan få del av, använda och delta i den forskning som bedrivs med offentliga medel" (Utbildningsdepartementet 2019: 17).

Det finns en start tilltro att open science kommer att leda till "knowledge discovery" (OpenAIRE 2021) som i slutändan också genererar innovationer. Dessa antas i sin tur ge Europa fördelar i den globala konkurrensen. Bara det faktum att EU-kommissionären ansvarar för forskning och innovation synliggör att dessa två områden betraktas som naturligt integrerade. Ett ytterligare argument är att open science handlar om att få valuta för satsade pengar, eller som ministerrådet uttrycker det i sina slutsatser, "boosting

the benefits of public investment in research" (Council of the European Union 2016). Ett likartat argument är att forskning som finansierats med offentliga medel bör nyttiggöras i så hög utsträckning som möjligt och att offentligt finansierad forskning bör betraktas som "public good" (Council of the European Union 2016: 8).

### PROBLEM SOM SÖKER LÖSNINGAR ELLER LÖSNINGAR SOM SÖKER PROBLEM?

I Vetenskapsrådets rapport (2020: 13) beskrivs på ett träffsäkert sätt hur den snabba tekniska utvecklingen förändrar förväntningarna på hur forskning ska och bör bedrivs och hur kunskap ska och bör tillgängliggöras. Teknikutvecklingen har gjort det möjligt att kombinera enormt stora datamängder. Visst finns det problem som söker efterlängtade lösningar i detta avancerade datakeri, men nog handlar det till viss del också om att det faktiskt *går* att göra. Fiffiga tekniska lösningar som kan hantera stora datamängder söker problem. Men vad händer när forskningsfrågorna formuleras utifrån vilken typ av data som finns och som går att kombinera? Vilka frågor är det som ställs och vilka kommer att lämnas obesvarade? Ett återkommande uttryck i de europeiska styrdokumenterna är data driven science. Det innefattar mycket mer än att data driver på forskning, mer än att bara kombinera datamängder för att, som det ibland heter i dessa sammanhang, "upptäcka" (se t.ex. Wilkinson 2016) nya kunskaper. Data-driven forskning betyder att maskiner inte bara analyserar datamängder utan också formulerar forskningsfrågorna. Artificiell intelligens med andra ord. Maskiner som överblickar och analyserar data på en nanosekund kanske ställer smartare, mer relevanta eller mer precisa frågor än mänskliga forskare? Den snöboll som är i rullning väcker många forskningsetiska frågor och i de viktiga samtalen måste det finnas även samhällsvetare och humanister, inte bara naturvetare, teknologer och medicinare.

## **GOD DATAHANTERING OCH DATAHANTERINGSPLANER**

EU:s medlemsländer bör i enlighet med beslutet i ministerrådet följa FAIR-principerna i forskningsprogram och finansiering (Utbildningsdepartementet 2017). Vetenskapsrådet fick i uppdrag av regeringen att utarbeta kriterier utifrån FAIR-principerna vilka skulle kunna vara allmängiltiga, teknikoberoende och beständiga över tid (Vetenskapsrådet 2018: 22). Jag avundas inte någon det minst sagt komplexa uppdraget. Vetenskapsrådet konstaterar att ”god datahantering redan i planeringsfasen är en nyckelkomponent för både öppen tillgång till och bevarandet av forskningsdata och en grundsten för FAIR (Vetenskapsrådet 2020: 23). Vetenskapsrådet tillsatte också en arbetsgrupp som fick i uppdrag att utarbeta en mall för datahanteringsplaner. Den mall som utarbetades utgår från den mall som Science Europe utarbetat utifrån FAIR-principerna och från och med 2019 ska alla som beviljas bidrag från Vetenskapsrådet utarbeta en datahanteringsplan (Vetenskapsrådet 2021).

God datahantering kan tyckas lovvärt men vad betyder det? Vad är motsatsen? Dålig datahantering, som när data från enskilda respondenter går att identifiera? När ifyllda enkäter ligger huller om buller på någons skrivbord? När obearbetade ljudfiler från intervjuer ligger öppet sökbara på internet? När observationsprotokoll hamnar i orätta händer för att en laptop blivit stulen? När forskningsarkivet är pappersåtervinningen? Det här skulle kunna vara exempel på dålig datahantering men det är sannolikt inte den här typen av oönskade scenarion som varit styrande i denna process. FAIR-principerna utgör normen för vad som kan betraktas som data (och inte) och vad som följaktligen är god datahantering. I det här sammanhanget handlar god datahantering om att planera hanteringen av data på ett sådant sätt att det går att öppet dela med andra som ska kunna återanvända den. Även lärosäten har börjat tillämpa datahanteringsplaner. Att ha en plan

för hur data ska insamlas, hanteras, förvaras och arkiveras förefaller som en klok strategi men det är inte hantering i allmänhet utan hantering för ett visst syfte. Vad betyder det för forskningsprocessen? Hur pass styrande blir dessa planer? Hur stora avsteg från planen får, bör, vågar enskilda forskare göra?

## **ETT FÖRÄNDRAT VETENSKAPSSYSTEM – (O)ÖNSKADE KONSEKVENSER?**

De krafter som driver open science vill åstadkomma genomgripande förändringar. Det beskrivs som att det handlar om ”change the ways of doing science” (Council of the European Union 2016: 6). I Vetenskapsrådets rapporter beskrivs det i termer av en normförändring. Fler delar av forskningsprocessen ska öppnas upp för samverkan, inte bara tillgången till publiceringen. Fler ska få tillgång till forskningsresultat och kunna ta aktiva roller i forskningsprocessen, inte bara forskare utan också intresserade medborgare. En av huvudfrågorna är just vem eller vilka som har rätt att göra anspråk på data och när? En komponent i open science är medborgarforskning som betyder att medborgare är medskapare i forskningsprocessen. Allt detta klingar vackert och vällovligt men innebär också enorma utmaningar och ställer rådande normer och vedertagna konventioner på ända. Hur många informanter/respondenter kommer att gå med på att bli studieobjekt när de förstår att deras livsberättelser och upplevelser ska kunna ”återanvändas” av andra? Vad händer med professionen när alla kan vara forskare? Vad blir det för forskningsresultat när ”data” kan frikopplas från det sammanhang inom vilket det ursprungligen genererades?

Vi kommer sannolikt att få se ett reformerat meriteringssystem där forskare som delar data med andra kommer att premieras. Ska den typen av meriteringskriterier gälla inom alla ämnesområden oavsett disciplinernas mycket varierade förutsättningar? Vetenskapsrådet konstaterar förvisso att förutsättningarna att uppfylla FAIR skiljer sig åt mellan olika forskningsområden och datatyper och att det inte är rimligt att all

forskningsdata alltid ska göras FAIR (Vetenskapsrådet 2018: 28). De konstaterar också att hänsyn alltid måste tas till gällande lagstiftning och till om data innehåller känsliga personuppgifter eller sekretessbelagda uppgifter men att då bör åtminstone metadata vara FAIR (Vetenskapsrådet 2020: 22). Jag kämpar fortfarande med att förstå hur forskningsmaterial som inte är hårda data och som inte går att förpacka i någon internationellt vedertagen källkod ska kunna passa in i FAIR-formatet. Hur ska till exempel konstnärliga verk som genereras i forskning kunna delas och återanvändas av andra och är det ens önskvärt? Frågorna är än så länge långt fler än svaren, men de forskningsetiska frågeställningar som denna process aktualiserar måste tas på allvar och samhällsvetenskapliga och humanistiska perspektiv måste inkluderas.

Sedan har vi de mindre allvarliga men för en enskild forskare nog så tröttsamma konsekvenserna av den här typen av processer. De tenderar nämligen att resultera i att en mängd nya administrativa processer och system måste tillskapas och sedemera också tillämpas. Det finns mallar som ska fyllas i, stödmaterial för att fylla i nämnda mallar och stödmaterial för stödmaterialet för att fylla i mallarna. Alla ska vi fylla i samma rutor oavsett om forskningen handlar om estetiska läroprocesser, malaria eller regeringsbildningar. Digitaliseringen har dessutom gjort det enklare än någonsin att skapa denna typ av administrativa processer och system och därmed är vi tillbaka där den här texten började.

## KÄLLOR

Coalition-S, 2021. *What is Coalition-S?* <https://www.coalition-s.org/about/>. Hämtad 2021-06-01.

Council of the European Union, 2016. *The transition towards an Open Science system – Council conclusions*, Adopted on 27/05/2016. 9526/16.

Enserink, M., 2016. "E.U. urged to free all scientific papers by 2020", *Sciencemag*, 2016-04-14. DOI 10.1126/science.aaf9911. <https://www.sciencemag.org/news/2016/04/eu-urged-free-all-scientific-papers-2020>. Hämtad 2021-06-01.

Enserink, M., 2016. "In dramatic statement, European leaders call for 'immediate' open access to all scientific papers by 2020", *Sciencemag*, 2016-05-27. DOI 10.1126/science.aag0577. <https://www.sciencemag.org/news/2016/05/dramatic-statement-european-leaders-call-immediate-open-access-all-scientific-papers>. Hämtad 2021-06-01.

European Commission, Directorate-General for Research & Innovation, 2016. *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*. Version 3.0. 26 July 2016.

FAIR, 2021. *Go FAIR Initiative*. <https://www.go-fair.org/go-fair-initiative/>. Hämtad 2021-06-01.

OpenAIRE, 2021. *Guides for researchers. How to make your data FAIR*. <https://www.openaire.eu/how-to-make-your-data-fair>. Hämtad 2021-06-01.

Svensk nationell datatjänst, 2021. *Hantera data*. <https://snd.gu.se/sv/hantera-data/fardigstallatillgangliggöra/FAIR-principerna>. Hämtad 2021-06-01.

Utbildningsdepartementet, 2017. Regeringsbeslut. 2017-12-21. U2017/05169/F. *Uppdrag att ta fram bedömningskriterier för att kunna följa utvecklingen mot ett öppet vetenskapssystem*.

Utbildningsdepartementet, 2019. Bilaga till regeringsbeslut 2019-04-18. *Nationell färdplan för det europeiska forskningsområdet 2019–2020*.

Vetenskapsrådet, 2018. *Kriterier för FAIR forskningsdata. Redovisning av regeringsuppdrag att ta fram bedömningskriterier för att följa vägen mot ett öppet vetenskapssystem*.

Vetenskapsrådet, 2020. *Samordning av öppen tillgång till forskningsdata. Statusrapport i Vetenskapsrådets uppdrag – summering av arbetet 2017–2019 och fortsatt arbete*.

Vetenskapsrådet, 2021. *Söka finansiering. Ta fram en datahanteringsplan*. <https://www.vr.se/soka-finansiering/krav-och-villkor/ta-fram-en-datahanteringsplan.html>. Hämtad 2021-06-01.

Vetenskapsrådet, 2021. *Öppen vetenskap*. <https://www.vr.se/uppdrag/oppen-vetenskap.html>. Hämtad 2021-06-01.

Wilkinson, M.D. et al., 2016. "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data* 3. Article number: 160018. DOI 10.1038/sdata.2016.18. <https://www.nature.com/articles/sdata201618>.

Katarina Roos är verksam vid Statsvetenskapliga institutionen, Umeå universitet.

E-post: katarina.roos@umu.se