

Spårbara forskningsdata: metoder och rekommendationer för ökad spårbarhet och synlighet för svenska datapubliceringar

Översiktsrapport

Urban Andersson (Chalmers) 

Lina Andrén (KTH) 

André Jernung (SND) 

Stina Johansson (GU) 

Olof Olsson (SND) 

Therese Tikkanen (Chalmers) 

Mattias Vesterlund (KTH) 



Versionshistorik

Version	Datum	DOI	Kommentarer/ändringar
1.0	2025-10-21	10.5281/zenodo.17365330	Första utgåvan

Innehållsförteckning

Sammanfattning	5
Inledning	7
Avgränsningar	7
Sammanfattning av bakgrund	8
Bakgrund	9
Tekniska förutsättningar	11
Sammanfattning av tekniska förutsättningar	11
Beskrivning av tekniska förutsättningar	12
Stödorganisation	19
Sammanfattning av stödorganisation	19
Beskrivning av stödorganisation	20
Tillgängliggörande och publicering	23
Sammanfattning av tillgängliggörande och publicering	23
Beskrivning av tillgängliggörande och publicering	24
Bevarande och arkivering	30
Tidigare studier kring datapublicering	31
Studier kring forskningsdatapublicering i Sverige	31
Internationella studier kring forskningsdatapublicering	35
Kartläggning	43
Metod	43
Analys och diskussion	46
Slutsatser	67
Definitioner	73
Källförteckning	77

Sammanfattning

Denna översiktsrapport är en del av ett arbete för att skapa metoder och rekommendationer som kan uppnå ökad spårbarhet och synlighet för forskningsdata publicerade vid svenska lärosäten och andra forskningsorganisationer.

Rapporten kartlägger nuläget genom att redovisa tekniska förutsättningar, stödorganisationer och processer för tillgängliggörande. Detta inkluderar en genomgång av några tidigare studier av svensk och internationell datapublicering samt resultatet av en enkätundersökning riktad till stödpersonal på svenska forskningsorganisationer.

Resultaten visar att även om de tekniska förutsättningarna är goda – med etablerade standarder för strukturerade, maskinläsbara metadata och beständiga identifierare (PID-system) som DOI, ORCID och ROR – är implementeringen ofta inkonsekvent hos både infrastrukturer och användare. Brister i metadatakvaliteten, tvetydig affilieringsinformation och avsaknad av väletablerade, systematiska rutiner för rapportering hos forskningsorganisationerna utgör stora hinder för effektiv spårbarhet. Trots att uttalade ambitioner för att ställa om till öppna vetenskapspraktiker funnits under många år räknas datapubliceringar sällan med som värdefulla forskningsprodukter i meriteringssystemen, vilket ytterligare försvårar spårbarhet och uppföljning.

Rapporten drar slutsatsen att en kombination av tekniska, organisatoriska och samordnande åtgärder är nödvändiga för att säkerställa att publicerade forskningsdata kan hanteras som självständigt citerbara forskningsprodukter och spåras. I Sverige krävs ett fortsatt arbete för att förbättra kvaliteten på metadata och förstärka stödstrukturerna kring datapubliceringar för att möta både nationella krav och internationella initiativ för öppen vetenskap.

Inledning

Denna översiktsrapport utgör en del av Svensk nationell datatjänsts flaggskeppsprojekt "Spårbara forskningsdata: metoder och rekommendationer för ökad spårbarhet och synlighet för svenska datapubliceringar". Projektets övergripande mål är att utveckla och tillhandahålla verktyg, metoder och rekommendationer som underlättar insamling av metadata om datapubliceringar. Genom att effektivisera processerna för att hitta, registrera och följa upp datapubliceringar, såväl lokalt som på nationell nivå, syftar projektet till att skapa bättre förutsättningar för att uppfylla FAIR-principerna¹ och nationella krav på forskningsdata och öppen vetenskap.

En central målsättning är att forskningsdata som publiceras i större utsträckning ska kunna följas upp och citeras. Att metadata om datapubliceringar kan synliggöras för meriterande syften, och användas i metaanalyser av forskningsfält, skapar långsiktigt stöd för den breddning av forskningsutvärdering som förespråkas av olika initiativ för öppen vetenskap. Denna synliggörandeprocess är avsedd att fungera på ett sätt som liknar hur metadata för traditionella vetenskapliga resultatpublikationer hanteras, men den bör också kunna dra nytta av de många nya verktyg och standarder som växt fram i det moderna öppna vetenskapslandskapet.

Rapporten ger en översikt av förutsättningarna för spårbar datapublicering genom att redovisa befintliga tekniska möjligheter, stödorganisationer och processer för tillgängliggörande. För att ge en bred och relevant bild av nuläget för svensk datapublicering redovisas dels en genomgång av några tidigare studier och kartläggningar, dels resultatet av en enkätundersökning riktad till stödpersonal på svenska forskningsorganisationer. Rapporten belyser särskilt frågorna: *I vilken utsträckning är datapubliceringar FAIR idag?* och *Vilka möjligheter och hinder finns för en effektiv och högkvalitativ uppföljning av publicerade data?*

Avgränsningar

Denna rapport omfattar i huvudsak spårbarhet hos forskningsdata som är tillgängliggjorda av forskare knutna till svenska lärosäten och offentligt finansierade forskningsorganisationer. Den omfattar inte hanteringen av forskningsdata hos privata forskningsaktörer. Vidare omfattar den enbart forskningsdata som har publicerats, eller kommer att publiceras, med öppen eller begränsad tillgång².

¹ Akronymen FAIR står för *Findable* (hittbart), *Accessible* (tillgängligt), *Interoperable* (interoperabelt) and *Reusable* (återanvändningsbart) och syftar på ett antal tekniska kriterier som först definierades av Wilkinson m.fl. [1]. Se även *Definitioner*.

² Datapubliceringar kan utgöras av öppna data, som är fritt nedladdningsbara, eller åtkomstbegränsade data, som är öppet beskrivna men kräver prövning innan datafiler kan lämnas ut. Se även *Definitioner*.

Sammanfattning av bakgrund

Svenska forskningsorganisationer har med anledning av svenska forskningspolitiska målsättningar kring öppen vetenskap och direkta krav med ursprung i öppna data-direktivet behov av att:

- Kontinuerligt spåra och inhämta metadata om nya och existerande datapubliceringar med koppling till organisationens forskare och forskningsaktiviteter.
- Etablera metadataregister över sådana datapubliceringar i sin förvaltning.
- Peka ut särskilt värdefulla datamängder i dessa metadata, med stöd av öppna data-direktivets definitioner.
- Överföra metadata om datapubliceringar till ansvarig myndighet, i dagsläget Digg – Myndigheten för digital förvaltning.
- Säkerställa att metadata om datapubliceringar tillgängliggörs på ett kontrollerat sätt som uppfyller relevanta lagkrav och utförs i linje med FAIR-principerna.
- Genom ovanstående åtgärder möjliggöra återanvändning, uppföljning och analys av datapubliceringar.

Bakgrund

Forskningsdata har ofta behandlats som något som är väsensskilt från andra produkter som skapas i forskningsprocesser. Det sätt som svenska forskningsorganisationer väljer att förhålla sig till forskningsdata på, och de förväntningar som finns på svenska forskare vad gäller publicerade data, tyder på att bilden av datapubliceringar som ett udda och till stor del oreglerat fenomen lever kvar.

Vem publicerar forskningsdata, och var hör den hemma?

Vid svenska lärosäten och andra forskningsorganisationer saknas ofta väletablerade arbetssätt för att publicera forskningsdata och koppla datapubliceringen till respektive organisation. Exempelvis finns inte alltid en vana att tydligt knyta datapubliceringen till respektive lärosäte med affilieringsinformation på samma sätt som forskaren skulle göra i en tidskriftsartikel. Publicerade forskningsdata med metadata som korrekt och otvetydigt pekar ut affiliering till forskares lärosäte eller organisation utgör sannolikt en begränsad del av det totala antalet datapubliceringar.

Sättet som data hänvisas till eller citeras på i annat material kan också variera stort beroende på situation. Yttre faktorer som riktlinjer från förlag, citeringsregler eller tekniska lösningar vid publicering kan påverka spridningen av metadata och information om datapubliceringen. Kultur och praxis inom den aktuella disciplinen kan avgöra hur hänvisningar till data görs och därmed påverka spårbarheten. Ofta saknas krav på att kontinuerligt rapportera in datapubliceringar till forskningsorganisationen eller finansören på samma sätt som övriga publikationer, såsom artiklar, böcker, bokkapitel, konferensbidrag osv. Utöver detta räknas forskningsdata sällan med som värdefulla forskningsprodukter i ställningstaganden kring meritering eller medelsfördelning. En stor mängd datapubliceringar kommer därför sannolikt aldrig till forskningsorganisationernas direkta kännedom.

Varför hålla reda på datapubliceringar?

Att publicerade forskningsdata lätt kan hittas och knytas till en forskningsorganisation har betydelse för lärosätens och andra forskningsorganisationers möjligheter att följa upp och rapportera forskningsdata enligt nationella och internationella krav. Det har också betydelse för synligheten för forskningsdata som fristående forskningsprodukter och forskarnas möjligheter att meriteras för sitt arbete med att ta fram, beskriva och tillgängliggöra data.

Det finns lagkrav på att svenska lärosäten ska redovisa och vidarerapportera datapubliceringar som skapats av forskare verksamma vid respektive organisation. Enligt öppna datalagen (SFS 2022:818) ska myndigheter upprätthålla en löpande förteckning över data som tillgängliggörs och dela denna med en utpekad myndighet, för närvarande Myndigheten för digital förvaltning (Digg) [2]. För detta ändamål har Digg skapat *Sveriges dataportal*, där alla inrapporterade metadata om datapubliceringar och andra myndighetsdata går att nå [3], [4]. Med öppna datalagen introducerades också krav på öppet tillgängliggörande och rapportering av myndigheters särskilt värdefulla datamängder, ett

begrepp som omfattar kategorier som exempelvis statistik och geodata, vilka ofta insamlas som forskningsdata [5].

Att forskningsdata ska tillgängliggöras i linje med FAIR-principerna är ett växande krav, drivet av politiska beslut, nationella och lokala riktlinjer och rekommendationer samt uttalade krav från finansiärer och förlag. Några viktiga FAIR-aspekter är exempelvis att data ska vara beskrivna med fristående maskinläsbara metadata, och att dessa ska innehålla detaljerad proveniens för hur datamaterialet har skapats och vem som skapat det [6, s. 7, s. 19].

I maj 2023 antog EU:s medlemsstater rådsslutsatser som understryker vikten av högkvalitativ, transparent, öppen och rättvis vetenskaplig publicering, där forskningsdata och metadata bör vara omedelbart och fritt tillgängliga med öppna licenser, i enlighet med maximen "så öppet som möjligt, så begränsat som nödvändigt" [7]. Detta understryker tidigare rådsslutsatser om öppen vetenskap från 2016 [8]. Slutsatserna kopplar direkt till behovet av att säkerställa att forskningsdata beskrivs med korrekta affilieringar och metadata, så att de kan knytas till forskningsorganisationer och finansiärer.

I den svenska forsknings- och innovationspropositionen 2024, "Forskning och innovation för framtid, nyfikenhet och nytta" (prop. 2024/25:60) fastställer regeringen att vetenskapliga publikationer och forskningsdata omedelbart ska göras öppet tillgängliga när så är möjligt, och att *omställningen till öppen tillgång (...) ska vara genomförd senast 2026* [9, s. 42]. Detta återkopplar till ett omställningsarbete för öppen vetenskap som efterfrågats i såväl svensk som europeisk forskningspolitisk nivå under många år. Vidare avser propositionen att Sverige ska fortsätta, och i viss mån öka, fördelningen av forskningsmedel via bibliometriska indikatorer, vilket innebär att citeringar fortsatt kommer att vara en viktig faktor vid tilldelning av anslag.

Samtidigt lyfts initiativ som CoARA (Coalition for Advancing Research Assessment) fram, vilka syftar till att bredda synen på vad som är meriterande inom forskning och förespråkar utvärdering av andra aktiviteter än enbart kvantitativa mått för resultatpublikationer och deras citeringar, däribland tillgängliggörande av forskningsdata [10]. Detta ligger i linje med andra internationella initiativ som *Barcelona-deklarationen för öppen forskningsinformation* och *DORA* (San Francisco Declaration on Research Assessment), vilka förespråkar ett mer nyanserat sätt att bedöma forskning, där även öppna vetenskapspraktiker och forskningsdata kan inkluderas i meriteringssystemen [11], [12]. Initiativen förutsätter ett systematiskt och öppet tillgängliggörande av forskningsinformation som beskriver forskningen och hur den hänger samman, vid sidan av öppna resultatpublikationer och öppna forskningsdata. Exempel på sådan öppen forskningsinformation är strukturerade, kvalitetskontrollerade och sammanlänkade metadata om forskningsprodukter som kan sammanställas och jämföras maskinellt.

En grundläggande förutsättning för att målsättningarna ovan ska kunna nås är etablerandet av välfungerande infrastrukturer och praktiker i Sverige för att registrera och spåra publicerade forskningsdata.

Tekniska förutsättningar

Sammanfattning av tekniska förutsättningar

- Strukturerade metadata som görs öppet tillgängliga och följer vedertagna metadatastandarder är avgörande för att omvärlden effektivt ska kunna förhålla sig till forskningsdata när datapubliceringar skapas och beskrivs.
- Beständiga identifierare (PID) som pekar ut datapubliceringar och andra digitala objekt kan användas i många former av forskningskommunikation, och är strategiskt viktiga för att publicera, peka ut och beskriva spårbara forskningsdata.
- Beständiga identifierare som innehåller fristående så kallade kärnmetadata, exempelvis DOI från DataCite, är särskilt viktiga för datapubliceringars spårbarhet.
- Olika system för beständiga identifierare kan hänvisa till många olika fenomen utöver forskningsmaterial som publiceras, exempelvis personer (ORCID), organisationer (ROR) eller projekt/aktiviteter (RAiD).
- Kontrollerade vokabulärer med uppslagbara auktoritetskällor är viktiga verktyg för att hänvisa till klassificeringar, ämnesord och andra fastslagna definitioner som kan beskriva datapubliceringar och göra dem sökbara och jämförbara.
- Affilieringsmetadata är sällan standardiserade, men tvetydigheter går att förebygga med konsekvent användning av beständiga identifierare som ORCID och ROR. Då affilieringsmetadata inte är standardiserade blir det nödvändigt att förvalta listor över olika möjliga varianter av organisations- och institutionsnamn som stöd vid sökningar.
- Formella relationer mellan olika objekt i forskningsekosystemet, exempelvis att A bygger på B, kan anges i flera metadatastandarder och system för beständiga identifierare.
- Forskningsorganisationer använder ibland CRIS-system för att förvalta information om forskningsaktiviteter och forskningsprodukter, och dessa kan ofta förteckna och dela öppen forskningsinformation om hela forskningsaktiviteten och dess sammanhang.
- Infrastrukturer som samlar in öppen forskningsinformation kan finnas både på nationell och internationell nivå, och har gradvis blivit mer betydelsefulla för att kunna spåra datapubliceringar och deras sammanhang vid sidan om proprietära informationskällor.
- Kvalitetssäkring av datapubliceringar och deras metadata kräver en god förståelse för hur den information som anges påverkar forskningsekosystemet, och ett systemstöd som i hög grad lotsar skaparen rätt, för att inte skapa hinder för spårbarheten.
- Olika typer av automatiserad utvärdering och validering kan användas för att få vägledning om vissa kriterier för hur en given datapublicering kan göras mer spårbar.

Beskrivning av tekniska förutsättningar

I arbetet med att följa upp datapubliceringar behöver svenska forskningsorganisationer förhålla sig till tekniska förutsättningar i forskningslandskapet, framför allt de som relaterar till sätt som öppen forskningsinformation³ skapas och tillgängliggörs på.

Strukturerade metadata

Olika forskningsprodukter, som tillgängliggjorda forskningsdata, kan ha sammanhörande metadata av många olika slag. I en bredare bemärkelse kan metadatabegreppet sägas kunna omfatta alla typer av relaterade beskrivningar eller dokumentation om materialet och dess proveniens.

Det är dock viktigt att särskilja metadata som huvudsakligen produceras för mänsklig läsning och manuell tolkning, som en narrativ dokumentationsfil, från strukturerade metadata som produceras enligt etablerade metadatastandarder. Sådana metadata tillgängliggörs i form av standardiserade filformat, inbäddade uppmärkningselement, öppna API:er eller tekniker som *signposting* på webbsidor [13]. De följer fastställda metadatascheman som kan tolkas maskinellt av många olika aktörer och tjänster, t ex det vanligt förekommande schemat Schema.org som ofta bäddas in på webbplatser [14].

Den strukturerade typen av metadata är alltså avsedd att kunna läsas in, tolkas, jämföras och sorteras på maskinell väg, samtidigt som ett snarlikt informationsinnehåll ofta också kan presenteras för mänsklig läsning i exempelvis ett webbgränssnitt. Det är också sådana strukturerade metadata som huvudsakligen delas mellan olika infrastrukturer, som databaser eller aggregatorer för forskningsinformation, och gör forskningsdata och andra forskningsprodukter sökbara och klassificerbara. Det medför också att brister i dessa metadata kan skapa stora följder för spårbarhet och återanvändning [15]. En ansvarsfull hantering av forskningsinformation ställer därför höga krav på konsekventa och kvalitetssäkrade processer för hur metadata blir till.

Beständiga identifierare och auktoritetskällor

Beständiga identifierare (ofta benämnda PID, *persistent identifiers*) fungerar framför allt som robusta länkar till material online, till exempel en artikel eller en datapublicering, som inte slutar fungera utan kan underhållas över tid. De enklaste PID-systemen fyller just en sådan funktion och inte mycket mer: den beständiga identifieraren pekar vidare till ett länkmål i form av en annan webbadress, som i sin tur kan bytas ut om materialet eventuellt flyttas. Enkla identifierare av denna typ erbjuds exempelvis av Handle-systemet, som också utgör en basinfrastruktur för många andra PID-system [16], eller av National Bibliography Number (URN:NBN) som bland annat används i det svenska DiVA-systemet [17].

³ Öppen forskningsinformation är sådan information som berättar om hur en forskningskontext hänger samman: vad det forskas om, vem som gör det, hur det går till, vilka resurser som finns och vad som skapats i form av resultat, data och andra adresserbara forskningsprodukter. Se även *Definitioner*.

Mer utvecklade PID-system erbjuder utöver själva vidarelänkningen till målet möjligheter att ange utförlig information i en kärna av metadatainnehåll för varje beständig identifierare, som lagras hos själva PID-tjänsten och kan avläsas oberoende av vad som finns i länkmålet. Dessa kärnmetadata (eng. *kernel metadata*) kan utgöras av sådant som materialtyper, upphovspersoner, affilieringar eller andra relationer. När sådana metadata används för att skapa öppen forskningsinformation kan de återanvändas till många olika syften såsom sökfiltrering, referenshantering, kartläggning av forskningsverksamhet eller bibliometriska analyser. Bland beständiga identifierare som fungerar på detta sätt bör särskilt nämnas DOI-systemet⁴, de två större aktörerna som levererar DOI-identifierare (Crossref och DataCite), och de unika DOI-metadatascheman som dessa i sin tur förvaltar [18], [19].

Att beständiga identifierare används för att uttrycka olika fenomen som relaterar till forskningsprocessen och på så sätt göra dem spårbara över tid har kommit att ses som ett viktigt framsteg för öppen, maskinläsbar och återanvändbar forskningsinformation, och som en grundstomme i det öppna vetenskapssystemet [20]. PID-system används dock inte bara för att peka ut forskningsprodukter, utan också för att ange robusta hänvisningar till många andra typer av kontrollerad information vars officiella definitioner finns på en annan plats.

I samband med att beständiga identifierare för forskningsändamål har vuxit fram har det blivit vanligt att PID-system används för att hänvisa till auktoritetskällor som beskriver unika fenomen såsom forskaridentiteter, forskningsorganisationer eller finansiärer. Detta har stora fördelar framför att endast mata in text på frihand. Till exempel kan ett person- eller organisationsnamn ha många olika variationer, förändras över tid och vara svårt att tolka eller jämföra för automatiska processer eller vid utökningar. Arbetssätten kompletterar andra internationella källor, såsom de system för att uttrycka auktoritetsinformation som sedan tidigare används inom biblioteksverksamhet. Viktiga exempel på denna typ av beständiga identifierare är ORCID⁵ för personer och ROR⁶ för organisationer [21].

Användningsområden för beständiga identifierare har efter hand dykt upp för allt fler adresserbara fenomen i forskningslandskapet [22]. Kontrollerade vokabulärer, såsom formella ämnesordssystem och ontologier, möjliggör ytterligare återanvändning av maskinläsbara metadata då även dessa kan matchas mot digitala auktoritetskällor till skillnad från fritexttermer. Praktiker där öppen forskningsinformation skapas genom att använda beständiga identifierare och kontrollerade vokabulärer för att beskriva forskningsprodukter på ett strukturerat sätt i linje med FAIR-principerna överlappar i stort med de allmänna principerna för länkade data och den semantiska webben såsom definierade av Tim Berners-Lee [23]. Samtidigt kan metoderna både användas för att beskriva fakta och att uttrycka att något är en tolkning, vilket kan svara mot disciplinspecifika behov.

⁴ DOI, *Digital Object Identifier*, är en vanlig beständig identifierare för exempelvis vetenskapliga artiklar och forskningsdata.

⁵ ORCID, *Open Researcher and Contributor ID*, är en beständig identifierare för individer som verkar i forskningssammanhang.

⁶ ROR, *Research Organisation Registry*, är idag den dominerande beständiga identifieraren för forskningsorganisationer efter att officiellt ha ersatt system som GRID och Crossref Funder ID.

Affilieringsmetadata

En förutsättning för att datapubliceringar eller andra forskningsprodukter ska kunna kopplas till en forskningsorganisation är att den innehåller någon form av relation till organisationen, och den vanligaste källan är affilieringsinformation. Traditionella författaradresser har formen av institutions- och organisationsnamn, ibland med ort och land, uttryckta i exempelvis parenteser eller fotnoter kopplade till individer i författarlistor. Alla publiceringsinfrastrukturer tillgängliggör dock inte informationen som strukturerade metadata där enskilda individers uppgifter kan särskiljas utan ytterligare granskning.

Idag finns ofta möjligheter att peka ut affilieringar direkt med beständiga identifierare som ROR, vilket skapar kopplingar till organisationer som inte kan misstolkas. Nyare publiceringsinfrastrukturer brukar ha stöd för forskaridentifieraren ORCID för att peka ut medverkande. Organisationsmetadata för personens nuvarande och tidigare affilieringar kan också ha lagts till i ORCID-posterna, och ibland har metadata om själva publiceringen tillförts posten. Det innebär att forskningsorganisationer som systematiskt bevakar ORCID-identifierare som pekar ut affilierad personal därmed har goda möjligheter att spåra forskningsprodukter som datapubliceringar [24].

I många fall har dock inga beständiga identifierare angetts vid publicering. Många forskare har heller inte något ORCID att ange. Det innebär att forskningsorganisationer fortsatt har ett behov av att underhålla listor med alla relevanta textkombinationer som kan förekomma i affilieringsmetadata som har matats in på frihand, inklusive vanliga felstavningar, missuppfattningar och äldre namnformer.

En problematik när olika forskningsprodukter ska knytas till forskningshuvudmän är också att en medverkandes affiliering inte i sig pekar ut organisationen som ansvarig för forskningsprodukten. För datapubliceringar finns det ett behov av att särskilja fall där forskningsorganisationen är ansvarig huvudman. Detta är ofta svårt att avgöra med endast affilieringsmetadata som grund.

Formella relationer och kunskapsgrafer för forskningsinformation

Flera andra typer av relationer kan uttryckas i metadata med öppen forskningsinformation. Exempelvis kan de uttrycka hur en forskningsartikel bygger på en datapublicering eller hur en programvara genererat ett resultat. I exemplet *DataCite Metadata Schema* uttrycks detta genom fälten *relatedIdentifier* och *relationType*, som anger den beständiga identifieraren för det relaterade objektet och den relation som råder mellan de två (t. ex. *IsSupplementTo*, *Cites*, *IsDerivedFrom*) [25].

Dessa relationer skapar riktade hänvisningar i en grafstruktur av öppen forskningsinformation. På så sätt kan metadata inte bara beskriva ett enskilt objekt, utan också väva samman en hel kontext av artiklar, forskningsdata, metoder och verktyg, men även återanvändning och citat. Genom att vandra mellan relationerna kan ytterligare information fås om kontexten. Principen att arbeta på detta sätt kallas *kunskapsgrafer för forskningsinformation* eller *scientific knowledge graphs*. Metodiken används redan av många internationella aktörer, även om låg kvalitet på metadata ibland ställer till hinder [26].

Förvaltning av information om forskningsaktiviteter

Utöver den hantering av forskningsinformation som görs i olika infrastrukturer för publicering finns ofta system som hanterar information om forskningsaktiviteter hos forskningsorganisationerna själva, eller hos forskningsfinansiärerna. En vanlig typ är så kallade CRIS-system⁷, som kan ha mer heltäckande samordnande och administrativa funktioner för forskningsverksamhet, eller vara av enklare karaktär där fokus främst ligger på uppvisning av forskningsaktiviteter och resultat för omvärlden. I de fall där systemen förvaltas av forskningsfinansiärer kan informationsomfånget ibland vara avgränsat till finansiärens huvudsakliga intressen, såsom projektsammanfattningar och tilldelade anslag.

Även i CRIS-systemen finns stora variationer i hur öppen forskningsinformation görs tillgänglig i form av maskinläsbara metadata. Vissa system är endast anpassade för att skapa statistik information som visas på en webbsida i exempelvis en forskningsportal, men levererar inte forskningsinformationen som standardiserade och jämförbara metadata. Det finns också en stor variation på graden av användning av beständiga identifierare. En metadatastandard för att skapa interoperabel information om forskningsaktiviteter i CRIS-system har tagits fram i form av CERIF (*Common European Research Information Format*), som samordnas av organisationen euroCRIS [27]. CRIS-systemen hanterar ibland datapubliceringar som supplementär information associerad med resultatpublikationer. Genom att registrera datapubliceringar som egna objekt, på motsvarande sätt som traditionella publikationer, är det möjligt att skapa goda förutsättningar för att de kan behandlas som självständigt spårbara forskningsprodukter knutna till forskningsaktiviteterna.

Vad som i praktiken beskrivits som avgränsade forskningsaktiviteter har traditionellt sett ofta varit forskningsprojekt med extern finansiering. Strävan mot öppen forskningsinformation har skapat ett behov av att även forskningsverksamhet utan anslag, såsom lönefinansierad forskning och mer småskaliga samarbeten, kan bli redovisade som forskningsaktiviteter. En mer inkluderande definition av konceptet i standarder och infrastrukturer kan antas främja öppen vetenskap och uppföljning genom att forskningsprodukter i så stort omfång som möjligt kan knytas till en spårbar forskningsaktivitet. Det finns också beständiga identifierare som kan peka ut och beskriva forskningsaktiviteter i sig. PID-systemet RAiD (*Research Activity Identifier*), en ny ISO-standard, kan tilldela identifierare till projekt och andra forskningsaktiviteter. RAiD lagrar kärnmetadata om fenomen som aktiviteten är relaterad till, som personer, organisationer, anslag, artiklar eller datapubliceringar, uttryckta med relationer till andra identifierare [28]. RAiD-identifieraren, som administreras av forskningsaktivitetens huvudman⁸, kan därmed skapa en knutpunkt med en maskinläsbar översikt av forskningsaktiviteten, ett tillvägagångssätt som enligt EOSC bör implementeras i Europa [29].

⁷ CRIS, *Current Research Information System*. RIMS, *Research Information Management System* förekommer också. Se även *Definitioner*.

⁸ Infrastruktur och organisering för att tilldela RAiD-identifierare till forskningsaktiviteter i Sverige har dock inte fastställts i dagsläget.

Infrastrukturer som samlar in öppen forskningsinformation

Information om forskningsprodukter som används för bedömningar, jämförelser och bibliometri har traditionellt sett ofta köpts som tjänster från kommersiella aktörer. För specifikt datapubliceringar tycks särbehandlingen i tidskrifters formalia ha lett till att kommersiell indexering fått begränsad täckning. Clarivates tjänst *Data Citation Index* har funnits sedan 2012, och dess indexering drar åt STEM-ämnena [30]. Elseviers motsvarighet *Data Monitor* lades ned 2025 på grund av lågt intresse [31].

I och med ökad tillgång till öppen forskningsinformation har dock fler internationella aktörer som utbyter metadata med varandra börjat sammanställa information om forskningsprodukter och dela dem fritt enligt öppna standarder [32]. Flera av dessa följer modellen för kunskapsgrafer för forskningsinformation, tar hänsyn till datapubliceringar, och möjliggör sökningar och översikter via webbgränssnitt och API:er. *OpenAlex* har tagit över och fortsatt berika det metadataunderlag som fanns hos nedlagda *Microsoft Academic Graph* [33]. Den europeiska infrastrukturen *OpenAIRE* har satt upp *OpenAIRE Graph* [34], som hämtar metadata från många källor, inklusive sådana som genererats via *Scholix*, ett tekniskt ramverk för att utbyta information om relationer mellan datapubliceringar och resultatpublikationer [35]. Flera aktörer samverkar också med *OpenCitations*, en infrastruktur som skapar beständiga identifierare som pekar ut enskilda citeringar [36]. I *Google Dataset Search* görs en indexering av datapubliceringar som har tillräckligt tydliga metadata [37].

I Finland och Norge har statliga uppdrag lett till skapandet av de nationella forskningsportalerna *Forskning.fi* och *Nasjonalt vitenarkiv (NVA)* [38], [39]. Dessa är fullskaliga nationella CRIS-system som tar emot metadata från forskningsorganisationer, även gällande datapubliceringar, och kan dela dem öppet med internationell indexering, exempelvis till *OpenAIRE Graph*. I Sverige finns ännu ingen motsvarighet. Kungliga bibliotekets index *SwePub* omfattar huvudsakligen andra forskningsprodukter än datapubliceringar [40]. Sedan 2025 finns dock forskningsdataportalen *Researchdata.se*, som bevakar vissa utpekade datakällor och samlar in information om datapubliceringar som kan knytas till svenska forskningsorganisationer och uppnår tekniska minimikrav på metadata. *Researchdata.se* rapporterar även denna metadata till Digg så att öppna datalagens informationskrav delvis kan efterlevas av huvudmännen [41].

Kvalitet och användbarhet hos forskningsinformation

Utgångspunkten för hållbart skapande av öppen forskningsinformation är alltså att informationen i så stor utsträckning det går tekniskt struktureras och definieras så att effektiv återanvändning kan ske. Användbarheten beror i stort på kvaliteten hos den metadata som skapas. För många användningsområden är det viktigt att möjliggöra automatiserad återanvändning, utan krav på tidskrävande manuell granskning eller att nödgas använda AI-verktyg och *text mining* för att försöka sälla i ostrukturerade uppgifter. Det är därför avgörande att metadatastandarder följs, att beständiga identifierare används i så stor utsträckning som möjligt och att den som skapar informationen har en förståelse för hur den kan komma att återanvändas. När öppen forskningsinformation och metadata

knutna till publiceringsinfrastrukturer, PID-system och CRIS-system skapas krävs således en analys av möjliga återanvändningsområden och ett systematiskt arbetssätt för att andra ska kunna ha full nytta av informationen. Stödmaterial, vägledning och verktyg bör utformas på sätt som gör att både allmängiltiga och disciplinspecifika behov ges hänsyn.

Variationen i kvalitet hos öppen forskningsinformation och metadata är i dagsläget mycket stor för flera typer av forskningsprodukter, inte minst för datapubliceringar. Brister som försvårar spårbarhet och återanvändning av metadata uppstår inte sällan redan på systemnivå. Vissa infrastrukturer, som förlag eller datarepositorier, levererar på grund av tekniska misstag eller brist på granskning regelbundet felaktiga eller begränsade kärnmetadata till PID-systemen. Detta kan exempelvis ske på grund av felkonfigurerad automatisering, eller att man inte uppdaterar metadata överallt efter ändringar, även om korrekt information faktiskt finns lagrad på andra platser i infrastrukturerna [42].

Goda praktiker kan också förhindras av regler som skapas av infrastrukturerna själva. Förlag kan ofta hänvisa datapubliceringar till vissa supplementformer som hindrar att datamaterialet kan spåras som adresserbara forskningsprodukter, då endast de sammanhörande resultatpublikationerna är beskrivna i metadata eller har fått en beständig identifierare [43]. Andra hinder kan skapas då möjligheten att ange en uppgift i metadata inte finns, eller att krav finns på att den efterfrågade informationen anges eller struktureras på ett sätt som i praktiken kräver manuell granskning för att uttolkas. Att kunna ange beständiga identifierare eller kontrollerade vokabulärer för att hänvisa till andra entiteter eller definitioner stöds inte heller i alla infrastrukturer, även om etablerade identifierare finns för ett givet fenomen.

Ett stort ansvar faller givetvis också på enskilda individer som fattar beslut om de metadata som anges i en given situation, inte sällan enskilda forskare. Ett hållbart skapande av metadata under hela forskningsprocessen kräver en god förståelse för hur beständiga identifierare används och hur metadatainnehållet påverkar forskningsekosystemet, både hos forskarkåren och hos forskningsstödjande funktioner [44]. Detta kräver självfallet också effektiva verktyg, ett gediget stöd och tydlig vägledning.

Verktyg för att utvärdera kvalitet hos metadata

Många av principerna för att skapa datapubliceringar som är FAIR har att göra med teknisk interoperabilitet och följsamhet till standarder. Då vissa av dem också kan utvärderas på automatiserad väg har en flora av verktyg växt fram, som arbetar med så kallade *FAIR Metrics* [45]. Till aspekter som kan provas hör förekomst av beständiga identifierare, strukturerade metadata, användning av standardiserade vokabulärer eller tydlig information om licenser och restriktioner. Då flera av dessa har bäring på just spårbarheten går det att arbeta förebyggande med FAIR Metrics för att etablera en lägstanivå av maskinläsbarhet, interoperabilitet och teknisk hållbarhet för spårbara datapubliceringar. Däremot behöver mänsklig granskning komplettera metoderna, för att kunna säkerställa innehållsmässig kvalitet och utvärdera praktisk användbarhet.

Stödorganisation

Sammanfattning av stödorganisation

- Det finns oftast stöd för forskare på lokal nivå för att skapa spårbara datapubliceringar, men stödets omfattning och inriktning varierar stort.
- Det finns ett antal nationella infrastrukturer som kan stötta forskare, som är uppbyggda på olika sätt och kan vara disciplinspecifika.
- Det nationella stödet för att följa upp datapubliceringar är inte utbyggt så att en enskild forskningsorganisation enkelt kan genomföra uppföljning.

Beskrivning av stödorganisation

Vetenskapsrådet understryker i sin årliga uppföljning om publicering av forskningsdata [46] vikten av att samordna hanteringen för att underlätta arbetet med hantering av forskningsdata. I Sverige finns stödfunktioner för forskningsdatahantering på flera olika nivåer. Flera av dessa är av relevans för skapandet av spårbara datapubliceringar i linje med FAIR-principerna.

Lokala stödfunktioner

De flesta svenska forskningsorganisationer erbjuder olika typer av stöd för hantering och publicering av forskningsdata. Stödets omfattning och karaktär är dock mycket varierande, beroende på organisationens storlek, inriktning och olika lokala omständigheter. Stödet kan bestå av såväl tekniska tjänster, programvaror och annan digital infrastruktur som stödpersonal för olika typer av frågor. Stödpersonal med relevant kompetens kan finnas i många olika delar av en organisation – särskilt i större organisationer är uppgifterna ofta utspridda på många personer och enheter. I mindre organisationer kan ett problem vara att det saknas stöd inom vissa områden, eller att det stöd som finns är begränsat och koncentrerat till några få befattningar.

De organisationer som ingår i Svensk nationell datatjänsts nätverk har i och med medlemskapet åtagit sig att organisera en så kallad *Data Access Unit (DAU)*, en lokal organisation för att stödja forskningsdatahantering [47]. I beskrivningen föreslås att följande kompetenser kan ingå, varav i princip alla kan bli involverade i frågor som på olika sätt handlar om tillgängliggörande av forskningsdata [47, s. 4]:

- *Granskning och kurering av data och metadata*
- *Publicering av data och metadata*
- *Juridiska kompetenser, särskilt med relevans för forskningsdata*
- *Forskningsetisk kompetens*
- *Praktiskt stöd för datahantering, inklusive planering av datahantering (t.ex. DMP)*
- *IT- och informationssäkerhet*
- *IT-kompetenser (som -drift, -utveckling, -arkitektur)*
- *Arkivering*
- *Forsknings- och innovationsstöd, eller motsvarande kompetenser*
- *Undervisning och kommunikation*
- *Bibliometri*

Nationella stödfunktioner

Nationellt finns ett antal stödfunktioner som forskare kan nyttja beroende på vad de har för fråga, disciplin, typ av forskning och typ av datamaterial.

Svensk nationell datatjänst (SND)

Svensk nationell datatjänst [48] är en nationell forskningsinfrastruktur finansierad via Vetenskapsrådets infrastrukturbidrag, med Göteborgs universitet som huvudman. SND står bakom forskningsdataportalen och metadatakatalogen Researchdata.se, och har också ett eget verktyg för publicering av forskningsdata (DORIS), där datapubliceringarna lagras i lokala datarepositorier kopplade till svenska lärosäten eller det CoreTrustSeal-certifierade repositoret SND-CARE. SND stöttar Sveriges lärosäten med bland annat tekniskt stöd för att skapa DOI för forskningsdata. Det kan ske genom att forskare skapar datapubliceringar via DORIS, eller att lärosäten på egen hand registrerar DOI för datapubliceringar via medlemskapet i SND:s nationella DataCite-konsortium.

National Academic Infrastructure for Supercomputing in Sweden (NAISS)

Även NAISS [49] är en forskningsinfrastruktur finansierad av Vetenskapsrådets infrastrukturbidrag, men med Linköpings universitet som huvudman. NAISS tillhandahåller beräkningskapacitet och tjänster för AI och datahantering med hjälp av superdatorer. Eftersom de beräkningar som görs hos NAISS ofta använder stora mängder data tillhandahåller de tjänster för lagring och hantering av data i anslutning till sina beräkningstjänster. De resurser som finns hos NAISS kan utnyttjas efter tilldelning till forskare som ansöker vid olika löpande utlysningar.

SciLifeLab

SciLifeLab [50] är en nationell infrastruktur för livsvetenskaper. Inom SciLifeLab organiseras också flera olika instrument och tjänster som gör olika typer av kemiska, biologiska eller medicinska analyser. För att stödja forskningen inom SciLifeLabs olika områden har organisationen byggt upp ett Data Centre och en dataplattform som erbjuder olika tjänster för datahantering, inklusive SciLifeLab Data Repository, ett repositorium för publicering av forskningsdata [51]. SciLifeLabs dataplattform erbjuder också en tjänst för att publicera direkt körbara dataapplikationer - SciLifeLab Serve - liksom ett antal färdiga arbetsflöden för olika typer av dataanalys. Dessa arbetsflöden gör det möjligt att på reproducerbara sätt förbereda data för visualisering och publicering i standardiserade format.

Huminfra

Huminfra [52] är en nationell infrastruktur inom humaniora, med Lunds universitet som samordnande organisation. Huminfra agerar som en övergripande infrastruktur som sammanställer och länkar till andra infrastrukturer och dataresurser för svensk forskning inom humanioraämnena. Infrastrukturen kan ge stöd i frågor om att strukturera, tillgängliggöra och göra humanistiska data mer begripliga. Vidare utvecklar organisationerna som samarbetar inom Huminfra också gemensamma kurser och utbildningsmaterial för digitala forskningsmetoder inom humaniora.

InfraVis

InfraVis [53] erbjuder stöd för visualisering av data till forskare vid svenska lärosäten. InfraVis arbetar med en bred definition av visualisering – målet är att skapa en grafisk representation av data men även analys på vägen dit räknas in. Visualiseringarna genomförs på reproducerbara sätt och kan kopplas till genomförda datapubliceringar. Forskare som bara behöver enklare konsultation kan kontakta InfraVis för hjälp, och de som behöver fler timmar kan ansöka om detta i en årlig utlysning.

Stödfunktioner för uppföljning

Vid uppföljning av forskning använder många svenska lärosäten internationella datakällor, ofta från privata aktörer. Det finns också några svenska organisationer som samlar in information om den svenska forskningens olika forskningsprodukter. Uppdragen om öppet tillgänglig publicering av forskningsprodukter i Sverige är delade mellan Kungliga biblioteket, som har ett uppdrag om öppen tillgång till publikationer, och Vetenskapsrådet, som har ett uppdrag om öppet tillgänglig publicering av forskningsdata. Det finns ingen enskild organisation som samlar in och följer upp information om datapubliceringar.

Vetenskapsrådet och uppföljning av öppen publicering av forskningsdata

Vetenskapsrådet har ett uppdrag att arbeta för öppen publicering av forskningsdata, och följer därför upp utvecklingen i Sverige [54]. Det arbetet görs dock i huvudsak genom enkäter och insamling av information från stödfunktionerna vid Sveriges lärosäten. VR genomför i dagsläget inte någon teknisk insamling av metadata om datapubliceringar på egen hand, däremot finansierar de Svensk nationell datatjänst (SND) via infrastrukturanslag.

Kungliga biblioteket och Swepub

Kungliga biblioteket tillhandahåller databasen Swepub [40]. Detta är en nationell publikationsdatabas som samlar data om publikationer från alla svenska lärosäten och andra forskningsorganisationer. Data i Swepub används för att följa upp till exempel andelen publikationer med öppen tillgång, som en del av KB:s uppföljning av sitt uppdrag på detta område [55]. Även om Swepub har tekniska möjligheter för indexering av datapubliceringar har detta inte gjorts i någon större utsträckning, främst på grund av det uppdelade samordningsuppdraget för resultatpublikationer och forskningsdata.

SND och researchdata.se

SND och ett antal samarbetsorganisationer tillhandahåller Researchdata.se – en portal med möjlighet att söka efter forskningsdata från svenska lärosäten och andra forskningsorganisationer [41]. Researchdata.se aggregerar information från flera kända datakällor, och gör det härigenom möjligt att söka ut och följa upp datapubliceringar som gjorts tillgängliga via dessa källor. Tjänsten utökas löpande med fler datakällor, som kan anmälas för bevakning så länge som de uppfyller minimikrav för metadata, som att kunna kopplas till en svensk forskningsorganisation och ett ämnesområde [56]. Metadata från Researchdata.se skördas av Digg och indexeras därmed också i Sveriges dataportal.

Tillgängliggörande och publicering

Sammanfattning av tillgängliggörande och publicering

- Datapubliceringar blir skapade på många olika sätt och formen kan påverkas av disciplinspecifika behov, yttre krav eller vilken arbetsinsats som är möjlig.
- Datapubliceringens form och infrastrukturen som används påverkar ofta dess spårbarhet.
- Supplementformat i tidskrifters webbinfrastrukturer är fortfarande ett av de vanligaste sätten att tillgängliggöra forskningsdata på, men detta är ofta förknippat med avsaknad av adresserbarhet och spårbarhet.
- Datarepositorier är vanliga infrastrukturer för att skapa datapubliceringar, som ofta ger goda möjligheter för spårbarhet, då de ofta erbjuder beständiga identifierare och möjligheter att beskriva data med rika metadata.
- Datarepositorier kan ofta vara ändamålsanpassade, med möjligheter att publicera både öppna data och data med åtkomstbegränsningar.
- Datarepositorier kan ha olika nivåer av kvalitetskontroll, automatiserad såväl som manuell, och kan vara certifierade med olika kvalitetscertifieringar.
- Data papers erbjuder ett sätt att beskriva och hänvisa till datapubliceringar som ger ökade möjligheter att få erkännande, och kan öka datapubliceringars synlighet då de kan bli uppfångade av indexering som i dagsläget exkluderar själva datapubliceringarna.
- Vissa former av datapubliceringar, som inrapportering till databaser, är svåra att adressera som enskilda datapubliceringar eftersom det inte alltid finns etablerade processer för att särskilja och spåra dessa.
- Registerforskning är ett tydligt exempel, där data förvaltas som en sammanhållen helhet och där nya inrapporterade data inte alltid dokumenteras som fristående forskningsprodukter.
- För data som förvaltas i databaser och register finns dock tekniska och konceptuella möjligheter att avgränsa och beskriva utdrag och bidrag, och fixera dem på spårbara sätt.
- Citering av datapubliceringar är möjligt i de flesta kontexter, men om och hur det faktiskt görs kan bero på flera olika omständigheter.
- Särskilda disciplinspecifika sätt att hänvisa till data, som exempelvis accessionsnummer, kan vara svåra att fånga upp på automatiserad väg.
- Referenshanteringsprogram kan hjälpa till med dataciteringar, men felaktiga metadata orsakar problem vid citering.
- Förlags och tidskrifters regler för att hänvisa till och beskriva datapubliceringar har skapat ett spretigt landskap utan standardisering, där olika åtgärder leder till att spårbarhet ibland kan främjas, men ibland också förhindras.

Beskrivning av tillgängliggörande och publicering

I dagsläget tillgängliggörs och publiceras forskningsdata på många olika sätt. Några är enkla att förhålla sig till, medan andra inte lika enkelt kan adresseras eller avgränsas som datapubliceringar. I många fall har detta inte att göra med datamaterialets karaktär, utan snarare med ramar som sätts av det format eller den infrastruktur som använts för tillgängliggörande och de begränsningar eller regler som lösningen medför. Dessa ramar kan ofta även påverka datapubliceringarnas spårbarhet.

Särskilda disciplinspecifika behov gör att det finns många olika specialiserade sätt för att tillgängliggöra och hänvisa till forskningsdata. Det finns också olika epistemologiska synsätt på vad som utgör data i en viss forskningskontext. Denna översikt kan därför inte göra anspråk på att vara heltäckande, eller uppge definitiva avgränsningar för vad som kan anses utgöra en datapublicering.

Supplementformat

Det sätt som kanske är vanligast att tillgängliggöra forskningsdata på idag är i form av supplementformat⁹ som är underordnade resultatpublikationer. Sättet som forskningsdata i supplementformat distribueras på kan ta sig olika former, men formen kallas vanligen *supplementary files*, *supporting information* eller liknande, beroende på den enskilda tidskriftens norm. Denna distributionsform är också ibland skild från hänvisningar till datamaterial som görs i redogörelser för dataåtkomst (*data availability statements*). I många fall går dessa supplementfiler inte att spåra på egen hand. De hittas via särskilda länkar i förlagens webbpresentationer av artiklar eller bokkapitel, ofta i särskilda spalter eller sektioner som är åtskilda från huvudpublikationen. De lagras ofta som enstaka filer i förlagets innehållssystem, och får sällan beständiga identifierare eller beskrivande metadata. Detta tillvägagångssätt används för många typer av behov, utöver att tillgängliggöra forskningsdata, såsom att bifoga en tabell eller figur som inte får plats av utrymmesskäl eller att ge ytterligare detaljer om insamlingsmetodik och arbetsflöden. Det är därför mycket svårt att systematiskt skilja ut forskningsdata som tillgängliggörs på detta sätt från övrig information, eller att bedöma om exempelvis ett kalkylark utgör forskningsdata eller i stället är en statistikredovisning som inte fick plats i artikeln. Det krävs således manuell granskning för att kunna fastslå innehållets karaktär.

Sammanfattningsvis kan konstateras att supplementformatet sällan tillgängliggörs i linje med FAIR-principerna. Fricke, Enslow & Shipman understryker de många sårbarheterna som är förknippade med detta system och förespråkar att förlagen ska övergå till att ge supplementärt material beständiga identifierare och låta det indexeras med separata metadata för att möjliggöra spårbarhet [43]. En trend kan dock skönjas där självpublicering i tjänster som tillåter flera resurstyper, exempelvis Figshare, i större utsträckning börjar användas för att tillgängliggöra material i supplementformat med beständiga identifierare.

⁹ Begreppet bör inte förväxlas med *tidskriftssupplement*, en benämning på ett slags specialnummer för tidskrifter.

Datarepositorier

Det traditionella datarepositoriet är en vanlig publiceringskanal för forskningsdata. Det är vanligt att datarepositorier följer en eller flera grundläggande metadatastandarder och tilldelar beständiga identifierare, såsom DOI, till datapubliceringar i form av versionerade paketeringar av data och dokumentation. Datarepositorier kan vara generella, det vill säga utformade för tvärdisciplinär användning och passa många ändamål. De kan också vara mer specialiserade för att svara mot disciplinspecifika behov, exempelvis vad gäller segmentering och adressering av data, hur data kan beskrivas och vilka metadata som kan genereras. Publicering i datarepositorier kan vara ett av de sista stegen i olika specialiserade och fastställda arbetsflöden, som kan ha olika grad av automatisering.

Ett datarepositorium är i regel byggt kring en eller flera lagringslösningar för datafiler och metadata. Tekniska lösningar, åtkomstprotokoll, lagringskapacitet, bandbredd, säkerhetsaspekter och följsamhet till olika standarder kan påverka vilka kategorier av data som repositoriet bör användas för.

En viktig fråga är vilka åtkomstnivåer för datapubliceringar som datarepositoriet är avsett för: öppet tillgängliga data eller data som har begränsad åtkomst. Flera repositorier möjliggör skapandet av båda formerna av datapubliceringar. Att skapa spårbara datapubliceringar i linje med FAIR-principerna är ofta möjligt både för öppet tillgängliga och åtkomstbegränsade data, förutsatt att datarepositoriets funktioner för t.ex. skapande av metadata ger ett grundläggande stöd för detta.

Kraven på information som anges i metadata och dokumentation för en datapublicering varierar dock mellan repositorierna. En skillnad mellan olika repositorier som kan påverka den faktiska metadatakvaliteten är i vilken utsträckning metadata och datamaterial granskas och kureras av andra än inlämnaren. Olika modeller finns, såsom helt autonom självpublicering, aktiv kvalitetsgranskning innan publicering, eller arbete med förädling av redan publicerade metadata i efterhand. Förekomst eller avsaknad av aktiv kvalitetsgranskning kan ha betydelse för spårbarheten hos datapubliceringar som görs i repositorierna.

Vissa principer och certifieringar har tagits fram för kvalitetssäkring av datarepositorier. TRUST-principerna är ett allmänt ramverk för hur datarepositorier bör arbeta för att vara tillförlitliga [57]. Begreppet *trustworthy digital repositories* (TDR) förekommer ofta i detta sammanhang.¹⁰ Tre certifieringar har i dagsläget relevans för kvalitetssäkring: *CoreTrustSeal*, som är ett samarbete mellan flera internationella forskningsdataaktörer och har ersatt den tidigare DSA-certifieringen [58], *nestor Seal* (DIN 31644), som främst använts för certifiering av tyska repositorier [59], och *ISO 16363*, som har sitt ursprung i datahantering för rymdfart och rymdforskning [60]. Certifieringarna förhåller sig till OAIS-modellen (ISO 14721), en arkitekturmodell för arkivliknande verksamheter [61].

¹⁰ Begreppet bör användas med försiktighet i beskrivningar och kravställningar då detta, beroende på kontext, både kan syfta på en generisk målbild och särskilda kvalitetscertifieringar.

Datatidskrifter och data papers

Fenomenet datatidskrifter eller *data journals* har uppkommit för att ge möjlighet att presentera forskningsdata i form av en särskild beskrivande artikel (*data paper*) som hänvisar till en datapublicering. Ett data paper och dess relaterade datapublicering kan vara fristående, och behöver inte hänga samman med ett särskilt tillgängliggörande av forskningsresultat. Ibland hör datatidskriften ihop med eller hänvisar till en särskild publiceringsinfrastruktur för datapubliceringar, såsom ett utpekad datarepositorium. Artikeln i data paper-format och datapubliceringen utgör dock i de allra flesta fall separata forskningsprodukter med egna beständiga identifierare.

Datatidskrifter kan ge större möjligheter att presentera aspekter som metodik, tillförlitlighet hos variabler, eller olika överväganden i datainsamlingsprocessen än vad som är praxis att bifoga i medföljande dokumentation. Detta kan vara värdefullt inom flera discipliner. Dessutom genomförs ofta referentgranskning som kan vara mer omfattande än vad som kan erbjudas av andra infrastrukturer.

Samtidigt kan datatidskrifterna också ses som ett fenomen som uppstått som en spontan reaktion på incitamentsproblematiken där datapubliceringar särbehandlas som forskningsprodukter och sällan citeras eller plockas upp av indexering. Även om verkshöjden för innehållet i ett data paper i många fall inte nämnvärt behöver skilja sig från en datapublicerings bifogade dokumentation, så utgör den en annan publikationstyp som i nuvarande forskningsekosystem oftare synliggörs och ger upphovspersonerna större chans att få erkännande [62].

Datapubliceringsformer med begränsad adresserbarhet

Ett antal former för tillgängliggörande av forskningsdata kan på grund av särskilda omständigheter vara svåra att adressera som datapubliceringar, även om tillgängliggörandet i övrigt motsvarar en publicering. Exempelvis kan data löpande inrapporteras till större infrastrukturer eller databaser i form av nya mätningar eller observationer. Ett exempel där detta är vanligt är geovetenskapliga databaser.

Ett typiskt exempel är också registerstudier, där särskilda register förvaltas över tid med data som har en viss avgränsning. Det kan till exempel röra sig om mätvärden från forskningspersoner med en viss egenskap. Registret förvaltas som en sammanhållen datamängd och när nya data rapporteras in till registret uppgår dessa i mängden. Det finns dock inte alltid etablerade processer där det nytillkomna datamaterialet dokumenteras och pekas ut för sig som en adresserbar och spårbar forskningsprodukt. Det finns dock goda möjligheter för konceptuella och tekniska beskrivningar av både utdrag från och bidrag till register. Exempelvis kan avgränsningar som pekar ut data beskrivas i generella termer i klartext eller mer specialiserat som exakta skript, instruktioner eller databasfrågor. Vetenskapsrådet tillhandahåller verktyget RUT som hjälper till att dokumentera urval av specifika variabler och dataomfång från flera svenska register [63]. För hälsodata som används i forskningsverksamhet utarbetas nu också krav på särskild metadataredovisning inom ramen för Förordningen om det europeiska hälsodataområdet (EHDS-förordningen) [64].

Gemensamt för denna kategori är att alltså avgränsningen av datamaterialet behöver fixeras och sparas för att kunna tolkas och återanvändas. Även datautdraget kan behöva bevaras för sig, om framtida återskapande inte kan garanteras. Rauber m.fl. har ställt upp en kravprofil för hur dynamiska data bör kunna avgränsas och adresseras för att kunna vara spårbara, vilket också resulterat i rekommendationer från RDA som vänder sig både till användare och infrastrukturer [65], [66].

Citering av publicerade data

Att citera och skapa referenser till datapubliceringar är möjligt inom många olika referensstilar, och flera av de vanliga stilarna har särskilda regler för källtyper såsom "Dataset". Det är dock inte självklart att en datapublicering citeras på samma sätt som andra källor, beroende på olika yttre förväntningar och omständigheter, och hur man refererar till egna och andras data kan skilja sig åt.

Hänvisningssätt för forskningsdata skiljer sig också åt på en övergripande nivå mellan forskningsdiscipliner. Inom humaniora är fotnoter ofta norm. Samhällsvetenskapliga forskare använder vanligen stilar som stöder författare-år i brödtextr, medan andra discipliner kan infoga dataciteringar i tabeller, bilder eller i referenslistan. De olika sätten att referera på kan i sin tur påverka vad som går att fånga upp och göra synligt beroende på uppföljningsmetodik, och på hur systemen för elektronisk publicering hanterar referenserna och relaterade metadata.

De skiftande praktikerna kring citering av datapubliceringar har resulterat i flera konkreta rekommendationer riktade till forskningsorganisationer och andra avnämare, exempelvis från CESSDA ERIC och FORCE11 [67], [68]. Rekommendationerna betonar att datapubliceringar ska behandlas som värdefulla forskningsprodukter, med tydlig citering, beständiga identifierare och metadata av hög kvalitet så att data blir hittbara, tillgängliga, verifierbara och återanvändbara. Rekommendationerna vänder sig i första hand till beslutsfattare som långsiktigt kan påverka infrastrukturer, system och policyarbete för att skapa hållbara citeringspraktiker. Forskarna själva lutar sig oftast mot tidskrifters instruktioner för forskningsdata och reglerna i de generella referensstilar (exempelvis APA, MLA, Chicago och IEEE) som hänvisas till av förlagen [69].

Unika disciplinspecifika hänvisningar till publicerade data

Flera forskningsdiscipliner har unika sätt att tillgängliggöra och hänvisa till data, som ofta existerade innan många andra digitala metoder för hänvisningar vunnit hävd. Ett exempel på detta är så kallade accessionsnummer som ofta används inom livsvetenskaperna, såsom genetik, proteomik eller mikrobiologi. Dessa utgörs av siffror eller textsträngar som tilldelas efter inrapportering av ett visst datamaterial i en databas eller liknande infrastruktur. Även om de inte fungerar som webblänkar på samma sätt som beständiga identifierare är de oftast möjliga att slå upp i de infrastrukturer som genererat dem.

När accessionsnummer används som hänvisning till resultatpublikationer förekommer de ofta ensamma i löptexten eller i en fotnot. Denna praktik kan leva kvar även om typen av accessionsnummer senare har fått en uppslagsfunktion som går att länka till, och därmed låter dem fungera som beständiga identifierare. Det kan därför vara svårt att tillförlitligt söka fram och spåra dessa hänvisningar till publicerade data, med mindre än manuell genomläsning eller mönsterigenkännande *text mining*-metodik applicerad på fulltexter. Liknande problem uppstår med andra typer av hänvisningar till data som fungerar på snarlika sätt, såsom olika katalog- och samlingsnummer som kan vara vanliga inom andra discipliner.

Referensstilar och referenshanteringsprogram

I system för referensstilar, som APA, Chicago och IEEE, avgör källtypen vilka referenskomponenter som blir tillgängliga och hur en referens får presenteras. För datapubliceringar innebär detta att valet av en passande källtyp som "Dataset" gör att centrala uppgifter som skapare, titel, år, version, utgivare och beständig identifierare kan exponeras på ett sätt som skapar citerbarhet och spårbarhet. De vanligaste referensstilarna stödjer i dag formella dataciteringar i referenslistan.

Ofta används olika referenshanteringsprogram (som EndNote, Zotero m.fl.) för att skapa referenser. När en referens till en datapublicering görs med en korrekt källtyp, och referensens olika komponenter har angetts i avsedda strukturerade metadatafält, kan referensen presenteras på likvärdig nivå som andra källor. Om fel källtyp valts eller ifyllda metadatafält saknas blir referensen ofullständig och svår att följa upp.

Vid användning av olika referenshanteringsprogram kan användaren också hämta färdiga metadata för referenserna. Detta kan göras i form av referensfiler som hämtas via exporter från publiceringsinfrastrukturers landningssidor i vanliga filformat som exempelvis *.ris* eller *.bib*. Metadata för referenser kan också hämtas genom uppslag mot beständiga identifierare såsom DOI där verktygen kan läsa av kärnmetadata från PID-systemen. När datapubliceringen har tilldelats en korrekt källtyp (såsom "Dataset") i referensfiler eller tillhörande kärnmetadata och uppgifter som version, utgivare, DOI osv. exponeras i rätt metadatafält resulterar det i en korrekt citering. I frånvaro av korrekt metadata krävs ofta manuell komplettering för att referenser ska bli korrekta och spårbara.

Praktikerna för digital referenshantering öppnar för flera möjliga felkällor, varav många orsakas av infrastrukturerna själva. En infrastruktur som erbjuder nedladdning av referensfiler för referenshanteringsprogram kan ha konfigurerat metadata så att de hamnar i felaktiga fält, vilket gör att felen ärvs av författarna om de inte upptäcker och åtgärdar dem. På samma sätt kan infrastrukturer skicka felaktiga uppgifter till PID-systemens kärnmetadata eller inte uppdatera den efter att ha gjort ändringar, vilket gör att felaktiga uppgifter också hämtas av användarna av referenshanteringsverktygen. När användare hämtar referenser från tredjepartsaktörer, som metadataaggregatorer och datakataloger, kan också ytterligare felkällor uppstå då även dessa infrastrukturer brukar ha unika processer för att generera egna referensmetadata.

Riktlinjer från förlag och tidskrifter

Kraven på att tillgängliggöra underliggande forskningsdata har stärkts efter hand som förväntningarna på öppna vetenskapspraktiker ökat, och dessa har i många fall formaliserats i förlags och tidskrifters egna policyer. Samtliga större förlag, exempelvis Elsevier, Springer Nature, Wiley, MDPI och Taylor & Francis, anger att författare ska eller bör redogöra för var data finns och hur de kan nås [70]. Skärpan på dessa krav eller rekommendationer varierar, men gemensamt är att datamaterialet och dess tillgänglighet ska beskrivas i samband med publicering.

Förlagens riktlinjer kan också styra över möjligheter att själv citera datapubliceringar. I flera publiceringsmiljöer är det uttryckligen praxis att formellt hänvisa till datamaterialet, exempelvis bland Elsevier, Springer Nature med flera [70]. Dessa riktlinjer innebär att praktiken att själv citera datapubliceringar i referenslistan inte bara är accepterad utan förväntad när data utgör underlag för resultatpublikationen. Samtidigt förekommer generella redaktionella policyer som avråder från överdriven själv citering utan saklig motivering. Sådana policyer syftar till att motverka irrelevanta själv citeringar och manipulation av citeringssiffror, men kan i praktiken bidra till osäkerhet hos författare kring när och hur egna datapubliceringar bör citeras formellt. Detta riskerar i sin tur att försvaga kopplingen mellan resultatpublikation och data när formella data citeringar uteblir, vilket kan göra datapubliceringarna mindre spårbara. Utöver detta kan datamaterialet vara tillgängliggjort i supplementformat, vilket innebär att en hänvisning till data kan vara av typen "se supplementfil 1", något som sällan blir sökbart och spårbart som datapublicering.

I praktiken redovisas datamaterialets tillgänglighet ofta genom redogörelser för dataåtkomst (*data availability statements*), som fungerar som en artikelsektion för information om hur data går att komma åt. Enligt förlagens anvisningar för redogörelser för dataåtkomst ska dessa vanligtvis innehålla beständiga identifierare eller länkar till repositorer, samt uppgifter om åtkomstbegränsningar och hur utlämnanden provas när öppen tillgång inte är möjlig. I vissa publiceringsmiljöer specificeras dessutom olika kategorier av data där datapublicering i vissa ämnesspecifika repositorer krävs [70].

De uppgifter som uppges i redogörelser för dataåtkomst kan vara allt från öppet tillgängliga datapubliceringar med beständiga identifierare till uppgifter om att data endast kan lämnas ut på begäran, eller inte kan delas alls. Följden är att en väl ifylld redogörelse för dataåtkomst inte innebär att data är öppna, utan i första hand klargör vilka villkor som gäller för åtkomst. Dessutom skiljer sig krav, förväntat innehåll och stödmallar mellan förlag och enskilda tidskrifter. Vissa tillhandahåller detaljerade textexempel och ställer krav på användning av särskilda repositorer för vissa datakategorier, medan andra nöjer sig med mer allmänna riktlinjer. Systemet gör att uppgifter om datapubliceringar ofta hamnar i fritext och inte alltid resulterar i strukturerade metadata, vilket försvårar maskinell spårbarhet och länkning mellan artikel och data. Samtidigt är rutiner för efterlevnad och kontroll inte alltid tydliga. Det är inte alltid klart om forskningsdata faktiskt görs tillgänglig med öppen tillgång eller på förfrågan även när detta uppges, och hur dataåtkomst kan beskrivas varierar stort mellan publiceringsmiljöer [70].

Bevarande och arkivering

För att vetenskapliga resultat ska kunna granskas och verifieras måste forskningsdata bevaras vid forskningsorganisationen och vara tillgängliga även efter att en forskningsaktivitet är avslutad. I lagen om ansvar för god forskningssed och prövning av oredlighet i forskning (SFS 2019:504) anges att forskningshuvudmän är ansvariga för att lämna de uppgifter som behövs för att utreda oredlighet i forskning, vilket omfattar allt material som ligger till grunden för forskningsresultat [71].

Forskningsdata som samlas in eller tas emot av en myndighet som bedriver forskningsverksamhet blir som de flesta andra handlingar till allmän handling enligt tryckfrihetsförordningen (SFS 1949:105) [72]. Dessutom omfattas forskningsdata som har tagits fram i forskning som bedrivs av en myndighet eller av ett i lagen särskilt omnämnt lärosäte av arkivlagen (SFS 1990:782) [73]. Detta innebär att forskningshandlingar ska hållas ordnade och vara tillgängliga så att exempelvis allmänhetens rätt till insyn kan tillgodoses. Dessa krav omfattar således både avgränsade datapubliceringar och de insamlade forskningsdata som de bygger på.

Utgångspunkten är alltså att forskningsdata ska bevaras. Praxis på de flesta lärosäten är på grund av lagkraven en gallringsfrist på 10 år för forskningsdata, ett minimikrav sprunget ur redogörelsekravet vid en potentiell oredlighetsutredning. Denna period ska räknas från senast publicerade resultatpublikation baserad på datamaterialet. Det ställs också krav på dokumentation av viss medicinsk forskning som medför en gallringsfrist på 15 år. Riksarkivets föreskrifter ger riktlinjer för när forskningsdata bör bevaras långsiktigt även efter gallringsfristen (RA-FS 1999:1) [74], och ytterligare vägledning för praktisk bedömning har tagits fram av SUHF:s expertgrupp för arkiv och informationshantering [75].

Sammanfattningsvis kan konstateras att datapubliceringar som genomförs inom ramen för myndigheters och vissa andra forskningsorganisationers forskningsverksamhet behöver kunna följas upp och arkiveras inom ramen för den löpande förvaltningen för att följa arkivlagstiftningen och andra lagkrav. För att möta de samlade lagkraven behöver publicerade forskningsdata kunna återfinnas och förstås, vilket i praktiken innebär att de behöver vara FAIR, både när de är tillgängliggjorda i en publiceringsinfrastruktur och när de är arkiverade i ett myndighetsarkiv.

Tidigare studier kring datapublicering

I det här avsnittet presenteras tidigare undersökningar som är relevanta för rapporten och kartläggningen. Syftet är att ge en kontext till de metoder och resultat som tidigare undersökningar har genererat och hur de relaterar till det föreliggande arbetet. Undersökningarna har belyst olika aspekter av publicering av forskningsdata i både svenska och internationella kontexter, och de har i redovisningen delats in i tre huvudsakliga teman: *Repositorier och metadata*, *Öppen tillgång och FAIR* samt *Uppföljning och incitament*.

Studier kring forskningsdatapublicering i Sverige

Repositorier och metadata

En central fråga för att organisera ett välfungerande stöd och en infrastruktur för öppen publicering av forskningsdata av god kvalitet är *var* forskningsdata publiceras samt hur detta i sin tur påverkar möjligheten att hitta, återanvända och följa upp forskningsdata.

Tidigare undersökningar vid Linköpings universitet (LiU) [76], Karolinska Institutet (KI) [77] och Stockholms universitet (SU) [78], som alla initierats av universitetens stödfunktioner, har funnit att användningen av repositorier för datapublicering är fragmenterad, och att det saknas enhetliga riktlinjer för hur forskare ska publicera sina data. LiU:s kartläggning, som granskade hur och var man kan hitta datapubliceringar och hur många datapubliceringar som gjorts tillgängliga av LiU-forskare, konstaterade att affilierade forskare publicerar sina data i såväl generella repositorier – såsom Zenodo, Figshare och Dryad – som i ämnesspecifika databaser, beroende på forskningsområde och datamaterialets karaktär. I rapporten sammanfattas att det inte finns en tydlig standard för vilket repository en forskare bör välja, vilket leder till att data sprids över många olika plattformar [76].

Viktiga förutsättningar för att kunna spåra forskningsdata är metadata och beständiga identifierare. Metadata gör det möjligt att spåra datapubliceringar över tid och underlättar sökbarheten, men kvaliteten på metadata varierar kraftigt mellan olika repositorier och publiceringsplattformar. En kartläggning av var KI:s forskare publicerar data, och undersökning av metodutveckling för att följa upp detta, fann att många av lärosätets datapubliceringar saknar fullständiga metadata, vilket gör dem svåra att identifiera. För att lokalisera och beskriva datapubliceringar används beständiga identifierare som DOI, ORCID och ROR, men dessa implementeras inte konsekvent i alla repositorier [77].

Vid SU har man utvecklat ett verktyg kallat *Harvest Combine* för att skörda metadata och konvertera datafiler från olika repositorier. Detta arbete har belyst tekniska utmaningar kopplade till långtidsbevarande av forskningsdata och dess metadata. I en artikel beskrivs hur metadata skördas från olika repositorier och transformeras för att följa arkivstandarder, något som förbättrar både

sökbarhet och interoperabilitet. Däremot identifieras stora variationer i hur repositorer strukturerar dessa metadata och huruvida de följer metadatastandarder eller ej, vilket försvårar integration mellan olika system [79].

Sammanfattningsvis visar tidigare undersökningar av forskningsdatapublicering vid svenska lärosäten att valet av repositorium och uppgifterna som anges i metadata spelar en avgörande roll för att kunna lokalisera och arkivera datapubliceringarna, men att bristande standardisering och den stora spridningen av datapubliceringar mellan olika repositorer och infrastrukturer gör det svårt att få en enhetlig bild av var data publiceras, hur de kan hittas och hur olika system kan kopplas samman för att stödja interoperabilitet och effektiv datahantering.

Öppen tillgång och FAIR

Öppen tillgång och FAIR-principerna är centrala begrepp i kravställning såväl i den svenska forskningspolitiska diskussionen och regelverket, som i flera lokala regelverk hos forskningsorganisationerna. Detta innebär att forskningsdata ska göras tillgängliga och återanvändbara. Svenska forskningsorganisationer förväntas redovisa hur dessa principer implementeras, men undersökningar visar att uppföljningen av öppen tillgång och FAIR står inför flera utmaningar.

När KI [77] och SU [78] undersökte datapubliceringar vid respektive lärosäte, framkom att många vetenskapliga publikationer saknar tydliga redogörelser för dataåtkomst (*data availability statements*), beständiga identifierare som DOI för datamaterialet eller länkar till externa källor. Detta minskar både sökbarhet och återanvändbarhet för tillhörande forskningsdata. SU:s kartläggning visar att även när en DOI anges är det inte alltid tydligt om den faktiskt leder till en datapublicering, och att osäkerhet råder kring vad som räknas som forskningsdata inom olika discipliner [78]. En stor utmaning som diskuteras i flera av undersökningarna är att det saknas standarder och riktlinjer kring tillgängliggörande av och hänvisningar till forskningsdata, vilket leder till att data publiceras med inkonsekventa metadata och i varierande format. Detta blev synligt i KI:s undersökning där man såg en hög grad av variation i formuleringar i redogörelser för dataåtkomst och hänvisningar till data [77].

Vetenskapsrådets (VR) årliga enkät om forskningsdata och öppen tillgång syftar till att följa upp svenska forskningsorganisationers implementering av öppen tillgång. Enkäten som genomfördes 2024 [91] gav även inblick i forskares perspektiv på öppen tillgång till forskningsdata. Enligt undersökningen uppger en stor andel forskare att de redan har (39 %), eller planerar att (51 %), tillgängliggöra sina forskningsdata öppet, och över 70 % upplever att det finns incitament att göra det. Sammantaget ger detta en bild av en positiv inställning till publicering av öppna forskningsdata. Samtidigt framgår det att många forskare (64 %) upplever att de saknar de praktiska förutsättningar som krävs för att prioritera datapublicering, såsom tillgång till systemstöd, tekniska lösningar och kompetensutveckling.

I samma undersökning anger flera lärosäten att de har tillgång till tillförlitliga och ändamålsenliga tekniska lösningar för arbetet, men majoriteten svarar att detta endast gäller *i viss utsträckning*. Det innebär att de befintliga lösningarna inte täcker de faktiska behoven. Därtill rapporterar organisationerna att tiden och resurserna för utsedd personal ofta inte räcker till. Denna problematik understryker behovet av att de tekniska resurserna är skalbara och tillgängliga i tillräcklig omfattning för att möta forskarnas växande behov. För att adressera dessa utmaningar rekommenderade Vetenskapsrådet i 2024 års rapport att kostnader och behov av resurstillskott för ett långsiktigt öppet tillgängliggörande av forskningsdata identifieras och beaktas i kommande planering.

Sammanfattningsvis framgår att öppen tillgång och FAIR-principerna utgör centrala riktmärken, men att tillämpningen i praktiken präglas av praktiska och strukturella problem. Trots forskarnas generellt positiva inställning till att publicera data kvarstår begränsningar i tid, resurser och systemlösningar, som krävs för att tillgängliggöra och hantera forskningsdata på ett öppet sätt i linje med FAIR-principerna. Resultaten pekar på behovet av mer samordnade och långsiktiga insatser för att målen om öppen tillgång och FAIR ska kunna uppnås i praktiken.

Uppföljning och incitament

Vetenskapsrådets årliga enkät *Öppen tillgång till forskningsdata 2025* [46], som följer upp svenska forskningsorganisationers arbete med att implementera öppen tillgång, visar att den nationella uppföljningen av öppna forskningsdata fortfarande är ofullständig även om vissa framsteg kan noteras sedan föregående år. Undersökningen visar att en stor majoritet av organisationerna fortfarande har svårt att uppskatta i vilken grad forskningsdata faktiskt har bedömts eller tillgängliggjorts öppet, vilket pekar på kvarstående svårigheter kring överblick och uppföljningsmöjligheter. Denna uppföljningsproblematik kan delvis förstås i ljuset av vad flera undersökningar visar, nämligen att det saknas etablerade standarder och riktlinjer för hur datapubliceringar ska formuleras och hänvisas till. Att organisationerna har svårt att rapportera kvantitativa siffror understryks också i rapporten, som lyfter behovet av samordnade processer och bättre systemstöd för att samla in uppgifter över tid.

KI:s undersökning [77] är ett konkret exempel på hur ett lärosäte försöker ta fram metoder för uppföljning av lärosätets datapubliceringar, delvis för att kunna besvara VR:s årliga enkät. I de rekommenderade metoderna organiserade efter VR:s frågeställningar konstaterar man att flera av VR:s nyckelfrågor inte går att besvara av två skäl: dels eftersom de juridiska och etiska bedömningar forskarna gör kring vilka data som kan delas öppet inte registreras centralt eller fångas upp i metadatafält, dels eftersom metadata i många repositorer är ofullständiga och varierar mellan plattformar. Detta gör att en ren repositoriesökning inte kan ge en heltäckande eller kvantifierbar uppföljning av vilka dataset som bedömts möjliga för öppen tillgång respektive publicerats enligt FAIR-principerna.

I LiU:s kartläggning (2023) [76] undersöktes bland annat DataCite Commons som ett möjligt verktyg för att identifiera och följa upp lärosätets datapubliceringar, i kombination med sökningar direkt i

repositorier som Zenodo, Figshare och Dryad. Studien påpekar att DataCite Commons endast omfattar poster med DOI registrerade via DataCite samt en begränsad del relaterade poster från Crossref. Dataset registrerade genom andra DOI-utfärdare inkluderas inte, vilket innebär att DataCite inte kan ge en fullständig bild av den samlade dataproduktionen. Därtill saknas ofta affilieringsmetadata¹¹ eller kopplingar till ROR-identifierare, vilket ytterligare begränsar möjligheten att söka på affilieringsnivå. Sammantaget visar studien att DataCite Commons är användbart för att lokalisera datapubliceringar med komplett DOI-registrering, men att det behöver kompletteras med andra källor för att uppnå mer heltäckande uppföljning.

När det gäller incitament befinner sig arbetet fortfarande i ett uppbyggnadsskede, men 2025 års VR-enkät visar vissa framsteg. Utifrån VR:s indikator 5 har 18 lärosäten (37 %) nu helt eller delvis integrerat öppna forskningsdata i sina incitamentssystem. Bland forskningsfinansiärerna uppger åtta av elva att de ställer krav på öppen tillgång till forskningsdata i vissa eller alla utlysningar, och två anger att de följer upp dessa krav genom datahanteringsplaner och FAIR-principer. I rapporten noterar VR att incitament för forskare att skapa öppen tillgång och FAIR-data fortfarande behöver stärkas, särskilt genom att göra sådant arbete meriterande. Flera aktörer lyfter att svårigheter kan kvarstå så länge forskningsmeritering främst baseras på vetenskapliga publikationer, men pekar samtidigt på ett växande nationellt arbete inom CoARA¹² för att utveckla mer ansvarsfulla incitamentsmodeller. Även i 2024 års rapport framhölls behovet av tydligare belöningsstrukturer, då 70 % av forskarna upplevde att det åtminstone delvis fanns incitament att dela data, medan resten svarade nej eller "vet inte", vilket tyder på en viss osäkerhet kring meriteringssystemen. Vetenskapsrådet betonar därför fortsatt vikten av att väva in öppna forskningsdata i lokala karriär- och meriteringssystem.

Tillsammans ger dessa resultat en bild av ett område i förändring, där både uppföljningsmetoder och incitamentsstrukturer successivt utvecklas. Flera undersökningar visar på några företeelser som försvårar tillförlitlig kvantifiering och uppföljning av datapublikationer; låg enhetlighet av metadata, skillnader mellan hur repositorier organiserar och beskriver data liksom avsaknad av standardiserade hänvisningssätt. Vetenskapsrådet betonar behovet av fortsatt samordning, tydligare belöningsstrukturer och stärkt systemstöd för att forskare ska ges incitament att tillgängliggöra och dokumentera sina data i enlighet med FAIR-principerna.

¹¹ Se även *Affilieringsinformation i Definitioner*.

¹² Akronymen CoARA står för Coalition for Advancing Research Assessment, se även *Bakgrund*.

Internationella studier kring forskningsdatapublicering

Repositorier och metadata

Internationella studier visar att även med etablerade infrastrukturer och standarder präglas metadatalandskapet av stora variationer i struktur, täckning och enhetlighet. Dessa skillnader påverkar möjligheterna att lokalisera, följa upp och förstå forskningsdata och relationer till andra forskningsprodukter. I detta avsnitt lyfts några centrala exempel på hur metadata och beständiga identifierare används och påverkar möjligheterna att lokalisera, följa upp och citera forskningsdata.

En studie av Johnston m.fl. [80] (*Seek and you may (not) find: A multi-institutional analysis of where research data are shared*) analyserade var forskare vid sex amerikanska forskningsintensiva universitet publicerar sina dataset. Studien byggde på metadata från DataCite och Crossref samt direkta sökningar i institutionella repositorier, och fann en ökad användning av unika disciplinspecifika repositorier och att de tio största repositorierna stod för 89–99,8 % av alla DOI-registrerade dataset. Trots detta missades många kända repositorier i de inledande sökningarna på grund av ofullständiga eller inkonsekventa affilieringsmetadata. Författarna föreslår ett antal praktiska förbättringar i rekommendationer som vänder sig till repositorier, metadataaggregatorer (som DataCite och Crossref) och skapare av datapubliceringar, för att främja spårbarhet hos datapubliceringar och förtydliga relationer mellan forskningsobjekt.

Liknande utmaningar noterades i en flamländsk fallstudie av Van Wettere m.fl. [81] (*Affiliation Information in DataCite Dataset Metadata: a Flemish Case Study*), som fann att även om användningen av ORCID och strukturerade affilieringsfält ökar, är länkar mellan dataset och relaterade publikationer ofta frånvarande. I ett tematiskt närliggande resultat i en finsk CRIS-kontext [82] visades att relationsmetadata mellan forskningsobjekt ofta saknas eller är ofullständiga, vilket innebär att de kan behöva kompletteras manuellt eller genom att läsa in ytterligare metadata från relaterade forskningsobjekt som redan finns utpekade med beständiga identifierare, s.k. PIDGraph¹³-metodik, för att fylla i luckorna. Detta överensstämmer med flera av källorna i den svenska kontexten, som med observationerna i LiU:s kartläggning [76] av fragmenterad repositorieanvändning och avsaknad av enhetliga riktlinjer, med KI:s [77] iakttagelser av ofullständiga affilierings- och relationsmetadata, och med SU:s [79] Harvest Combine-erfarenheter av stora variationer i hur repositorier strukturerar metadata.

Bristen på standardiserad PID¹⁴-användning lyfts även i en fransk undersökning av Bassinet m.fl. [83] (*Large-scale machine-learning analysis of scientific PDF for monitoring the production and the openness of research data and software in France*), där syftet var att utveckla den nationella resursen Open Science Monitor genom storskalig maskininlärningsbaserad textanalys av vetenskapliga

¹³ PIDGraph motsvarar kunskapsgrafer för forskningsinformation, se även *Definitioner*.

¹⁴ Akronymen PID står för *persistent identifiers*, beständiga identifierare på svenska. Se även *Definitioner*.

fulltexter. De refererar bland annat till tidigare resultat, som visar att under tio procent av artiklar med dataomnämningen innehåller beständiga identifierare för data, medan motsvarande identifierare för mjukvara inte påträffades i slumpurval. Detta illustreras i deras granskning av OpenAIRE Observatory, där de noterar att täckningen av datapubliceringar och mjukvara på land- och organisationsnivå är mycket begränsad och att metodbeskrivningen är knapphändig och brister i transparens. Författarna föreslår att *text mining* kan komplettera befintliga metadatakällor för att ge en mer heltäckande bild av datalandskapet. Fynden stämmer med KI:s [77] och SU:s [78] observationer av inkonsekvent PID-implementering och bristande länkar mellan data och publikationer. KI påvisar också att kompletterande metoder kan ge bättre täckning.

Ett praktiskt hinder på verktygsnivå är referenshanteringsprogrammen, där Vrouwenvelder m.fl. [42] (*Obstacles to dataset citation using bibliographic management software*), testar sju vanliga referenshanteringsverktyg (som EndNote och Zotero) på dataset från 14 repositorer. De visar att en majoritet av verktygen inte hanterar metadata för dataset tillräckligt väl vid import eller export, vilket ger ofullständiga eller felaktiga datareferenser och försvårar därmed efterlevnaden av etablerade riktlinjer för data citering. Endast ett fåtal verktyg importerade minst hälften av de centrala uppgifterna i datareferensen korrekt. Samtidigt identifieras betydande brister på repositornivå: rekommenderade citat på landningssidor saknar ofta flera centrala citatkomponenter och kan avvika från kärnmetadata i DataCite, och standardiserade exporter i .bib-format saknas ofta eller tilldelas fel källtyp i metadata. Författarna menar att detta utgör ett praktiskt hinder för utbredningen av data citering och i förlängningen begränsar möjligheten att spåra genomslag och skapa incitament för att publicera data. Författarna noterar också att etablerade riktlinjer för data citering är sporadiska och inkonsekventa, vilket försvårar implementeringen i befintliga forskningsinfrastrukturer. De rekommenderar att referensstilar och format uppdateras för att inkludera dataspecifika källtyper och viktiga citatkomponenter som åtkomstdatum och versionsinformation. Vidare bör repositorer säkerställa att exporter av referensmetadata harmoniserar med DataCite-metadataschemat och att de tillhandahåller maskinläsbara .bib-filer där källtypen för dataset är angiven i metadata.

Trots etablerade standarder präglas metadatalandskapet kring forskningsdata av stora variationer i struktur och enhetlighet, vilket hindrar spårning av data och formella relationer mellan forskningsprodukter. Kända repositorer missas ofta på grund av ofullständiga eller inkonsekventa affilieringsmetadata, och relationsmetadata, såsom länkar mellan dataset och resultatpublikationer, saknas ofta. Användningen av beständiga identifierare är inkonsekvent. Problem finns också på verktygsnivå, då majoriteten av referenshanteringsprogrammen inte hanterar metadata för datapubliceringar optimalt vid import eller export, något som också förvärras av felaktiga referensmetadata som levereras av infrastrukturer. Vidare avviker repositoriernas rekommenderade citat på landningssidor ofta från metadata som lagras hos de motsvarande DataCite-posterna. Dessa brister utgör tillsammans ett praktiskt hinder för utbredd data citering och spårbarhet.

Öppen tillgång och FAIR

Internationella studier pekar på en tydlig klyfta mellan policy och faktisk tillgång, samt på infrastrukturella och praktiska hinder som försvårar öppen tillgång och full FAIR-efterlevnad.

I en studie författad av Gabelica m.fl. [84] (*Many researchers were not compliant with their published data sharing statement: a mixed-methods study*) undersöktes forskares efterlevnad av deras redogörelser för dataåtkomst (*data availability statements, DAS*). Författarna kategoriserade dessa i åtta kategorier (sex enligt Springer Natures riktlinjer, kompletterade med två tillägg: "inte tillgängliga under några omständigheter" och "övrigt"). Av 3 416 artiklar med DAS var den vanligaste kategorin att data kan lämnas ut på rimlig begäran (1415, 41 %), och en mindre andel DAS behövde dubbel- eller trippelkodas (152 respektive 2 av 3416, totalt cirka 4,5 %). När författarna följde upp samtliga artiklar vars DAS angav att data kunde delas levererades data i endast 6,8 % av fallen. Detta ger en kompletterande bild av frågan om tillgängliggörande: formuleringen "på rimlig begäran" resulterar sällan i faktisk delning. SU:s undersökning [78] visar samtidigt att även när en beständig identifierare eller länk förekommer leder länkningen inte alltid heller till datasetet. Trots ett åttadelat klassificeringsschema behövde Gabelica m.fl. dubbel-/trippelkoda cirka 4,5 % av DAS, vilket speglar KI:s observation [77] av en stor variation av redogörelser för dataåtkomst och andra hänvisningar.

I undersökningen av Johnston m.fl. [80] bekräftas de stora skillnaderna i hur data i praktiken tillgängliggörs. Författarna noterar med hänvisning till relaterade delstudier att forskningsdata ofta delas på sätt som inte syns i DataCite eller Crossref, till exempel i supplementformat, på personliga webbplatser eller i repositorer som använder andra beständiga identifierare än DOI, vilket gör att dessa dataset inte fångas upp av DataCite eller Crossref. I diskussionen lyfter de behovet av bland annat gemensamma arbetssätt för databeskrivningar, för att underlätta sökning och åtkomst, samt konsekventa riktlinjer för hur DOI:er tilldelas och hänvisas till och hur olika relationstyper används för att uttrycka formella relationer mellan forskningsprodukter. Detta är nödvändigt för att tydliggöra samband mellan studiens komponenter och att utveckla möjligheterna att hitta data och öka kvaliteten i hur den delas.

Ett kompletterande användarperspektiv är forskarnas förståelse för och användning av beständiga identifierare (PID). Macgregor m.fl. [44] visar i sin undersökning *Measuring the concept of PID Literacy: User perceptions and understanding of PIDs in support of open scholarly infrastructure* att bland forskare finns en ojämn förståelse för PID-användning, även bland vanligt förekommande identifierare, där felkopplingar mellan PID-typer och entiteter var vanligt förekommande bland alla forskare oavsett disciplin. Samtidigt är attityderna till tekniken i grunden positiva, men när användningen kräver aktiva insatser upplevs beständiga identifierare ofta som arbetskrävande och komplexa. Författarna understryker att forskare inte bara är användare av beständiga identifierare utan även viktiga producenter av dessa, och föreslår riktade utbildningar och outreach till forskare för att öka återanvändningen av beständiga identifierare och på så sätt förebygga att olika infrastrukturer

som använder sig av PID-system, exempelvis kunskapsgrafer för forskning, förlorar viktiga hänvisningar till existerande forskningsprodukter och fenomen.

Variationen kring hur forskningsdata återanvänds, delas och kommuniceras kan även förklaras av forskningsdisciplinära skillnader. I en stor tvärdisciplinär enkät av Khan m.fl. [85] (*Data sharing and reuse practices: disciplinary differences and improvements needed*) framgår att delning och återanvändning skiljer sig kraftigt mellan forskningsämnen och påverkas av erfarenhet. 46,8 % uppger att de delar data online, med högre andelar inom t.ex. naturvetenskap (73 %) och geovetenskap (70 %) samt lägre inom ekonomi och företagsekonomi (33 %) och medicin (38 %). Forskare med längre forskarerfarenhet är mer benägna att återanvända data, och de forskare som återanvände data var således även mer benägna att dela data. Samtidigt rapporteras svårigheter att hitta återanvändbara datapubliceringar av de flesta forskare, med undantag för forskare inom fysik. Vid sökning efter datapubliceringar lyfter respondenterna främst god dokumentation (67 %), öppna data (52 %) och information om användbarhet (42 %) som avgörande, och det mest nämnda hindret är brist på kännedom om var och hur man söker. Många forskare (60,9 %) letar efter forskningsdata genom att läsa artiklar, och författarna hänvisar till tidigare studier som indikerar att avsaknad av meningsfulla åtkomstlänkar försvårar detta.

På ett liknande sätt varierar även vilka objekt som forskare hänvisar till när de citerar data. I en artikel av Gregory m.fl. [69] (*Tracing data: A survey investigating disciplinary differences in data citation*) undersökte författarna hur forskare citerar forskningsdata och såg där disciplinära skillnader. Enkätsvaren visar att forskare ofta refererar till resultatpublikationer när de vill peka ut relaterade data, men de hänvisar även till datakällor som repositorer eller datacenter och till faktiska data i form av dataset eller datafiler. Hänvisningsmönstren skiljer sig mellan discipliner: inom ingenjers- och teknikvetenskaper uppges hänvisningar till relaterade artiklar och data papers vara vanligare, medan forskare inom samhällsvetenskap och humaniora i högre grad anger hänvisningar som pekar ut själva datamaterialet.

Sammantaget visar studierna upp heterogena och fragmenterade praktiker för var och hur forskningsdata delas, återanvänds och citeras. Ambitioner om öppen tillgång och FAIR krockar ofta med bristande faktisk åtkomst, delning utanför PID-baserade infrastrukturer blir ofta osynlig och metadata görs sällan enhetliga, vilket försvårar sökbarhet och uppföljning.

Uppföljning och incitament

I detta avsnitt presenteras internationella studier som belyser olika sätt att följa upp forskningsdata samt hur incitament och citeringspraxis påverkar datapublicering. Genom exemplen visas hur metodval, metadata och forskningskultur tillsammans formar möjligheterna att mäta och synliggöra forskningsdata.

Två uppföljningsinsatser av forskningsdata på multiinstitutionell och nationell nivå visar hur man på två olika sätt kan fånga olika dimensioner av vad som går att mäta kring forskningsdata. I den multiinstitutionella flamländska fallstudien av Johnston m.fl. är uppföljningsmetoden grundad i att söka efter forskningsdata via affilieringsinformation (namnsträngar och beständiga identifierare för organisationer) i metadata. Trots att organisationsidentifierare gick att söka efter kunde författarna inte förlita sig enbart på denna utsökningsmetod, då användningen av dessa inte var utbredd när de utförde undersökningen. De såg dock ett ökat användande av dessa metadatafält i sina resultat.

I den andra nationella uppföljningsinsatsen i Frankrike använde författarna Bassinet m.fl. i stället en maskininlärningsbaserad textutvinning av fulltextartiklar (vars korpus de byggt baserat på DOI), med motiveringen att PID-baserade metoder inte ger realistiska utvärderingar och indikatorer på grund av låg användning av och kännedom om de beständiga identifierarna. Författarna menar att metoden de använder sig av har fördelen att man utifrån omnämnden i fulltexterna kan följa hur policyer för öppen vetenskap påverkar forskningspraktiker på nationell och institutionell nivå. Samtidigt nämner författarna tydliga begränsningar med metoden: den förlitar sig på tillgång till fulltexter med DOI där runt två tredjedelar av artiklarna kunde laddas ned. Metoden för att känna igen omnämnden av forskningsdata hade också endast stöd för engelsk text, och i och med att metoden bygger på omnämnden i artikeltexten missas data som publicerats utan koppling till (minst) en artikel.

Ovanstående undersökningar har visat hur man med hjälp av metadata, beständiga identifierare och textutvinning kan följa upp aktiviteten kring forskningsdata: det vill säga hur mycket som delas, skapas och används. Ett närliggande kompletterande uppföljningsperspektiv handlar om i vilken utsträckning forskare får erkännande och genomslag för sina datapubliceringar genom citeringar. Sådana uppföljningar kräver i regel tillgång till databaser med citeringsindex som kan spåra kopplingar mellan datapubliceringar och resultatpublikationer. I en jämförande studie av dessa typer databaser har Gerasimov m.fl. [86] (*Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Crossref, and DataCite*) undersökt hur väl publikationer som citerat cirka 11 000 dataset i Earth Observing System Data and Information System (EOSDIS) fångats upp. I sin undersökning identifierade de 17 095 publikationer som citerade datapubliceringarna. Resultaten visar att ingen enskild databas gav en fullständig bild av citeringslandskapet, då endast omkring en tredjedel av de citerande publikationerna återfanns i mer än en källa, medan varje databas innehöll ett betydande antal unika träffar. Kombinationer av två databaser, som Google Scholar och Scopus, ökade täckningen, men först när tre eller fler användes tillsammans närmade sig resultaten en mer heltäckande bild. Google Scholar hade högst täckningsgrad (cirka 70 %), medan DataCite hade lägst (cirka 19 %). Crossref uppvisade något större täckning (33 %) än DataCite men fångade endast omkring en tredjedel av alla citeringar, trots att majoriteten av dess poster innehöll referenser till datapubliceringar med DOI. Författarna förklarar detta med ofullständig referensdeponering från förlagen och att de tekniska system som används för att identifiera citeringslänkar inte återger alla registrerade referenser. Eftersom DataCite hämtar sina citeringslänkar från Crossref speglar dess

täckning samma begränsningar, men båda databaserna bedöms ha stor utvecklingspotential i takt med att fler förlag börjar registrera referenser till datapubliceringar i Crossref och direkta citeringsrelationer successivt tillförs DataCite. Mot denna bakgrund rekommenderar författarna att analyser av dataciteringar baseras på flera bibliografiska källor parallellt, snarare än på enskilda index, för att ge en mer heltäckande och tillförlitlig bild av datasetens genomslag.

Vidare kan även olika discipliners citeringspraxis påverka den information om forskningsdata som går att utvinna. Som tidigare nämnts undersöker Gregory m.fl. i sin artikel [69] hur forskare hänvisar till forskningsdata, med resultat som visar tydliga skillnader i kunskap om datacitering och var hänvisningar skrivs. I materialet framgår att forskare inom humaniora oftare placerar hänvisningar till data i fotnoter medan samhällsvetenskapliga forskare oftare hänvisar i brödtext, till skillnad från forskare i övriga discipliner som uppger att de hänvisar till forskningsdata i referenslistan.

Undersökningen visar också på en begränsad kännedom om standarder för datacitering generellt. Respondenterna känner bäst till och använder riktlinjer från tidskrifter och förlag samt etablerade referensstilar som APA och MLA, medan särskilda rekommendationer för datacitering från till exempel DataCite eller vetenskapliga samfund är mindre kända. Även här finns disciplinära skillnader: forskare inom samhällsvetenskap uppvisar lägre kännedom om särskilda dataciteringsstandarder, med undantag för väletablerade referensstilar som de känner till och använder i högre grad. Forskare inom naturvetenskap och jordbruk har däremot högre kännedom och användning av olika citeringsrekommendationer.

I samma artikel har författarna även undersökt vad som motiverar forskare till att citera forskningsdata. Generellt svarar forskare i linje med vad som kan anses vara god forskningssed, det vill säga att ge erkännande, visa tydligt var forskningsmaterialet finns och kunna validera sina forskningssatsar. Över lag är det inte heller vanligt att forskare citerar forskningsdata för att kritisera eller korrigera någon annans data, förutom inom humaniora där det är mer vanligt än inom andra discipliner. Både inom denna undersökning och Dorta-González m.fl:s undersökning [87] om datadelning och motivering (*To what extent is researchers' data-sharing motivated by formal mechanisms of recognition and credit?*) finns det hos forskare en starkt positiv inställning till att få sina forskningsdata citerade. I den senare nämnda undersökningen finner författarna att formella erkännanden hänger ihop med faktisk delning. Ju högre värde forskare tillskriver datacitering, desto oftare delar de data. Forskarna som uppger att de har fått erkännande för delning rapporterar också en högre delningsfrekvens än de som aldrig fått erkännande. En majoritet av forskarna anser samtidigt att datadelning får otillräckligt erkännande, även om den uppfattningen minskar bland de forskare som delar allra mest. Studien visar även att erfarenhet av att återanvända öppna data samvarierar tydligt med hur ofta man gör sina egna data öppet tillgängliga.

Internationella studier visar att uppföljning av forskningsdata är metodologiskt och infrastrukturellt komplext. Metadata, beständiga identifierare och citeringspraxis påverkar i hög grad vad som går att mäta och hur väl forskningsdata kan spåras över tid. I de analyserade studierna framkommer att

kombinationer av flera datakällor krävs för att närma sig en heltäckande bild av dataciteringar och användning, och att både PID-baserade och textutvinningsbaserade metoder erbjuder olika styrkor och begränsningar för detta ändamål. Samtidigt belyses att incitament och erkännande för att publicera data fortfarande är svagt utvecklade, även om forskare som får formellt erkännande tenderar att dela data oftare. Disciplinära skillnader i citeringspraxis och varierande metadatakvalitet kvarstår som centrala hinder för tillförlitlig uppföljning och synliggörande av forskningsdata.

Kartläggning

Denna kartläggning har haft som målsättning att undersöka hur svenska forskningsorganisationer idag arbetar med att spåra datapubliceringar, och även att fånga in mer kvalitativa redogörelser för vilka problem som upplevs och hur respondenterna ser att de kortsiktigt eller långsiktigt kan få en lösning.

Metod

För kartläggningen konstruerades en enkät som i första hand riktade sig till personer som på olika sätt arbetar med forskningsorganisationernas stödverksamheter. Enkäten fick titeln *"Att publicera och hitta forskningsdata som skapats vid svenska lärosäten"*. Länken till enkäten distribuerades via olika professionella nätverk, funktionsadresser och sändlistor för att särskilt kunna nå personer som arbetade med forskningsdata, publiceringsfrågor, bibliometri och forskarstöd, och enkäten gick att besvara mellan 2025-03-03 och 2025-03-17. Efter enkätperiodens slut hade 54 respondenter inkommit med svar.

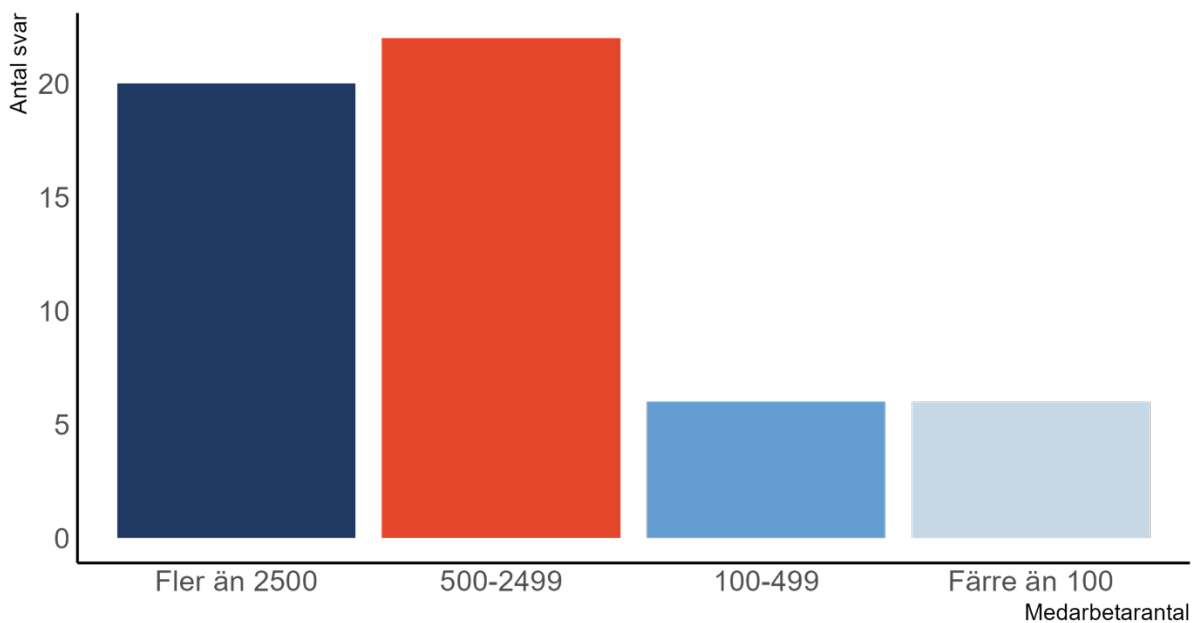


Fig. 1. Hur många medarbetare har den organisation som du är verksam vid?

En stor del av svaren inkom från större forskningsorganisationer med fler än 500 medarbetare. Detta kan delvis förklaras av valen av distributionskanaler som användes för att rekrytera respondenter, men också av att större organisationer kan ha fler medarbetare som aktivt arbetar med eller intresserar sig för den typ av frågor som enkäten handlade om och således kan vara mer benägna att svara.

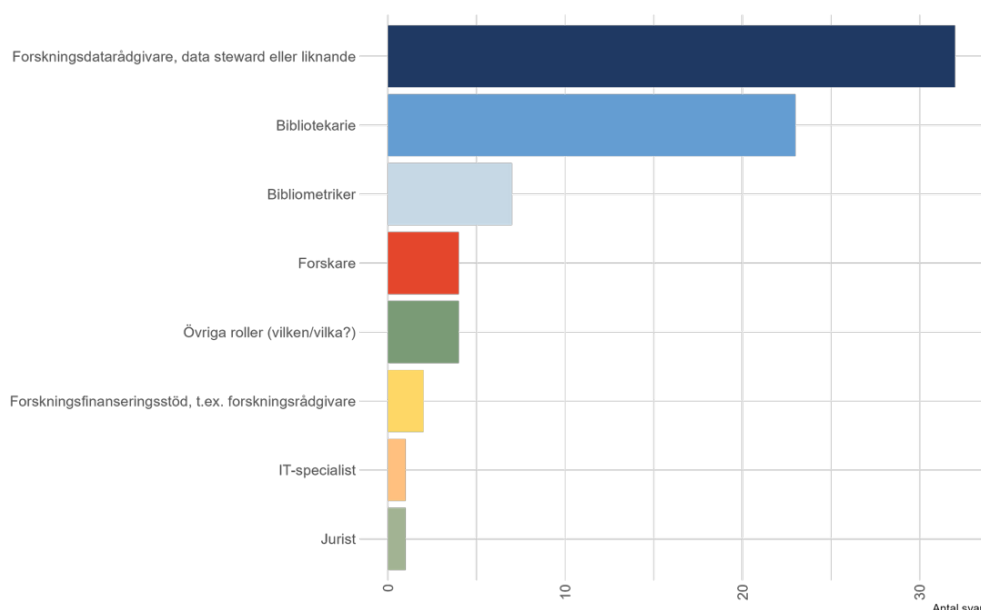


Fig 2. Respondenternas självrapporterade roller. Det var möjligt att ange en eller flera roller.

De roller som respondenterna uppgav utgjordes till stor del av befattningar som har eller kan ha forskningsdatafrågor som en del av sitt uppdrag, som bibliotekarier eller data stewards. Flera bibliometriker svarade också, vilket var väntat med tanke på frågeställningen. Endast ett fåtal svar inkom från forskare. Bland övriga roller fanns fritextangivna titlar som chef eller koordinator.

Insamling och design

Enkätkonstruktion och administration genomfördes i Göteborgs universitets instans av Sunet Survey [88]. Frågorna formulerades i stor utsträckning som fritextfrågor, för att kunna utgöra ett komplement till den mer kvantitativt inriktade enkät baserade på förbestämda indikatorer på organisationsnivå, som Vetenskapsrådet skickar till alla lärosäten varje år [46]. I vår enkät fanns sju kategoriska frågor med, medan de resterande sju frågorna gav möjlighet att utveckla sig utförligt i fritext. Frågeformuläret finns bifogat som dokumentation till datapubliceringen [89].

Enkäten designades med inställningar för anonym svarsfunktion och hade endast enstaka frågor som yrkesroll och organisationsstorlek som bakgrundsvariabler. Utöver detta fanns dock stora möjligheter att utveckla fritextsvar om lokala förhållanden. På grund av att respondenter nämnde unika tjänster och funktioner knutna till en organisation, och i vissa fall organisationens namn i klartext, var det ibland möjligt att spåra vissa respondenter till arbetsplats. Med särskilda förkunskaper kan det sannolikt också gå att bakvägsspåra vissa individer genom enkätsvaren.

För att minska exponeringen för identifierande uppgifter under analysfasen genomfördes ett förberedande genomläsningspass där fritextuppgifter från enkätsvaren delvis maskerades genom att utpekande uppgifter ersattes med generiska formuleringar, exempelvis för lärosätesnamn eller namn på unika lokala tjänster.

Analysmetod

Enkätanalysen genomfördes i form av en tematisk analys, som inleddes genom en kvalitativ tematisk kodningsmetodik där enkätsvaren lästes igenom och koder utarbetades för olika teman. I arbetet med att ta fram kodning användes den öppna kvalitativ dataanalys-mjukvaran *Taguette* i version 1.4.1, som installerades på en server på Göteborgs universitet och möjliggjorde att flera projektdeltagare kunde samarbeta via ett webbgränssnitt med kodningen [90].

Problem.Ledning
Problem.Ledning.AvsaknadInrapportering
Problem.Ledning.AvsaknadPolicy
Problem.Ledning.AvsaknadRekommendationer
Problem.Ledning.AvsaknadStrategi
Problem.Ledning.AvsaknadUppdrag
Problem.Ledning.AvsaknadUppföljning
Problem.Ledning.ComplianceFokus
Problem.Ledning.KunskapsbristOrganisation

Fig. 3. Exempel på en av grupperna av tematiska koder som användes under analysfasen (Problem.Ledning).

Arbetet med att skapa tematiska koder resulterade i ett hierarkiskt träd med tre nivåer. På den högsta nivån hamnade koder för *Krav*, *Problem* och *Lösning*. De tematiska koderna gav sedan stöd för att gruppera anteckningar kring hur respondenterna beskrev eller förhöll sig till olika fenomen relaterade till frågeställningarna, och de olika problembeskrivningar och lösningsförslag som de lade fram i enkätsvaren. Slutligen sammanfattades en reflekterande analystext som en sammanställning av de olika teman som identifierats. Kodboken finns bifogad som dokumentation till datapubliceringen [89].

Datapublicering

Enkätsvaren och den tematiska kodningen som applicerats på dem finns tillgängliggjorda som en datapublicering med begränsad åtkomst och öppen dokumentation [89]. Åtkomstbegränsningen gjordes med hänsyn till möjligheterna att bakvägsidentifiera respondenter.

Analys och diskussion

Enkätsvarens olika teman har här blivit indelade i följande breda kategorier:

- Infrastruktur och processer
- Standardisering och interoperabilitet
- Nationell samordning
- Ledning och resurser
- Uppföljning och incitament
- Vägen mot en öppen vetenskapskultur

Det bör dock understrykas att många teman och ämnen ofta går in i varandra och kan ha bäring på flera olika aspekter av exempelvis en problembild. Dessa kategorier bör därför inte ses som exklusiva klassificeringar, utan är huvudsakligen ett hjälpmedel för att strukturera texten.

Infrastruktur och processer

Flera av respondentsvaren anknyter till problematik som orsakas av processerna för hur datapubliceringar ofta skapas och används i praktiken. En del av problematiken styrs också av de infrastrukturer som skapar eller begränsar valmöjligheterna för hur datapubliceringen kan genomföras.

Ett problemområde som kan identifieras med hjälp av enkätsvaren är att forskare, som citerar till exempel en vetenskaplig artikel, inte med samma självklarhet citerar en datapublicering. Detta kan ha många orsaker, men kan delvis antas bero på till att det saknas utbredda riktlinjer och praktiker för hur forskaren gör sin datapublicering citerbar. Andra faktorer kan vara särskilda regler från förlag som begränsar spårbarheten, eller hävdvunna traditioner för hänvisningar till data inom olika forskningsfält. Att göra citeringsanalyser på publikationstypen datapubliceringar ger därför mindre information än samma analys för vetenskapliga resultatpublikationer.

Inglehart, R. (1997). *Modernisation and Postmodernisation: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.
ISSP Research Group (2023). *International Social Survey Programme: Environment IV - ISSP 2020*. GESIS, Cologne. ZA7650 Data file Version 2.0.0, <https://doi.org/10.4232/1.14153>.
Karadja, M., Mollerstrom, J., & Seim, D. (2017). Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution. *The Review of Economics and Statistics*, 99(2), 201–212. https://doi.org/10.1162/REST_a_00623

Fig. 4. Ett exempel på en referens till en datapublicering inlagd i en resultatpublikations referenslista. Att forskningsdata blir citerade på samma sätt som andra källor är inte en självklarhet.

I en vetenskaplig artikel kan exempelvis hänvisningar till data som används helt saknas, enbart göras med omnämnande i löptext, eller endast göras indirekt genom hänvisningar till andra publikationer som i sin tur relaterar till data. Formella referenser kan saknas såväl till data som man själv har samlat in som till data som tillgängliggjorts av andra, även i fall då en citerbar datapublicering faktiskt finns tillgänglig. Även om fullgoda referenser till data faktiskt har gjorts anger respondenterna att

datapubliceringarna ändå ofta blir exkluderade ur sådana metadata om forskningsprodukter som gör dem spårbara. Detta kan bero på orsaker som begränsningar i funktionalitet och täckning hos metadataindexerande tjänster, eller på att tidskrifter särbehandlar referenser till insamlade data och hänvisar dem till särskilda redogörelser för dataåtkomst (*data availability statements*).

"När forskare publicerar data (som inte är publicerade i SND) och vi inte är kontaktade så har vi ingen koll alls."

"Metadata samlas inte in på samma sätt som för publikationer där vi ofta hittar metadata [...] via Web of Science och Scopus."

Läget innebär att datapubliceringarna av flera orsaker systematiskt blir exkluderade ur de informationsflöden som respondenterna har möjlighet att kontinuerligt bevaka forskningsprodukter i. Samtidigt finns det oftast inte någon policy där individuella datapubliceringar aktivt måste redovisas hos forskarens organisation. Det leder till att den som vill spåra datapubliceringar kan behöva använda tidskrävande granskningsmetoder som sällan är realistiska för mindre organisationer, såsom *text mining* eller omfattande manuell granskning av resultatpublikationers innehåll.

Sammantaget antyder respondenterna alltså att nuvarande praktiker och processer leder till stora svårigheter för forskningsorganisationer att leva upp till lagkraven om arkivering och öppna data på grund av problem med att kartlägga och följa upp datapubliceringarna, eller kort sagt svårigheter att alls få reda på att dessa publiceringar existerar.

"Svårt att hitta data som delas som supplementary data."

"Hur ska vi förhålla oss till data som "bara" ingår i artiklar eller böcker, kontra data som har egen DOI? Vill man verkligen ha hela bilden skapar den "osynliga" datan som bakats in i publikationerna utmaningar."

En problematik knuten till etablerade processer inom vetenskaplig publicering är systemet med data i supplementformat som på olika sätt bifogas en artikel, men inte blir självständigt adresserbart, inte får beständiga identifierare och egna metadata och därmed inte dyker upp i någon sökindexering, även om det skulle kunna klassificeras som en datapublicering.

Här påtalar några respondenter också en avgränsningsproblematik där det blir oklart vad som räknas, är supplementet en del av resultatpublikationen? Inom vissa discipliner är det också vanligt att hela datamängden redovisas just som en integrerad del av en ordinarie publikation på samma sätt som av tradition gjorts i trycksaker, exempelvis form av tabeller, kataloger och liknande. Vad ska i slutändan räknas som en datapublicering?

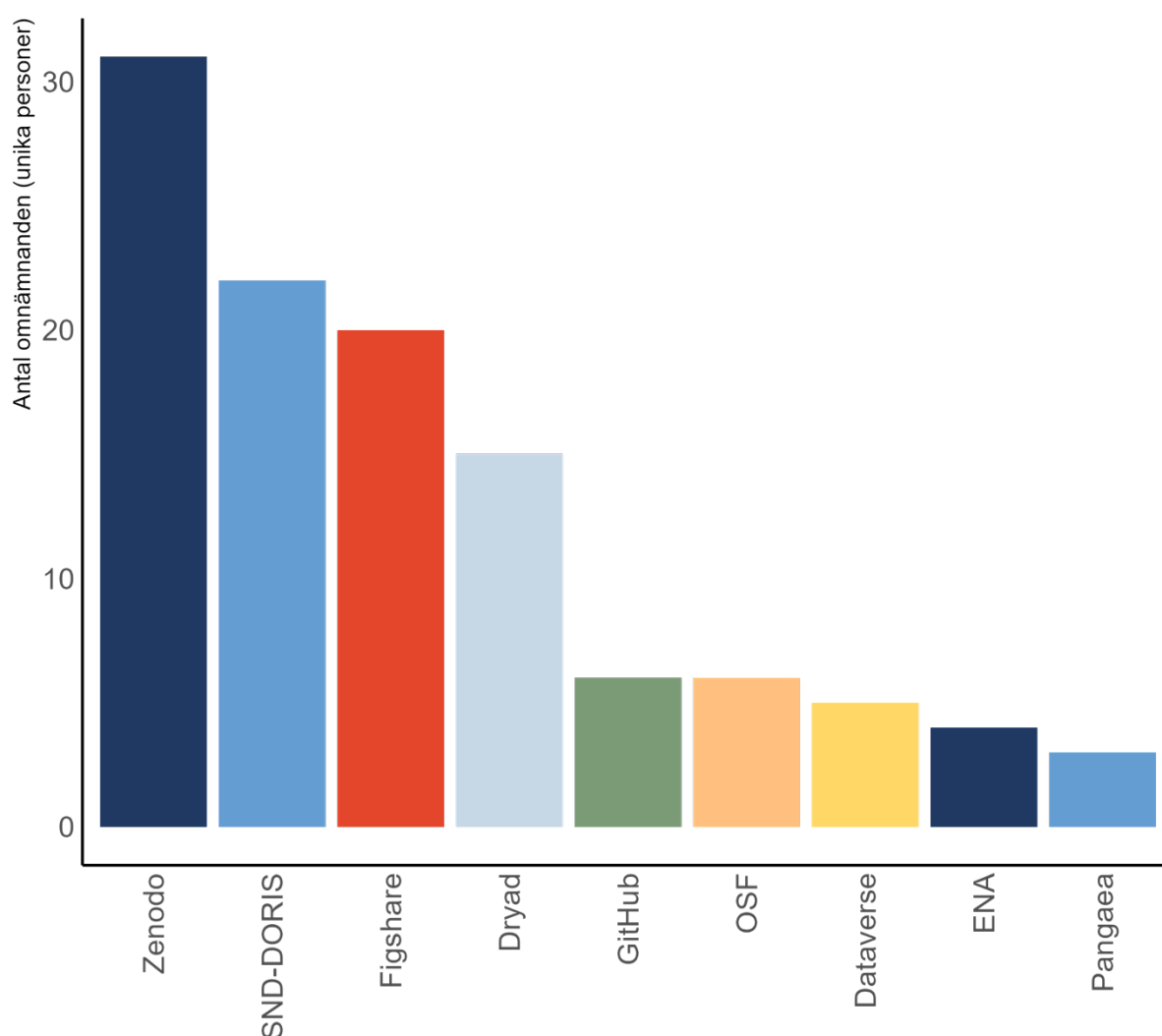


Fig 5. Vilka repositorier eller andra publiceringsplattformar för forskningsdata som era forskare publicerar sig i känner ni till? (Figuren visar omnämningen från unika personer. Flera svar per person var möjliga).

Infrastrukturerna för publicering som omnämns i enkätsvaren följer en distribution där populära repositorier som tillåter självpublicering utan granskningsprocess, såsom Zenodo och Figshare, hamnar i toppen. Flera respondenter nämner dock att de i dessa använder sig av, eller planerar att använda sig av, funktionalitet som communities eller organisationsspecifika instanser som kan ge ökade möjligheter för lokal granskning och kvalitetskontroll. Många nämner också användningen av SND:s publiceringsverktyg DORIS för att tillgängliggöra data i de olika repositoriellösningar som är integrerade med verktyget. Disciplinpassade repositorier, som ENA för genetiska data eller Pangaea för geovetenskapliga data omnämns också, men de mer specialiserade tjänsterna har en mer spridd och varierad distribution, som kan antas bero på lokala verksamheters disciplinspecifika behov och den individuella respondentens expertis.

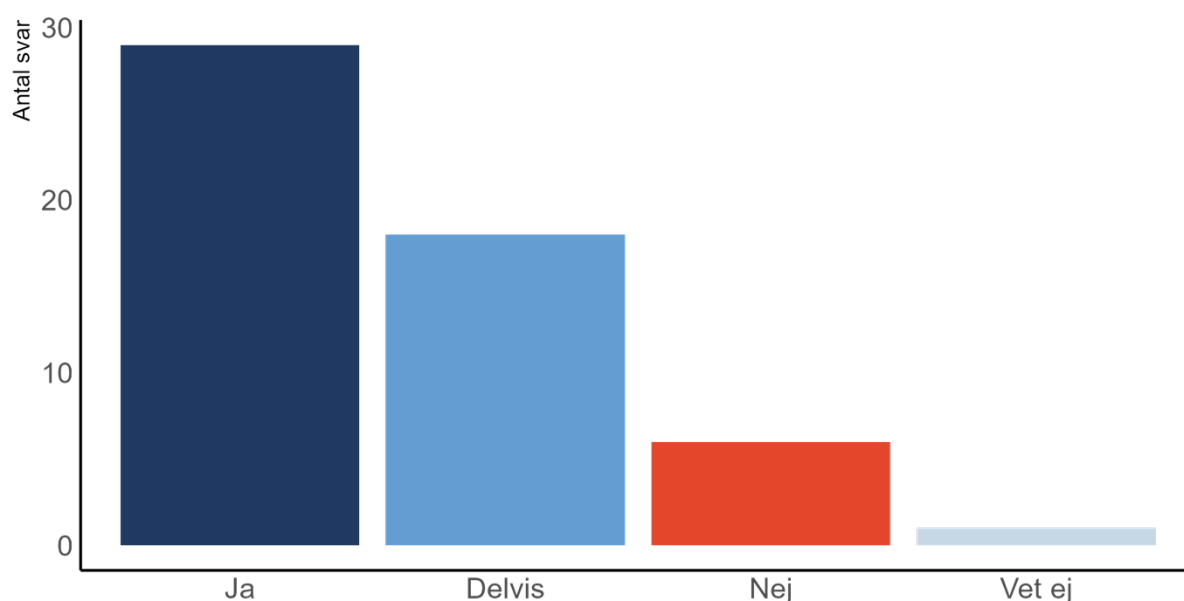


Fig 6. Arbetar din organisation med aktiv kurering eller granskning av datapubliceringar och tillhörande metadata?

En majoritet av respondenterna uppger att deras organisation på något sätt arbetar med aktiv och granskning och kurering av datapubliceringar på central nivå. Detta kan dock med utgångspunkt i fritextsvaren antas vara avgränsat till sådana infrastrukturer som centrala funktioner aktivt kan bevaka och har behörigheter i, exempelvis SND:s DORIS-verktyg, en lokal Figshare-instans eller Zenodo-communities för organisationerna som forskare själva anmäler material till.

”Det kan uppfattas som onödigt omständligt och krångligt att publicera, alternativt att man inte känner till möjligheten att publicera metadata trots att man har känsliga data som inte kan publiceras öppet.”

Flera respondenter lyfter att infrastrukturer och processer för datapublicering generellt upplevs som svåra att förstå sig på och att systemen är tidskrävande att arbeta i för forskarna. Problematiken kan dels handla om UX- och gränssnittsfrågor, men dels också om svårigheter att förhålla sig till PID-system, strukturerade metadata och kontrollerade vokabulärer utan att ha arbetat systematiskt med koncepten tidigare i forskningsprocessen eller karriären. Detta kan leda till sämre kvalitet på metadata, begränsad spårbarhet för datapubliceringar och att FAIR-principerna inte efterlevs.

”Ett bättre flöde för forskare så att det inte upplevs som krångligt att fylla i metadata - kanske kan information hämtas från andra system/dokumentation de redan tidigare i projektet skapat, tex från datahanteringsplaner?”

”Jag tror att det är viktigt att system som stödjer datahantering, som verktyg för projekthantering och datahanteringsplaner, aktivt ledsagar forskarna på ett sätt som gör att forskningsdata får t.ex. PID:ar och grundläggande metadata när de publiceras”

Problemen med att gränssnitt och processer för datapubliceringar upplevs krångliga föreslås av respondenterna kunna angripas från flera håll, utöver arbete med renodlade UX-frågor. Dels krävs vägledning och stödfunktioner för datahanteringen genom hela forskningsprocessen, oavsett om det gäller praktiska datahanteringsfrågor, teknik, juridik eller ansvarsfrågor. Dels finns också förslag om att i större utsträckning knyta samman system och verktyg som används för att hantera forskningsinformation, såsom projekthantering, CRIS-system och verktyg för datahanteringsplaner på ett sätt som gör att grundläggande metadata av rutin skapas tidigare i forskningsprojektets livscykel. Dessa metadata kan sedan återanvändas för att underlätta skapandet av spårbara datapubliceringar i publiceringsverktygen. En ökad integration av organisationernas lokala system med publiceringsverktyg kan också bidra till att ansvarsförhållanden och arbetsflöden förtydligas, samtidigt som det skapar ett naturligt informationsflöde där datapubliceringar kan bevakas.

”Det är krångligt och tidskrävande att tillgängliggöra forskningsdata med begränsning, därför att den infrastruktur som behövs för det är bristande eller saknas.”

Den infrastruktur som finns för att skapa helt öppet tillgängliga datapubliceringar är generellt sett mest utvecklad. Men flera respondenter framhåller att stödinfrastrukturen för att kunna genomföra datapubliceringar med just *begränsad åtkomst* upplevs vara bristfällig, vilket leder till merarbete och bristande spårbarhet. Exempelvis har inte alla forskningsorganisationer etablerat tydliga arbetsflöden och vägledning för åtkomstbegränsade publiceringar av skyddsvärda data, så att forskarna på egen hand måste försöka utreda hur processen eventuellt kan genomföras. Utöver en osäkerhet kring juridik finns således tveksamhet kring hur man praktiskt går till väga.

Utan infrastrukturstöd, godkänd lagring och tydlig vägledning ger många forskare upp försöken att genomföra datapubliceringar med begränsad åtkomst och hemfaller kanske till svepande formuleringar i redogörelser för dataåtkomst (*data availability statements*). Alternativt följer man kollegors förslag på utländska repositoriellösningar, som även om de resulterar i en spårbar datapublicering kan stå i konflikt med krav som finns på forskningsverksamhet i Sverige, exempelvis gällande personuppgiftshantering, sekretess och informationssäkerhet. Utöver detta så skapar lösningarna för att publicera data med åtkomstbegränsningar inte alltid digitala objekt som är FAIR och kan citeras och spåras separat.

Många respondenter nämner bristen på lämpliga och långsiktigt finansierade lagringslösningar, en problematik som drabbar alla led av forskningsdatahanteringen. För publicering av åtkomstbegränsade data saknas ofta bestående lagringslösningar där skyddsvärda data som hör till en datapublicering kan förvaltas över tid. Men problemet är genomgående för hela livscykeln för forskningsdata, där forskare ofta hänvisas till att finansiera tillfällig lagring med projektmedel, samtidigt som det ofta saknas utpekade lösningar för att bevara data med finansiering som långsiktigt uppfyller forskningsorganisationernas arkiv- och förvaltningsansvar.

Ett antal respondenter påtalar hur sårbart det är att den egna forskningsorganisationen saknar ett lokalt CRIS/RIMS-system, som kan förvalta och dela information om datapubliceringar vid sidan av andra forskningsprodukter. Ett återkommande tema i enkätsvaren är också just stödet för skapande av återanvändbara metadata för att förenkla administration och ge bättre vägledning. Dessa metadata kan också ofta delas i form av öppen forskningsinformation, som kan visa hur forskningen hänger samman, och ge grundförutsättningar för synlighet och erkännande som kan skapa incitament för att göra bra datapubliceringar. Ett återkommande önskemål på förbättring som förmodas förenkla arbetsflöden och bidra till skapandet av viktiga metadata som sedan kommer till användning i hela forskningsekosystemet är att tydliga kopplingar görs mellan planering i verktyg för datahanteringsplaner, beskrivningar i lokala CRIS-system och den slutliga datapubliceringen i ett rekommenderat repository tillsammans med arkivering i lokalt e-arkiv.

Standardisering och interoperabilitet

Flera respondenter tar upp frågor som handlar om interoperabilitet, jämförbarhet och följsamhet med olika standarder. Dessa påverkar direkt hur forskningsinformation kan återanvändas, exempelvis för att spåra datapubliceringar.

"Olika system har olika krav på metadata och fullständighet"

"Det finns inga rutiner för att skörda metadata om dataset och de standarder som finns (t.ex. DataCite) implementeras olika av olika repositoryer".

"en majoritet av publicerade dataset saknar [...] grundläggande metadata kring "användbarhet", som t.ex. licenser, ursprungligen tänkt användningsområde, etc."

Forskningsdatainfrastrukturer anges lida av en brist på standardisering, särskilt vad gäller metadata. Affilieringsinformationen är inte alltid obligatorisk eller sökbar. Det saknas också ofta metadatafält för centrala uppgifter som ansvarig forskningshuvudman eller primär affiliering för medverkande. Olika aktörer tillämpar ofta egna lösningar och prioriterar olika, vilket försvårar återanvändning och delning av metadata. Även etablerade standarder, såsom *DataCite Metadata Schema*, tolkas subjektivt och implementeras olika. Detta leder till variation i omfång, innehåll och kvalitet hos metadata, och således också till att informationen inte alltid blir jämförbar och systematiskt återanvändbar.

Skillnader i minimikraven i olika infrastrukturer kan leda till att grundläggande information som ursprung, syfte, licenser och återanvändningsvillkor saknas. Detta skapar hinder för både spårbarhet och praktisk återanvändning av datapubliceringen. De metadatastandarder och verktyg som används kan ibland endast stödja inmatning av uppgifter som är allt för grundläggande för att fylla de unika krav på klassificering, urval och beskrivning som kan behövas inom en specifik disciplin. Några respondenter pekar också ut att det för automatiserade processer inom AI och maskininlärning är särskilt viktigt med metadata för licenser och restriktioner, utöver de uppgifter som gör att processen kan klassificera datapubliceringen. Dessa uppgifter är direkt avgörande för om datamaterialet inkluderas eller exkluderas i automatiserade urval.

Tillgången till data kopplade till en vetenskaplig artikel beskrivs oftast i en så kallad redogörelse för dataåtkomst (*data availability statement*). Dessa beskrivningar är inte standardiserade och sällan tillgängliga som maskinläsbara metadata, vilket leder till att de ofta är otydliga och svåra att spåra och följa upp. Olika förlag kan ha egna riktlinjer kring specifika former för dessa beskrivningar och villkoren i dem. Detta kan skapa konflikter, exempelvis med sättet som datapubliceringen är genomförd på eller med restriktioner grundade i svensk lagstiftning. Olika praktiker kring att aktivt citera datapubliceringar kan leda till att detta beskrivningsavsnitt i praktiken blir den enda platsen där datapubliceringen faktiskt omnämns, samtidigt som inga strukturerade metadata som hänvisar till datapubliceringen görs tillgängliga.

"våldigt få dataset är idag beskrivna med hjälp av semantisk webb/linked dataresurser som t.ex. kontrollerade vokabulärer, thesaurier m.m. – detta utgör ett allvarligt hinder för att kunna såväl särskilja som kombinera olika dataresurser"

"Det är också ofta som man inte hänvisar från datamaterialet till eventuella publikationer som datamaterialet hänger ihop med[...]"

Formella relationer till andra entiteter saknas ofta i datapubliceringars metadata, exempelvis relationen till en tillhörande resultatpublikation. Även när möjlighet finns att uttrycka relationer till andra objekt och definitioner med beständiga identifierare eller kontrollerade vokabulärer så utnyttjas detta sällan. Vägledning för att använda sig av formella relationer och av etablerade kontrollerade vokabulärer, såsom ämnesord knutna till disciplinspecifika behov och avgränsningar, saknas ofta vilket gör att sådana sällan används för att beskriva datapubliceringar på ett konsekvent sätt. Då flera aktörer använder dessa relationer och definitioner för att klassificera material och länka ihop sammanhanget mellan forskningsdata och andra objekt som beskrivs i öppen forskningsinformation innebär detta att spårbarhet och synlighet, men också jämförbarhet med andra forskningsdata i många fall kan påverkas.

Nationell samordning

Flera av svaren anknyter till lösningar som är i behov av att samordnas nationellt, både i form av konkret teknik och infrastrukturer men också i form av mer mjuka frågor där samsyn behöver uppnås för att undkomma problem.

"Information om offentlig finansierad forskning i Sverige bör vara lättillgänglig, sökbar och komplett."

"...det finns ingen samordning eller samlad kommunikation alls kring detta"

En tydlig brist upplevs vara nationell samordning av öppen forskningsinformation. Om öppen, tydlig och maskinläsbar information om forskningsprojekt och de forskningsprodukter som skapas i dem finns tillgänglig, och i så fall i vilken utsträckning, kan variera stort mellan svenska forskningsorganisationer. Detta påverkar följaktligen också spårbarheten hos datapubliceringar.

Tillgängligheten till forskningsinformationen kan vara knuten till avgränsningar som gjorts av huvudmän, finansiärer, förlag och andra beslutsfattare, men också begränsas av mer indirekta skäl som val av mjukvaror, leverantörer, eller utformningen av organisationers webbplatser. Strukturerad forskningsinformation där relationerna mellan forskningsaktivitet (projekt), personer, organisationer, anslag, resultatpublikationer och forskningsdata går att utläsa finns sällan att ta del av i standardiserade och internationellt erkända metadataformat¹⁵. Inte minst saknas ofta kontinuerliga processer där hela spannet av forskningsinformation samlas in, kompletteras, kvalitetssäkras, förvaltas och delas vidare av forskningsorganisationerna själva. De processer som finns kan ofta vara avgränsade till enstaka områden, såsom information om resultatpublikationer. Detta begränsar möjligheterna att systematiskt arbeta med översikter och jämförelser på såväl det nationella som det internationella planet vilket pekar på ett stort behov av samordning.

Att konsekvent arbeta med beständiga identifierare framhävs som en viktig lösning för att koppla datapubliceringar till forskningshuvudmännen. Att PID-system som ORCID och ROR ID används systematiskt av forskare och organisationer ses därför av många respondenter som centralt för att kunna arbeta hållbart såväl med datapubliceringar som med andra forskningsprodukter. Hur formella relationer mellan forskningsprodukter och andra fenomen uttrycks i forskningsaktiviteternas relaterade metadata, exempelvis genom hänvisningar till just beständiga identifierare, är avgörande för hur svensk forskning visas i andra system. Det finns dock en mycket stor variation i hur dessa relationer praktiskt uttrycks, vilket gör att representationen av svensk forskningsverksamhet kan bli missvisande eller ofullständig hos återanvändare av metadata. Här kan nationell samordning vara till nytta för att fastställa arbetssätt och vilka uppgifter som bör tillföras eller hänvisas till i normalfallet, så att forskningsinformationen kan bli jämförbar och otvetydig.

¹⁵ Här kan exempelvis CERIF-standarderna vara aktuella. Se även kapitlet *Tekniska förutsättningar*.

De kärnmetadata som skickas in till PID-systemens infrastrukturer när en beständig identifierare skapas eller uppdateras, exempelvis via publiceringsprocesser, uppges också ofta i kvalitet och kan också sakna maskinläsbara relationer till andra poster och definitioner i den aktuella forskningskontexten. Då det ofta är dessa kärnmetadata som återanvänds av internationella infrastrukturer som arbetar med forskningsinformation gör detta att svenska forskningsprojekt i många fall saknar täckning i existerande informationsverktyg eller exkluderas från sådana som är under uppbyggnad. Ett tydligt exempel på detta är kunskapsgrafer för forskningsinformation, som baseras på sådana öppna metadata.

"De största svårigheterna är att forskare självständigt publicerar data utan några kvalitetskrav på tjänster eller den beskrivning och paketering som görs av datamaterialet"

"Om möjligt också kolla på metadatakvaliteten genom validering av output från olika providers. Det skulle kunna bidra till att rådgivningen till forskarna var de bör föredra att publicera sina forskningsdata"

Ett initiativ för att skapa en samsyn på grundläggande metadatakvalitet lyfts som en nationell angelägenhet, då en sådan skulle kunna hjälpa till att samordna svenska forskningsorganisationers arbete och säkerställa interoperabilitet med olika aktörer och standarder globalt. Ett sådant initiativ skulle sannolikt också indirekt peka ut olika infrastrukturer som bättre och sämre på grund av vad de rent tekniskt klarar av. Samtidigt kan det bli svårt att samordna annat än kvalitetskrav för sådana metadata som är allmängiltiga, på grund av olika disciplinspecifika krav.

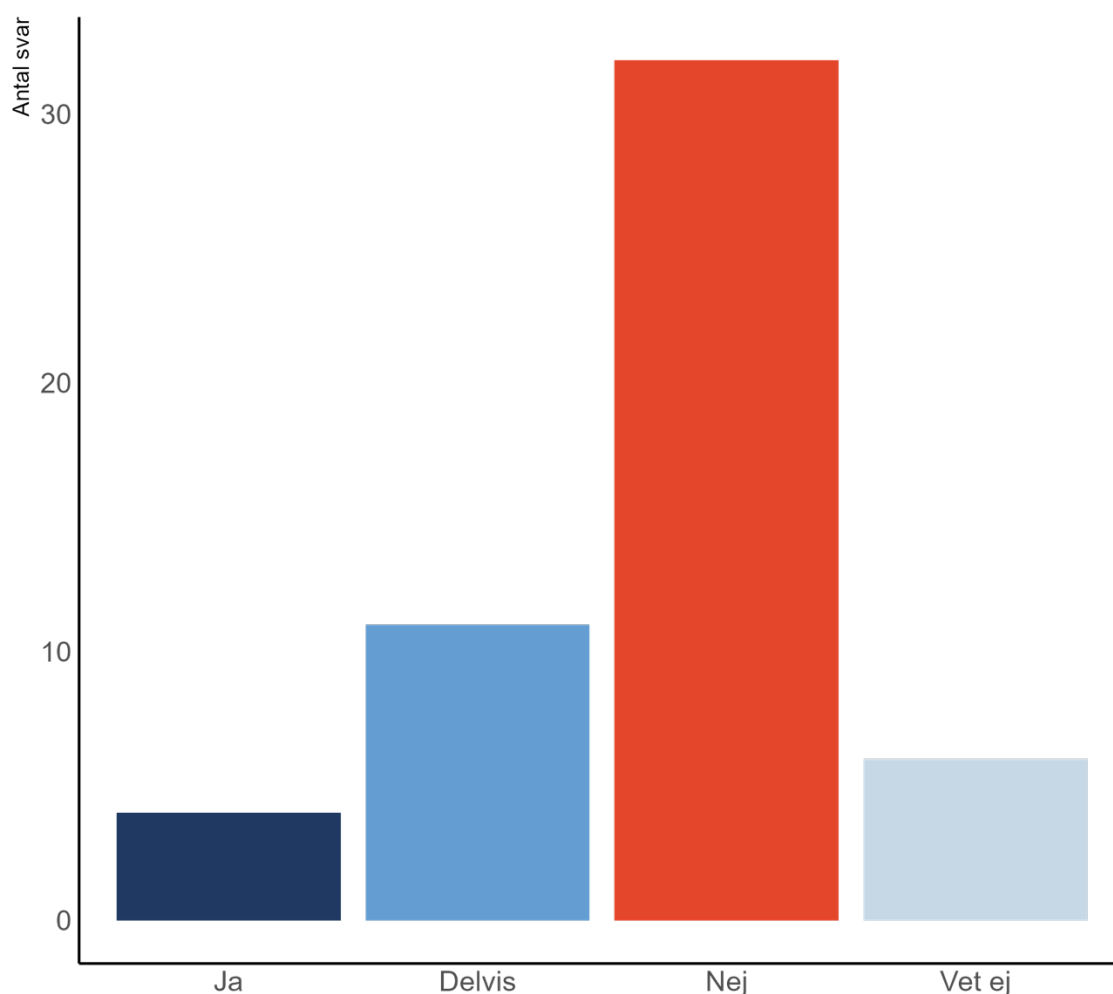


Fig 7. Gör ni någon registrering eller katalogisering av datapubliceringar som kommer till er kännedom, t.ex. i ett CRIS-system, institutionsrepositorium, publikationsdatabas eller annat liknande lokalt system? (Antal respondenter för den här frågan: 53).

En grundläggande förutsättning för att kunna förvalta information om organisationens datapubliceringar och dela den med andra är att datapubliceringarna på något sätt registreras eller katalogiseras hos organisationen. En majoritet av respondenterna svarar att detta inte görs aktivt, vilket bekräftar bilden av datapubliceringar som en udda och delvis oreglerad typ av forskningsprodukt. Den bristande täckningen för publicerade forskningsdata hos många indexerings tjänster, som i andra fall är användbara för att hitta olika forskningsprodukter, kan förklara svårigheterna att följa nya datapubliceringar systematiskt. Datapubliceringar tycks också ofta vara exkluderade från de olika typer av forskningsprodukter som forskare anmodas att aktivt rapportera in till forskningsorganisationerna när de har publicerats.

”Sverige har inget Cris-system med forskningsinformation som möter kraven för Euro-Cris och OpenAIRE och skickar inte heller sin forskningsinformation till OpenAIRE så att de blir sökbara i OpenAIRE Graph.”

”Tycker allmänt att det är synd att vi jobbar i stuprör kring olika delar av öppen vetenskap när vi istället kan dra nytta av redan befintlig infrastruktur för att samla all publicerad forskningsinformation. Risken är att vi blir beroende av kommersiella tjänster eller ökar antalet system och resurser som läggs på arbete med dem istället för stöd till forskare.”

Flera respondenter understryker vikten av att Sverige får en gemensam lösning för att samla in och göra information om offentligt finansierad forskning öppen och tillgänglig. Skapandet och förvaltandet av öppen forskningsinformation beskrivs som en nationell angelägenhet som långsiktigt bör samordnas i ett övergripande nationellt CRIS/RIMS-system eller liknande infrastruktur, som kan beskriva projekt och andra forskningsaktiviteter med dess relaterade forskningsprodukter som artiklar eller datapubliceringar¹⁶. Ett sådant system bör kunna ta emot öppen forskningsinformation från forskningsorganisationer, och leverera den vidare till andra, enligt internationella metadatastandarder¹⁷. OpenAIRE nämns som en viktig aktör som samlar in och sammanställer sådana uppgifter på internationell nivå. En nationell lösning förmodas minska bördan för varje enskild organisation, förutsatt att det finns tydliga regler och arbetsflöden så att varje forskningsorganisation vet vad den ska leverera. Ett möjligt flöde är att forskare rapporterar in forskningsdata som publicerats till sin organisation, på samma sätt som hos många organisationer görs för resultatpublikationer, och att denna information sedan kan återanvändas på nationell och internationell nivå.

Problemet handlar således inte bara om att kunna spåra datapubliceringar, utan om att göra all relevant forskningsinformation öppen och tillgänglig för återanvändning och sammanställning. Undantag från detta bör därför endast göras då unika förutsättningar gör det olämpligt att dela informationen. I ett digitalt ekosystem av forskningsinformation är datapubliceringar viktiga komponenter, som bäst förstås (och blir mest användbara) när de kompletteras med annan information, exempelvis om kopplade resultatpublikationer, projekt eller anslag.

På nationell nivå finns idag SwePub som samlar in information om publikationer och SweCris som beskriver projekt finansierade av vissa svenska forskningsfinansiärer. I SwePub ingår i dagsläget i praktiken inte datapubliceringar som publikationstyp, även om det finns tekniskt stöd för det i leveransformaten. Möjlighet finns dock att ange datapubliceringar som relaterade objekt till publikationer. SweCris tillhandahåller ett register över anslagen från några av de statliga finansiärerna med tillhörande projektbeskrivningar, men beskriver eller hänvisar inte till några relaterade forskningsprodukter. Publikationsrepositoriet DiVA, som används av flera svenska forskningsorganisationer för att förteckna och i vissa fall tillgängliggöra forskningsprodukter, har

¹⁶ Ett exempel på ett sådant system är *Forskning.fi* i Finland.

¹⁷ CERIF kan exempelvis läsas in av OpenAIRE, och DataCite Metadata Schema förstås av många metadataaggregatorer.

nyligen exkluderat datapubliceringar som möjlig publikationstyp och hänvisar sådana till Researchdata.se. DiVA är strikt publikationsorienterat och inte uppbyggt för att fungera som ett CRIS-system, även om det stöder hänvisningar till vissa beständiga identifierare. Diggs katalog på Dataportal.se är myndighetsgemensam och förhåller sig endast till enskilda datapubliceringar som inrapporterats vid sidan av andra myndighetsdata. Respondenterna uttrycker en frustration över att bristen på samordning på nationellt plan gör att potentiella dataflöden och sammanlänkning uteblir, vilket i sin tur kan leda till att metadata och annan forskningsinformation fastnar i olika kommersiella tjänster¹⁸.

"Researchdata.se som en ingång för svenska forskningsdata – kan det bli en portal för all forskningsdata som publiceras av forskare som är affilierade till svenska [organisationer]?"

Den nylanserade forskningsdataportalen Researchdata.se tas upp som en kanal som kan hjälpa till att fånga upp metadata om existerande datapubliceringar genom aktiv bevakning och skördning, eller som kan ge vägledning och verktyg för att skapa nya datapubliceringar. Flera nämner möjligheterna att hämta metadata från Researchdata.se till den egna forskningsorganisationen och lokal indexering. Det föreslagna informationsflödet relaterar också till processer för utbyte av forskningsinformation med det nationella CRIS-system som flera föreslår.

Flera respondenter lyfter behovet av gemensamma riktlinjer och mer enhetliga rekommendationer för forskare, gärna på nationell nivå. Det gäller generellt för forskningsdatahantering, där hjälpmedel som checklistor och andra stöddokument omnämns vid sidan av råd och verktyg för formatering av data eller skapande av metadata.

Riktlinjer efterfrågas dock särskilt för att skapa spårbara datapubliceringar, exempelvis lämpliga repositorer för olika typer av datamaterial och hur dessa bäst kan användas. Råd angående hur datapubliceringar framtidsäkras och görs redo för AI-användning och annan automatisering lyfts också som ett behov. Några respondenter nämner möjligheten att samordna metadatavalidering i form av riktlinjer och valideringsverktyg.

"Att ni jobbar för nationella incitament-strukturer"

"En nationell standard för yrkesmeritering för de som vågar satsa på att skapa bra data för underlag, på detta sätt kan fler också öppna upp data eller dela med sig av datan."

"Meritering på lokal och nationell nivå, t.ex. hos finansörerna. ("VR vill att man har en DMP men kollar aldrig på dem" är inte direkt ett incitament till att publicera forskningsdata.)"

¹⁸ Vissa proprietära lösningar för lokala CRIS-system eller publikationsrepositorier kan vara förknippade med metadataskördning till kommersiella databaser som har samma ägare.

En annan samordningsfråga som lyfts av flera respondenter är hur nationell samsyn kan skapas för hur publicering av forskningsdata av hög kvalitet kan relatera till incitamentsstrukturer och meritering. Detta kan röra frågor som hur datapubliceringar bör bedömas i forskningsutvärdering och yrkesmeritering, såsom vid tjänstetillsättningar eller tilldelning av anslag, eller vilka indikatorer som följs när forskningsaktiviteter och datapubliceringar kartläggs. Det kan vara svårt att som ensam forskningsorganisation eller finansiär skapa standarder för kvalitet och utvärdering om den inte blir jämförbar med den utvärdering som görs av andra.

"...komma ifrån den hårdaste tolkningen av GDPR i hela EU när det gäller anonymisering,..."

"...nationella juridiska riktlinjer för att minska belastning på lokala jurister"

Respondenterna beskriver en avskräckningseffekt kopplad till lagstiftningen i Sverige, särskilt vad gäller personuppgiftshantering i forskning och GDPR, som verkar hämmande i datapubliceringsfrågor. Identisk problematik tolkas olika inom olika forskningsorganisationer, där samma typ av situation kan hanteras restriktivt på ett ställe men erbjudas större handlingsfrihet på ett annat. Detta uppges leda till osäkerhet hos forskarna kring datahantering och datapubliceringar, särskilt i samarbeten över organisationsgränserna. Samtidigt är kunskapsnivån om förberedelser och metoder för att genomföra formella, spårbara publiceringar av skyddsvärda data med begränsad åtkomst generellt sett låg.

Samordning av tolkningar av och bedömningsunderlag för persondatahantering på en nationell nivå efterfrågas därför för att motverka osäkerhet och ett upplevt godtycke i lokala bedömningar, och för att tydliggöra metoder för hur datapubliceringar kan skapas med olika tillgänglighetsnivåer.

Några respondenter nämner också problem som skapas av olika tolkningar av terminologi relaterad till datapublicering och öppen vetenskap. Exempelvis kan vad som anses vara *forskningsdata* vara disciplinberoende och begrepp som *publicering* eller *tillgängliggörande* kan laddas med olika innebörd. Även här kan nationell samordning och samsyn vara till nytta, såsom i form av terminologiresurser som definitioner och kontrollerade vokabulärer.

"En teknisk lösning för metadatahöstning som fungerar för olika system och tjänster och som i sin tur kan leverera data till lokala system (ex CRIS)."

"Automatiserade standardiserade metoder för skördning av metadata från de mest vanligast förekommande repositorie-typerna."

"Vad finns det för möjlighet att utveckla tillgängliga verktyg för att hitta och identifiera svenska publicerade forskningsdata? Kan vi utveckla det som redan finns genom att bygga utifrån befintliga verktyg som t.ex. OpenAlex och DataCite?"

En nationell samordning och standardisering av den praktiska metodiken för att leta efter datapubliceringar som inte forskningsorganisationerna redan känner till efterfrågas också. Utsökningar på enskild forskningsorganisation bör vara möjliga. Som exempel på metod föreslås att

börja med att koncentrera sig på metadatakällor från stora infrastrukturer, och jämföra med redan insamlade metadata hos andra aggregatorer som OpenAIRE. Flera lyfter att disciplinspecifika publiceringsmönster gör att det blir svårt att skapa generiska metoder som fångar allt, och att riktade insatser med kompletterande metodik kan krävas. Avgränsningsproblematiken för vad som ska anses vara en datapublicering lyfts också som en försvårande aspekt.

”Underliggande plattformar för att arkivlagra stora mängder data vid svenska [organisationer].”

Som lösning på lagringsproblematiken, som är knuten till långsiktiga arkivkrav och behov av att genomföra öppna eller åtkomstbegränsade datapubliceringar, föreslås att svenska forskningsorganisationer skapar effektiva gemensamma lagringslösningar. Samtidigt uttrycks en medvetenhet om att det parallellt ofta kan behöva finnas lagringslösningar som helt kontrolleras av enskilda organisationer både vad gäller drift och administration, exempelvis för datamaterial av det känsligare slaget.

Ledning och resurser

För att publicera spårbara datapubliceringar behövs resurser, i form av kompetens, tid och lämplig teknisk infrastruktur. Enkätsvaren visar att respondenterna upplever resursbrist, både hos forskningsorganisationerna själva, och i den resursfördelning som görs inom individuella forskningsprojekt.

Få respondenter nämner ett uppdrag från ledningen att samla in och förvalta information om datapubliceringar, och en majoritet registrerar dem inte aktivt i dagsläget, trots kravet i öppna datalagen. Detta kan jämföras med en enkät som genomfördes av SND under våren 2024, där inga av de 27 deltagande svenska forskningsorganisationerna sade sig ha implementerat åtgärder för att samla in och rapportera information om data enligt lagkravet [91, s. 27].

Tidsbrist hos både forskare och stödpersonal pekas av respondenterna ut som en betydande utmaning. Dessutom lyfts brist på personal med rätt kompetens och otillräcklig budget för att hantera och tillgängliggöra forskningsdata fram som viktiga problemområden.

Resursbristen inom forskningsorganisationerna uppges delvis handla om kompetensbrist, exempelvis att det finns få medarbetare som kan ge stöd i frågor rörande datahantering, datapubliceringar, sammanlänkad forskningsinformation och öppen vetenskap. Den allmänna medvetenheten om metoder för datapublicering beskrivs som begränsad hos forskarna, exempelvis i fråga om att publicera skyddsvärda data med begränsad åtkomst, eller om att systematiskt arbeta med strukturerade metadata och beständiga identifierare i linje med FAIR-principerna.

"The repositories that exist are fine. People just don't know, don't care, don't have the time or budget to publish data."

Brist på tid att lägga på datapubliceringsprocessen uppges vara ett av de största hindren för forskare. När tid, budget och personalresurser ska fördelas inom ett forskningsprojekt prioriteras sällan datapubliceringar. En delförklaring till detta uppges vara en incitamentsstruktur där en väl genomförd datapublicering sällan ger fördelar eller blir uppmärksammas i forskningsutvärdering. Det beskrivs också att stödpersonalen inte får möjlighet att lägga tillräcklig arbetstid på uppgifterna, vilket begränsar deras möjligheter att aktivt uppmuntra, utbilda om och hjälpa till praktiskt med datapublicering.

"Vi behöver verktyg för detta för att minska kostnaden för att hitta var data publiceras. Det vore tidskrävande och kostsamt att använda metoder som att utgå från andra forskningspublikationer."

Själva problemställningen om hur forskningsorganisationer systematiskt ska kunna arbeta med att spåra nya och existerande datapubliceringar uppges också vara svår att ta sig an på grund av resursbrist, i synnerhet vad gäller att hitta och förteckna existerande datapubliceringar som faller under organisationens ansvar.

Uppföljning och incitament

Ett genomgående tema i enkätsvaren är incitamenten för att publicera forskningsdata och hur dessa påverkar såväl forskares praktiker som forskningsorganisationernas möjligheter att spåra materialet samt att följa upp hur väl policyer och riktlinjer efterlevs. Många uppger att datapubliceringen görs som ett direkt krav på att tillgängliggöra data på ett visst sätt, som kommer från en tidskrift eller en finansiär. Men när datapubliceringen väl genomförts enligt kravställarens grundkrav finns dock sällan några processer där den vetenskapliga gärningen att skapa datapubliceringen får något erkännande.

Datamaterialet uppges sällan bli föremål för granskning och kommentarer i referentgranskningsprocessen. Många forskare väljer därför redan på planeringsstadiet sådana infrastrukturer som möjliggör självpublicering av datapubliceringar utan någon kvalitetsgranskning. När publiceringen är slutförd uppges att ingen särskild uppföljning ges från exempelvis finansiärer eller andra aktörer, och forskarens organisation begär sällan in någon information om att datapubliceringen har skett.

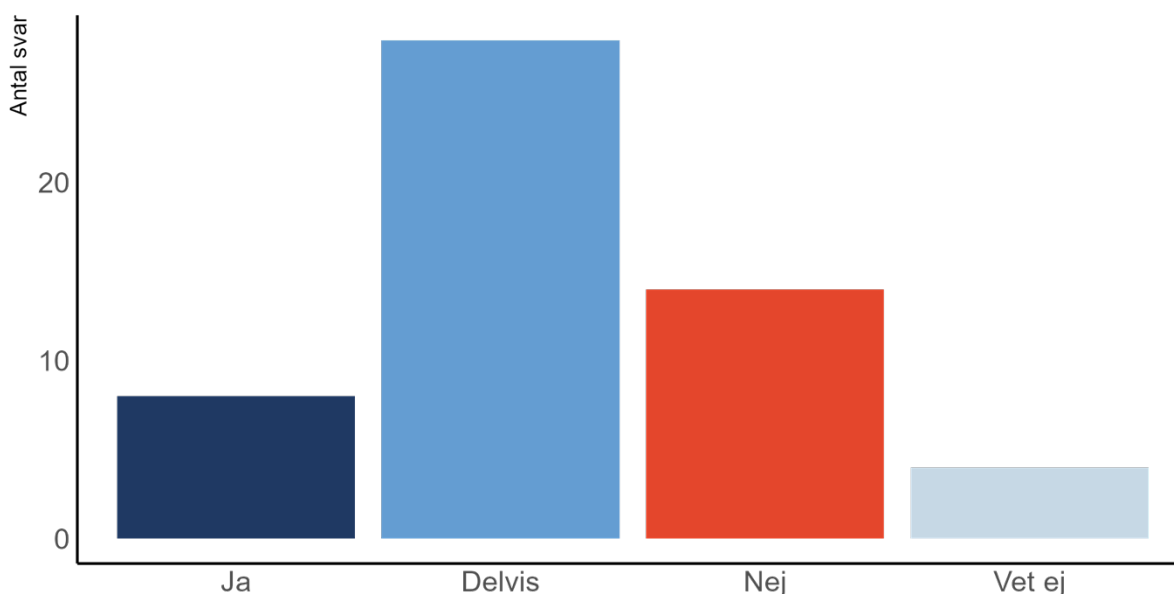


Fig 8. Upplever du att det finns incitament för att publicera forskningsdata (oavsett tillgänglighetsnivå)?

En majoritet av respondenterna upplever att tillräckliga incitament för att publicera data saknas eller att incitament för det bara delvis kan anses finnas. Den samlade bilden verkar vara att yttre kravställning är den drivande faktorn för de flesta datapubliceringar som skapas.

”[Forskare] avsätter sällan tid till att i övrigt få till en bra datapublicering. Detta för att arbetet som läggs ned trots alla fina ord om öppen vetenskap ofta inte tas hänsyn till när forskningen skall utvärderas.”

Datapubliceringar får på grund av citeringspraktiker och brister i infrastrukturer sällan något genomslag som kan fångas upp av mätmetoder som bibliometri. Till följd av detta anges att datapubliceringar sällan granskas och bedöms som ett underlag för forskningens kvalitet som kan vara meriterande för forskarens fortsatta karriär, exempelvis vad gäller nya forskningsanslag, tjänstetillsättningar eller befordran.

Att endast genomföra datapubliceringarna för att agera i linje med generiska målsättningar om öppen vetenskap upplever många som ett för diffust incitament, som inte i någon större grad motiverar forskarna att lägga ned tid och arbete på datapubliceringarnas kvalitet, inklusive aspekter som spårbarhet och sammanbundenhet. Sammantaget uppges detta medföra att ansträngningarna sällan går utöver kravställarens miniminivå för datapubliceringar, exempelvis en tidskrifts anvisningar till författare.

”Att öppen vetenskaplig praktik premieras i utvärderingar och tjänstetillsättningar”

” Det borde vara meriterande att publicera dataset likaväl som artiklar. Både på lokal nivå på [organisationen] och på nationell nivå.”

Lösningarna som förs fram för att åtgärda incitamentsproblematiken, och således även möjligheten att följa upp datapubliceringar, är spridda. Övergripande ses ett behov av att datapubliceringar behöver få ett meritvärde på samma sätt som resultatpublikationer, men hur man når dit är inte samstämmigt. Vissa resonemang hos respondenterna betonar kvalitativ bedömning där datapubliceringen räknas vid tjänstetillsättningar, befordran och nya forskningsanslag och där meriteringssystemet till exempel breddas enligt de kriterier som lyfts av CoARA. Andra resonemang om uppföljning och incitament efterfrågar kvantitativa indikatorer, mått som ger en fingervisning om kvalitet för datapubliceringar. Även om samsyn saknas kring exakt vad som ska mätas är önskan om transparens tydlig, indikatorerna behöver vara erkända och jämförbara mellan lärosäten. I beskrivningarna av båda förhållningssätten framhävs också av respondenterna att kvalitet är svårt att mäta och att kvantitativa och kvalitativa bedömningar måste backas upp av sakkunnig expertis samt ett ramverk för ansvarsfull forskningsbedömning.

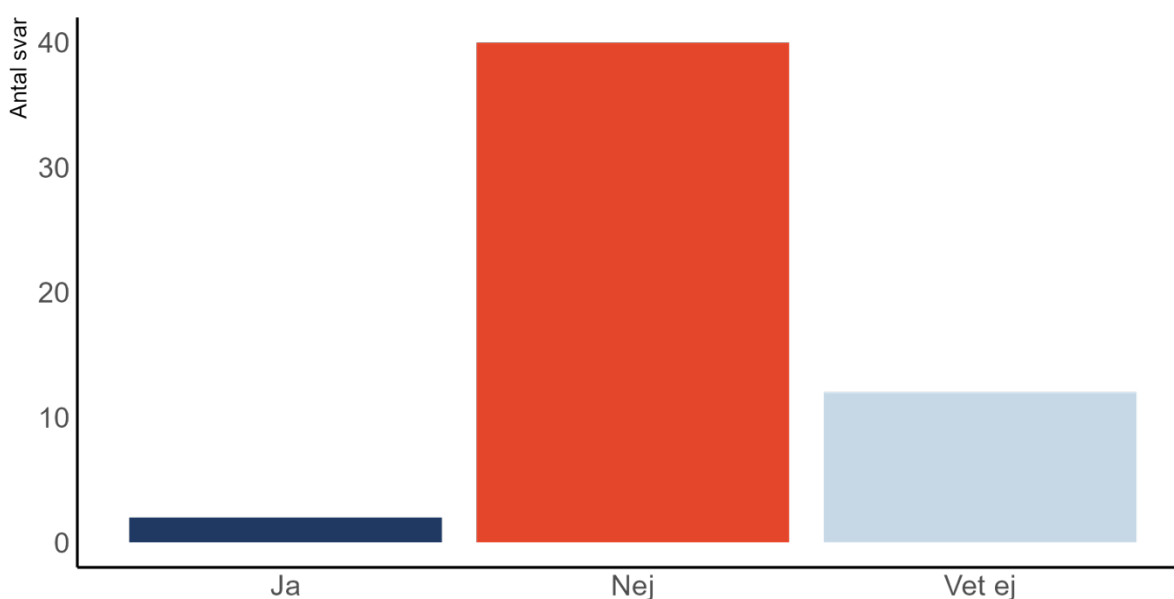


Fig 9. Har din organisation tagit fram någon form av indikator/KPI för forskningsdata?

Få respondenter svarar dock att de har arbetat fram specifika indikatorer för datapubliceringar som kan användas i det löpande arbetet för att spåra, följa och bedöma publiceringarna. Några förklaringar till detta kan vara ett undermåligt informationsunderlag på grund av spårbarhetsproblematiken, den låga graden av aktiv inrapportering av datapubliceringar till organisationerna, och den osäkerhet kring vilka datapubliceringar som kan följas rent metodmässigt som många uttrycker.

"Att publicera forskningsdata är ett sätt att synliggöra forskningen och att visa att man följer god forskningssed. Det underlättar också delning av data med andra forskare."

Trots svårigheten att konkretisera incitamenten återkommer tanken att datapublicering i sig är ett varumärkesbyggande moment som signalerar god forskningssed, underlättar delning och stärker synligheten för både forskare och resultat, vilket i sin tur förväntas leda till fler citeringar och samarbeten. Denna potentiella avkastning anses underutnyttjad och beroende av fungerande infrastrukturer och utvecklade praktiker för att uppnå ytterligare synlighet och genomslag.

"Att informationen [om existerande datapubliceringar] delas i Sverige är en grundläggande förutsättning för att kunna analysera med hjälp av t.ex. bibliometri för forskningsutvärderingar"

Uppföljningen beskrivs alltså som ett delat ansvar. Organisationen förväntas skapa processer och infrastrukturer som säkrar spårbarhet och sammanbundenhet, kopplar datapubliceringen till projekt och publikationer och gör uppföljning möjlig. Vissa menar att dessa strukturer behöver förstärkas med tydligare krav och sanktioner eftersom utebliven eller dåligt genomförd datapublicering bör få konkreta konsekvenser. Kärnfrågan är att datapubliceringen måste vävas in i det akademiska meriteringssystemet och att uppföljningen kräver en modell med både kvalitativa och kvantitativa redskap.

Vägen mot en öppen vetenskapskultur

Vi ser tydliga tecken på att respondenterna upplever att forskningsdata borde publiceras, men att olika kulturella faktorer ofta bidrar till att så inte sker. Datapublicering är inte alltid en självklar del av den vetenskapliga kulturen inom olika forskarkluster. Respondenterna uppfattar det som att data publiceras när det finns tydliga externa krav på det, men att publiceringen då ofta görs på ett sätt som uppfyller minimikraven men inget mer.

"Även om t.ex. stödmaterial kring datahanteringsplaner berättar om hur man från början ska planera för att göra det material som sedan ska tillgängliggöras FAIR så omsätts detta mycket sällan i praktiken i ordentliga beskrivningar och metadata, utan datapubliceringen är något som görs i panik inför en deadline från en tidskrift eller liknande"

Systematiska förberedelser för datapubliceringen genomförs sällan som en integrerad del av forskningsprocessen. När data publiceras är det vanligare att sammanställningen av datapubliceringen sker med ont om tid nära en deadline i projektets slutfas, exempelvis i samband med processen att publicera en artikel. Även detta uppges ibland kunna leda till att data tillgängliggörs på sätt som gör att de inte blir fristående och spårbara forskningsprodukter, exempelvis för att de saknar tydliga kärnmetadata eller hamnar i supplementfiler som bara kan hittas i en tidskrifts webbgränssnitt. Just supplementfilerna lyfts av flera respondenter som en särskilt problematisk del av kulturen som upprätthålls av förlagen och som inverkar menligt på forskningsdatas spårbarhet.

Respondenterna uppger också att systemet med redogörelser för dataåtkomst (*data availability statements*) knutna till tidskriftsartiklar ofta används för att ange olika ursäkter till varför en formell datapublicering inte kan genomföras, även om det faktiskt hade varit fullt möjligt att publicera spårbara data, antingen öppet eller med begränsad åtkomst. En vanlig praktik uppges också vara att skriva ett mer allmänt hållet yttrande om att data kan göras tillgängliga på begäran, utan att för den skull formellt förbereda, beskriva och sammanställa data för återanvändning. Det är svårt att utröna i vilken omfattning data aktivt delas på dessa sätt i en svensk kontext, men praktiken får tydliga negativa följder då de data som använts och hänvisas till inte blir till spårbara forskningsprodukter som är öppet beskrivna enligt FAIR-principerna. Vad som också är oklart är i vilken utsträckning sådana praktiker uppmuntrar till informell "skuggdelning" av data och hur ofta delningen i så fall genomförs i enlighet med svensk lagstiftning.

"Olika kulturer kring hur man citerar data kan ställa till det, eftersom många kanske hellre citerar ett metodkapitel eller liknande även när data finns tillgängliggjorda som självständig datapublicering. Det är en kulturell fråga som behöver förändras i samverkan med forskare"

En aspekt som lyfts är att skapandet av en kultur där citering av data blir norm skulle bidra till att öka antalet spårbara datapubliceringar. I nuläget kan det framstå som krångligt att citera data, eller som att det inte är den accepterade formen för hänvisningar till materialet inom en viss forskningsdisciplin. Systemet med redogörelser för dataåtkomst (*data availability statements*) motarbetar också självcitering av datapubliceringar i relevanta delar av resultatpublikationer genom att särbehandla datamaterialet i formalia. Även om en citerbar datapublicering faktiskt finns tillgänglig som primärkälla, uppges att andra relaterade publikationer kan bli citerade som substitut för att citera forskningsdata. Att forskningsdata ofta exkluderas ur citatindexerande tjänster antas också kunna påverka forskares benägenhet att aktivt citera primär- eller sekundärdata i form av egna eller andras datapublikationer.

Bristen på uppföljning och incitament för att skapa välgjorda datapubliceringar nämns som en del av problemen som befäster dessa praktiker. Utveckling av uppföljningsprocesser och ett aktivt arbete för att skapa konkreta incitament knutna till meritering nämns som åtgärder som skulle kunna bidra till att publicering av data blir något som forskare förväntas göra som en del av forskningsarbetet. Inom vissa fält upplever respondenterna att det redan finns relativt starka normer för detta, medan det inom andra fält är mer ovanligt med datapublicering.

"Många av de ställningstaganden som görs inom t.ex. CoARA och DORA kan ligga till grund för en bättre incitamentsstruktur för att tillgängliggöra forskningsdata"

"inom forskningsområdet (bibliometri) finns det en stark strävan mot öppen forskningsinformation (Barcelondeklarationen)"

Flera respondenter lyfter de nya ambitionerna för samarbete kring öppen forskningsinformation, implementation av öppna vetenskapspraktiker och en breddad syn på forskningsbedömning som

formuleras av initiativ som Barcelonadeklarationen för öppen forskningsinformation och CoARA. Det finns ett stort stöd för reformer både från delar av forskarkåren och hos olika stödfunktioner.

Många av dessa förändringar kräver dock konkreta insatser av beslutsfattare för att kunna bli verklighet. Så länge forskningsorganisationer fortsätter att arbeta på oförändrade sätt vad gäller rapportering och beskrivning av forskningsprodukter, kvalitetssäkring av metadata, eller hänsynstagande till datapubliceringar i meritutvärdering och forskningsbedömning, upplevs det svårt för enskilda individer att åstadkomma förändring. [OBJ]

Slutsatser

Arbetet för att kunna skapa och följa upp spårbara datapubliceringar kan ses som en del av en bredare rörelse för öppen vetenskap som började ta form under 1990-talet. Inledningsvis fokuserade rörelsen på arbetet med att skapa möjligheter för öppna resultatpublikationer. Det är dock först nu som vi börjar se resultaten av en faktisk kulturförändring där öppen tillgång till dessa börjar bli standard.

På samma sätt kan rörelsen som förespråkar öppna forskningsdata, och att skapa återanvändbara och spårbara datapubliceringar som är beskrivna i linje med FAIR-principerna, ses arbeta mot ett långsiktigt mål. Detta mål är att en kulturförändring har inträffat som faktiskt påverkat beteenden hos forskare och forskningsorganisationer genom nya kollegiala förväntningar på forskningsdata. Det breda arbetet för öppna forskningsdata är dock betydligt yngre, även om det nu har funnits uttryckliga politiska målsättningar för det på EU-nivå och svensk forskningspolitisk nivå i snart ett decennium.

Att genomföra en kulturförändring med krav som kommer uppifrån är svårt, och resulterar ofta i förvirring och fragmenterade resultat när insikterna om hur och varför något bör implementeras är begränsade. Nyttan är heller inte alltid tydlig, och flera krav kan uppfattas som att de går emot egna intressen eller tär på existerande resurser. Det rör sig heller kanske inte om en, utan flera kulturförändringar som parallellt försöker införas i olika kulturella kluster.

Vi kan exempelvis se en sådan fragmenterad implementation hos forskarkulturella kluster, där man skapar en datapublicering om en finansiär eller ett förlag begär det, men inte alltid med självklarhet citerar datapubliceringarna. Den går också att se i olika organisationskulturer. Hos forskningsorganisationerna har man skapat stödverksamheter för hantering av forskningsdata, men i stort inte brytt sig om att följa data när de har publicerats. Hos förlagen har man uppfattat att det finns en förväntning på öppna data och pliktskyldigt skapat särskilda regler för forskningsdata, men inte reflekterat över att särbehandling av datapubliceringar kan ställa till problem i andra led. Hos finansiärerna kräver man gärna datahanteringsplaner och öppna data, men tar sällan hänsyn till dessa i granskning, uppföljning och forskningsbedömning. I samarbetskulturerna i forskningssektorn på nationell nivå har man gjort många satsningar på öppna forskningsdata, men inte kommit överens om att datapubliceringar ska kunna följas på samma sätt som andra typer av forskningsprodukter och hur det i så fall ska göras.

Sett till de ambitiösa kraven på införandet av ett öppet vetenskapssystem redan 2026 [9, s. 42] verkar de förändringar som gjorts vara för spridda och otillräckliga för att nå målet så snart. Hur skapar man då en långsiktig kulturförändring i Sverige där alla eller nästan alla datapubliceringar går att följa? Lösningen är kanske att inte bara försöka forcera fram kulturförändringarna med krav som kommer uppifrån.

En strategi för kulturförändring

Organisationen Center for Open Science har formulerat en strategi för hur öppna vetenskapspraktiker kan införlivas i existerande kollegiala forskarkulturer [92]. De menar att även om många skriver under på grundvärderingarna om transparens och öppenhet är det svårt att lösgöra sig från existerande praktiker, traditioner och belöningssystem som på olika sätt kan motarbeta förändringarna. För att möjliggöra en kulturförändring behöver man i stället bygga nedifrån, genom att göra insatser som skapar en bättre grund att stå på och förebygger friktion och osäkerhet.



Fig. 10. Interventionspyramiden för kulturförändring¹⁹ som tagits fram av Center for Open Science [92].

Den ordning av interventioner som föreslås av Center for Open Science i strategin illustreras av en pyramid, där basen är fungerande infrastruktur och processer, och nästa steg i pyramiden är användarvänlighet hos tjänster och arbetsflöden. Med detta som grund kan kollegialt samförstånd börja skapas, exempelvis genom goda exempel och föregångspersoner som sår frön till nya normer. För att befästa de nya praktikerna behöver också konkreta incitament skapas där sådant som erkännande och meritering ger faktisk återkoppling på förändrade kulturella beteenden. Först när alla dessa förutsättningar finns på plats går det att ställa krav som är förankrade i kulturen och har goda utsikter att efterlevas.

Även om strategin från början är tänkt att appliceras på forskares disciplinspecifika kulturkluster är den också användbar för att resonera kring kulturförändringar i forskningsorganisationer, infrastrukturer eller på nationell samordningsnivå, och vilka insatser som kan främja och underlätta förändringarna.

¹⁹ Grafik baserad på ett bildkoncept från Center for Open Sciences webbplats, CC-BY 4.0 [93]

Nedan görs nedslag i hur olika uppgifter från kartläggningen kan kopplas till konkreta behov och insatser som på olika sätt relaterar till strategin för kulturförändring. Det är inte en heltäckande redogörelse, men uppgifterna kan tjäna som exempel på utvecklingsmöjligheter och nya förhållningssätt som långsiktigt kan bidra till en bättre spårbarhet för datapubliceringar.

Gör det möjligt

Flera av de uppgifter som framkommer i kartläggningen belyser infrastrukturella brister som kan behöva insatser för att möjliggöra vissa arbetssätt.

- På forskarnivån är det inte säkert att de disciplinspecifika behov man har för att tillgängliggöra data motsvaras av en lämplig tjänst eller infrastruktur som också ger goda möjligheter för att göra data spårbara, citerbara och FAIR. Detta kan kräva nyutveckling av tjänster. I de fall där tjänster finns kan dessa behöva utveckla funktionalitet eller följa standarder på bättre sätt, som att skapa och använda beständiga identifierare med tydliga kärnmetadata.
- På forskningsorganisationsnivån ser vi att flera inte har lokala system som kan registrera och följa datapubliceringar som rapporteras av forskarna, och att man ofta inte heller har ordnat lagring som möjliggör publicering av åtkomstbegränsade data eller långsiktig arkivering av datapubliceringar. Detta kan kräva utveckling eller anskaffning av infrastruktur.
- Hos förlagen kan man ha instiftat formaliaregler eller byggt system som orsakar hinder för att spåra eller citera forskningsdata. Att komma runt detta kan kräva dialog med förlag och tidskrifter eller val av andra publiceringskanaler.
- På nationell nivå finns inte ett sammanhållet registreringsflöde för alla datapubliceringar och inte heller en samordning och aggregering av all öppen forskningsinformation i ett nationellt CRIS-system eller forskningsportal, även om lanseringen av Researchdata.se i viss mån utökat registreringsmöjligheterna för datapubliceringar. En sådan samordning och ett sådant system kan därför behöva utvecklas.

Gör det enkelt

Kartläggningen visar också på olika insatser som kan göras för att vägleda och göra processer mer lätthanterliga.

- På forskarnivån behövs beslutsstöd och vägledning för många av de ställningstaganden som behöver göras vid datapublicering: dels hur man avgör vad som är möjligt och lämpligt juridiskt och ansvarsmässigt, dels hur man förbereder data praktiskt och arbetar med de olika verktyg, system och standarder som relaterar till publiceringsprocessen, inte minst vad gäller FAIR, metadata och beständiga identifierare, för att uppnå en grundläggande kvalitetsnivå. Det finns också behov av tydlig och lättförståelig vägledning för hur man praktiskt bör arbeta i specifika tjänster när datapubliceringar skapas för att göra dem spårbara, då arbetsflöden och funktionalitet kan skilja sig åt.

- På forskningsorganisationsnivån finns ett behov av att skapa enkla flöden där metadata om nya datapubliceringar kan registreras av forskare på samma sätt som resultatpublikationer, eller på andra sätt inhämtas till exempelvis ett CRIS-system. Det finns också ett behov av att använda projekthanteringsverktyg, verktyg för datahanteringsplaner och liknande på ett sätt som gör att grundläggande, kvalitetssäkrade metadata kan återanvändas och inte behöver återskapas vid varje givet tillfälle av forskaren, exempelvis vid publiceringstillfället. Kopplat till detta finns också ett behov av insatser där beständiga identifierare konsekvent implementeras i forskningsprocesser och administration, och där funktionaliteten hos PID-systemen också används för att lagra öppen forskningsinformation som kan återanvändas. Processer för metadatavalidering kan införas för att göra det enkelt att uppfylla grundkvalitet.
- På nationell nivå finns behov av att hitta samförstånd för och standardisera helhetslösningar för hur öppen forskningsinformation, inklusive information om datapubliceringar, samlas in och delas i Sverige. Kopplat till detta finns ett behov av att definiera vad som ska anses vara en adresserbar datapublicering som bör spåras. Vidare finns tydliga behov av att skapa en nationell standard och gemensamma arbetssätt för hur man kvalitetssäkrar arbetet med grundläggande metadataproduktion, beständiga identifierare och deras tillhörande metadatastandarder. Detta kan inkludera sådant som minimikrav, harmonisering av terminologi, standardisering av formella relationsuttryck och valideringsprocesser. Det finns också behov av ytterligare samordning av olika ställningstaganden som relaterar till datapublicering, i synnerhet en samsyn på juridiska bedömningar av vanligt förekommande situationer som kan ge trygga underlag för beslutsstöd.

Gör det till förebild

Kartläggningen visar också att kollegiala förväntningar spelar stor roll för hur forskare och forskningsorganisationer praktiskt arbetar med datapubliceringar.

- Att forskare förändrar vissa beteenden kring forskningsdata och datapublicering för att säkerställa kvalitet och spårbarhet är ett långsiktigt mål, men också det som är svårast att göra direkta insatser för, då forskare själva fritt väljer metoder och publiceringssätt. Goda förebilder och föregångspersoner kan hjälpa till att skapa nya kollegiala förväntningar. Nära knutet till detta är behovet av att väl genomförda datapubliceringar får återkoppling i utvärderingar, vilket också kan agera katalysator för skapandet av förebilder.
- Forskningsorganisationer kan behöva göra insatser för att lyfta fram väl genomförda datapubliceringar i forskningskommunikation och använda olika sätt för att nå ut, exempelvis genom att ta hjälp av *data champions* som kan driva utvecklingen av datahantering och spårbar datapublicering inom organisationen, eller genom att göra riktade utbildningsinsatser. Organisationen kan också behöva lägga extra vikt vid att främja goda praktiker för datacitering och att själva följa upp citering och återanvändning av datapubliceringar på sätt som tydligt kan uppfattas.

- På nationell nivå behövs ett ökat fokus på att visa upp datapublicering i excellent forskning och hur värdefulla forskningsdata görs öppet tillgängliga. Datapubliceringar bör följas som forskningsprodukter i nationella översikter. Uppföljning bör inte bara göras med indikatorer på makronivå utan även genom att granska kvalitetsaspekter och lyfta goda exempel.

Gör det meriterande

Flera respondenter i kartläggningen betonar att meriteringsfrågorna är några av de viktigaste incitamenten för att skapa spårbara datapubliceringar av hög kvalitet.

- Forskningsorganisationer bör göra insatser för att väl genomförda datapubliceringar kan tas i beaktande vid utvärdering av forskares prestationer, såsom vid tjänstetillsättningar och befordran. Bedömningar av datapubliceringar bör ske i linje med de nya kriterier för kvalitativ utvärdering av forskningsverksamhet som lyfts inom exempelvis CoARA. Organisationerna bör också följa samtliga datapubliceringar som skapas av affilierade forskare och utveckla metoder för bibliometrisk uppföljning som även drar nytta av nya tekniker för att följa öppen forskningsinformation.
- Forskningsfinansiärer bör i större utsträckning ta hänsyn till datapubliceringar i arbete med granskning och forskningsutvärdering. Att forskare kan visa upp ett ansvarsfullt och kvalitetsmedvetet arbete med datapubliceringar, FAIR-aspekter och andra öppna vetenskapspraktiker bör vara av betydelse vid tilldelning av forskningsanslag.
- På nationell nivå finns ett behov av att skapa en samsyn för hur skapandet av högkvalitativa datapubliceringar kan bli en del av det akademiska meriteringssystemet och tas hänsyn till i forskningsutvärdering och yrkesmeritering. Öppna vetenskapspraktiker bör ges tydliga erkännanden för att vi långsiktigt ska kunna nå de nationella målen kring öppen vetenskap. Öppen forskningsinformation bör samordnas och samlas in nationellt så att detaljerade forskningskontexter kan delas med omvärlden i form av metadata, och på så sätt sammanlänkas med andra processer som följer upp forskningsverksamhet internationellt.

Gör det obligatoriskt

Strategin för kulturförändring betonar att krav kan förväntas bli uppfyllda på meningsfulla sätt först när de andra beståndsdelarna finns på plats, och att kraven ofta annars resulterar i splittrade åtgärder utan förankring eller sammanhang.

- Kartläggningen visar på att publicering av data i huvudsak sker på forskares egna initiativ eller på grund av specifika krav från forskningsfinansiärer eller tidskrifter.
- Flera forskningsfinansiärer har uttryckliga krav på datapublicering, exempelvis Horizon Europe.
- Forskningsorganisationer kan ha interna regler och riktlinjer för datapublicering, även om vi inte granskat dem individuellt i denna kartläggning.

- På nationell nivå regleras publicering av forskningsdata av att det i forskningspropositionen *Forskning och innovation för framtid, nyfikenhet och nytta* [9, s. 42] fastslås att forskningsdata ska publiceras i enlighet med principer för öppen vetenskap. Till stöd för detta finns riktlinjer och vägledande information från Vetenskapsrådet, Kungliga biblioteket och Digg.
- Konkreta lagkrav på att följa och förteckna datapubliceringar som tillgängliggörs av en myndighet finns i öppna datalagen.

I realiteten innebär detta att många forskare inte upplever något obligatorium att skapa datapubliceringar. Detta observerades också av flera respondenter, men svaren fokuserade mindre på kravställning och mer på behoven av att utveckla incitament till att skapa spårbara datapubliceringar, och att utveckla tekniska och organisatoriska möjligheter för att göra det på effektiva och ändamålsenliga sätt.

Det framstår dock som att det finns en brist på policyer som konkretiserar förväntningarna på forskarna och forskningsorganisationerna, till exempel hur data ska publiceras och hur FAIR-principerna bör implementeras. De krav som finns är ofta löst formulerade vilket gör att för många aktörer är svårt att förstå hur de ska uppfyllas i praktiken. Detta leder även till att stödfunktioners uppdrag ofta är oklara, vilket försvårar deras arbete.

Några förslag på konstruktiva kravställningar som kan användas för att skapa en högre mognadsgrad i arbetet med att skapa spårbara datapubliceringar kan vara:

- Minimikrav för att ange affilieringsinformation.
- Krav på inrapportering av alla datapubliceringar som skapas till forskningsorganisationen.
- Krav på att skapa, underhålla och konsekvent använda ORCID.
- Minimikrav på angivna metadata i olika utpekade infrastrukturer för datapublicering.
- Minimikrav på information som framgår i citeringar och hänvisningar till datapubliceringar.
- Minimikrav på information som anges i redogörelser för dataåtkomst.

Avslutning

Aktivt arbete för att främja öppen datapublicering i Sverige har pågått i över ett decennium, men har huvudsakligen resulterat i olika generiska förhållningssätt för hur man kan arbeta med datapublicering i forskningsprocessen. Det betyder inte nödvändigtvis att data i praktiken publiceras på spårbara sätt.

Eftersom skapandet av högkvalitativa, spårbara datapubliceringar sällan ger resultat i form av erkännande i forskningsutvärdering och meritering har det faktiska genomförandet ofta blivit splittrat, då tid och resurser hellre läggs på andra uppgifter.

För att skapa en hållbar kultur kring datapublicering krävs därför en aktiv inrapportering vid publicering, bättre metoder för att spåra datapubliceringar som redan existerar och att datapubliceringar tillmäts ett större värde som forskningsprodukter i uppföljning och utvärdering.

Definitioner

Affilieringsinformation

Affilieringsinformation är detaljer om forskares organisatoriska tillhörigheter som lämnats ut vid publiceringstillfället, exempelvis i uppgifter om vem som har skapat en datapublicering. Denna information kan användas för att spåra och koppla ihop en datapublicering med en forskningsorganisation. För konsekvent spårning krävs att tydliga affilieringsuppgifter förekommer i metadata, helst med beständiga identifierare som ORCID och ROR ID.

Beständiga identifierare (PID)

Beständiga identifierare, eng. *persistent identifiers* (PID), är unika och varaktiga identifierare som används för att referera till fenomen eller resurser som dataset (t.ex. DOI), forskare (t.ex. ORCID), resultatpublikationer (t.ex. DOI), och forskningsorganisationer (t.ex. ROR ID). PID-system ger möjlighet att underhålla metadata kopplad till identifieraren, och det går därför att göra ändringar som att byta ut länken till en fil eller ändra på en namnform. På så sätt säkerställer beständiga identifierare att fenomen och resurser kan hittas och hänvisas till på ett korrekt sätt över tid.

CRIS-system

Current Research Information Systems (CRIS) kan användas av lärosäten för att hantera forskningsinformation, såsom metadata som beskriver forskningsprojekt, resultatpublikationer och datapubliceringar. Ett CRIS-system kan fylla en mer heltäckande forskningsadministrativ funktion, eller vara mer avgränsat och exempelvis endast fokusera på synliggörande av forskningen för omvärlden. Förkortningen RIMS (Research Information Management System) förekommer också, med i stort sett samma innebörd. CRIS- och RIMS-system kan kopplas till andra system för att dela metadata och förbättra synligheten hos organisationens forskningsaktiviteter.

Datakataloger

Datakataloger fungerar som register över datapubliceringar och innehåller metadata som hjälper användare att hitta och förstå dem. De kan länka till olika repositorer eller andra infrastrukturer där data faktiskt finns. Ett exempel på en datakatalog är Google Dataset Search.

Datapublicering

Publicering av forskningsdata kan ske som öppet tillgängliga dataset eller som dataset med begränsad tillgång där metadata hänvisar till skyddade filer som förvaltas på en behörighetskontrollerad lagringsyta. Datasetet fastslås i en specifik version med aktuella filer och datapubliceringen blir en fristående upprättad handling. Både öppna och åtkomstbegränsade dataset räknas som datapubliceringar, om de formellt har tillkännagivits via en ändamålsanpassad

infrastruktur och kan adresseras och citeras självständigt. Att exempelvis endast ange att forskningsdata är tillgängliga på förfrågan i en notis i en vetenskaplig artikel räknas dock inte som en datapublicering.

Begreppet datapublicering som sådant är inte alltid enkelt tillämpligt på allt tillgängliggörande av forskningsdata. Vissa discipliner och behov leder till att data tillgängliggörs på sätt som inte enkelt kan avgränsas på detta vis, exempelvis genom strömmande dataflöden eller genom inrapportering av enstaka datapunkter i större datamängder utan bevarad dokumentation att hänvisa till.

Datarepositorier

Datarepositorier är digitala lagringssystem där forskningsdata kan lagras, hanteras och göras tillgängliga. I dessa kan datapubliceringar skapas. Datarepositorierna kan vara generella eller disciplinspecifika. Generella datapubliceringar är öppna för datapubliceringar från alla forskningsområden och saknar ämnesbegränsningar. Specifika datapubliceringar är anpassade för datapubliceringar inom ett visst forskningsområde och kan följa särskilda metadatastandarder eller ha disciplinspecifika krav på den information som anges om datamaterialet.

FAIR

FAIR-principerna (Findable, Accessible, Interoperable, Reusable) syftar till att göra forskningsdata mer sökbara/hittbara, tillgängliga, kompatibla och återanvändbara. Dessa principer har till stor del att göra med hur forskningsdata är beskrivna via öppet tillgängliga metadata och dokumentation, och berör även hur data förbereds, struktureras och lagras. FAIR-principerna är viktiga i arbetet för att förbättra datahantering och nå målsättningar kring öppen vetenskap. FAIR är dock inte synonymt med öppen tillgång, och forskningsdata kan göras FAIR även om datafiler inte är tillgängliga för allmänheten. Exempelvis kan en viktig FAIR-aspekt då vara en tydlig beskrivning av hur data kan efterfrågas och vilka villkor som finns för en utlämning.

Institutionsrepositorier och publikationsdatabaser

Ett institutionsrepositorium är ett digitalt lagringssystem som förvaltas av exempelvis ett lärosäte för att tillgängliggöra eller hänvisa till material som relaterar till verksamheten. Detta är vanligtvis forskningsprodukter som artiklar eller avhandlingar. Förutom att lagra, beskriva och tillgängliggöra digitala objekt har institutionsrepositorier oftast möjlighet att skapa metadata som hänvisar till material som är tillgängliggjort någon annanstans, i form av hänvisningsposter. Forskningsdata kan vara lagrade i institutionsrepositorier, men det är vanligare att data tillgängliggörs i dedikerade datapubliceringar. Om systemet uteslutande hänvisar till material tillgängliggjort på annan plats med hänvisningsposter kan det kallas publikationsindex eller publikationsdatabas. Ibland är institutionsrepositorier och publikationsdatabaser separata system inom en organisation. Systemen är ofta integrerade med, eller en del av, CRIS-system och bibliotekstjänster.

Kunskapsgrafer för forskningsinformation

Det allmänna begreppet kunskapsgrafer syftar på samma tekniker som också har kallats öppna länkade data eller den semantiska webben, och som beskriver hur man skapar hållbara decentraliserade arbetssätt för metadata på internet.

Det specifika konceptet kunskapsgrafer för forskningsinformation syftar till möjligheten att kunna koppla samman fenomen som forskningsprojekt, forskare, anslag, resultatpublikationer, forskningsdata, forskningsinfrastrukturer, osv. med relationer som bildar ett grafnätverk.

Förutsättningen är att dessa har länkats samman i otvetydiga metadata med hög tillförlitlighet, helst via maskinläsbara beständiga identifierare. Ytterligare metadata för de sammankopplade objekten eller fenomenen kan vara förvaldade på andra ställen där de i sin tur kan slås upp och återanvändas. Konceptet gör det möjligt att navigera i kunskapsgraferna och skapa rika översikter av forskningsverksamhet utan att sammanställa all information för hand, exempelvis genom automatisering eller AI-bearbetning. Internationellt kallas begreppet bland annat *Science Knowledge Graphs* (SKGs) eller *PID Graphs*.

Redogörelser för dataåtkomst (data availability statements)

Redogörelser för dataåtkomst eller *data availability statements* är uppgifter om tillgängligheten hos den forskningsdata som ligger till grund för resultat i en resultatpublikation, som anges enligt särskilda formaliaregler från förlag och tidskrifter.

Dessa redogörelser är ofta hänvisade till en särskild spalt, sektion eller metadatafält i förlagens digitala publiceringsinfrastrukturer som är skild från publikationens textinnehåll. Redogörelserna kan innehålla referenser och beständiga identifierare, kontaktuppgifter eller generiska formuleringar om tillgänglighet eller hinder för tillgänglighet.

Text mining

Text mining är ett begrepp som syftar på olika automatiserade tekniker för att utvinna specifik information ur textmassor. Olika metoder finns för att känna igen olika mönster och formuleringar och på så vis extrahera uppgifter som man letar efter.

Text mining av forskningsprodukter brukar oftast handla om att man letar efter information i resultatpublikationers textinnehåll (fulltextsökning). Detta kan göras av olika anledningar, som i forskningssyfte, eller för att hitta ytterligare information som inte finns i publikationernas metadata och annars skulle kräva manuell genomläsning för att hitta.

Öppen forskningsinformation

Öppen forskningsinformation är ett begrepp som huvudsakligen syftar på beskrivande metadata som redogör för hur forskningsverksamhet hänger samman. Dessa metadata är information som lagras

och förvaltas utanför innehållet i resultatpublikationer, datapublikationer och andra forskningsprodukter.

Uppgifterna kan till exempel beskriva vad som det forskas om, vem som gör det, vilka forskningsprodukter detta resulterat i, vilka forskningsorganisationer som är ansvariga, vem som finansierar forskningen, vilka verktyg och infrastrukturer som används, och så vidare. Uppgifterna utgörs delvis av bibliografisk information som beskriver resultatpublikationer och andra forskningsprodukter, men moderna tekniker och standarder ger betydligt större möjligheter att beskriva de olika delarna av en forskningsaktivitet, inklusive nyttiggörande, återanvändning och citeringar.

Öppen forskningsinformation som koncept hör nära samman med initiativ för att skapa och dela sådan information utan att den läses in och behöver köpas tillbaka från olika kommersiella aktörer. Tillvägagångssättet förespråkas exempelvis av *Barcelonadeklarationen för öppen forskningsinformation*.

Begreppet bör inte förväxlas med den generiska användning av "forskingsinformation" som tidigare varit vanlig i myndighetstexter och syftat på resultatpublikationer. Denna användning har dock numera ofta ersatts av andra formuleringar som "kommunikation av forskningsresultat".

Öppen tillgång och åtkomstbegränsad tillgång till datapubliceringar

Öppet tillgängliga data är data som fritt kan laddas ned av allmänheten. Data med begränsad tillgång är formellt tillkännagivna och beskrivna, men har specifika åtkomstbegränsningar.

Både öppet tillgängliga data och data med åtkomstbegränsad tillgång kan göras till fristående och citerbara datapubliceringar som tilldelas unika beständiga identifierare. Även om forskningssektorn strävar efter att tillgängliggöra öppna data kan det vara nödvändigt att genomföra datapubliceringar med begränsad tillgång för att uppfylla lagkrav eller andra restriktioner, då möjligheten till utlämning av datafiler kan behöva bedömas från fall till fall i en sekretessprövning.

Oavsett om data publiceras med öppen eller begränsad tillgång kan öppna metadata och öppen dokumentation oftast tillgängliggöras för allmänheten. Filer som ska ingå i datapubliceringen kan också i båda fallen förberedas aktivt med relevanta åtgärder för framtida återanvändning. På så sätt kan datapubliceringar i de flesta fall genomföras i linje med FAIR-principerna oavsett eventuella åtkomstbegränsningar. Detta sammanfattas i maximen "så öppet som möjligt, så begränsat som nödvändigt".

På samma sätt kan data ha gjorts öppet tillgängliga utan att uppfylla FAIR-principerna, exempelvis genom att beskrivningar och metadata saknas eller är bristfälliga, eller att filer bara finns i format som är svåra eller dyra att återanvända.

Källförteckning

- [1] M. D. Wilkinson *m.fl.*, "The FAIR Guiding Principles for scientific data management and stewardship", *Sci Data*, vol. 3, nr 1, s. 160018, dec. 2016, doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- [2] Finansdepartementet, *Lag (2022:818) om den offentliga sektorns tillgängliggörande av data*, vol. 2022:818. 2022. Åtkomstdatum: 10 februari 2025. [Online]. Tillgänglig vid: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-2022818-om-den-offentliga-sektorns_sfs-2022-818/
- [3] Myndigheten för digital förvaltning (Digg), "Lämna metadata till Sveriges dataportal". Åtkomstdatum: 17 februari 2025. [Online]. Tillgänglig vid: <https://www.digg.se/kunskap-och-stod/oppna-och-delade-data/offentliga-aktorer/vagledning-om-oppna-datalagen/lamna-metadata-till-sveriges-dataportal>
- [4] Myndigheten för digital förvaltning (Digg), "Sveriges dataportal". Åtkomstdatum: 17 februari 2025. [Online]. Tillgänglig vid: <https://www.dataportal.se/>
- [5] Myndigheten för digital förvaltning (Digg), "Värdefulla datamängder". Åtkomstdatum: 02 maj 2025. [Online]. Tillgänglig vid: <https://www.digg.se/kunskap-och-stod/oppna-och-delade-data/offentliga-aktorer/vardefulla-datamangder>
- [6] Vetenskapsrådet, *Vägledning för implementering av kriterier för FAIR forskningsdata*. Stockholm: Vetenskapsrådet, 2021. Åtkomstdatum: 17 februari 2025. [Online]. Tillgänglig vid: <https://www.vr.se/analys/rapporter/vara-rapporter/2021-10-28-vagledning-for-implementering-av-kriterier-for-fair-forskningsdata.html>
- [7] Europeiska unionens råd, "Rådets slutsatser om högkvalitativ, transparent, öppen, tillförlitlig och rättvis akademisk publicering", Europeiska unionens råd, 9616/23, maj 2023. [Online]. Tillgänglig vid: <https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/sv/pdf>
- [8] Europeiska unionens råd, "Rådets slutsatser om övergången till ett öppet vetenskapssystem", Europeiska unionens råd, 9526/16, maj 2016. [Online]. Tillgänglig vid: <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/sv/pdf>
- [9] Klimat- och näringslivsdepartementet och Utbildningsdepartementet, *Forskning och innovation för framtid, nyfikenhet och nytta (Prop. 2024/25:60)*. 2024. [Online]. Tillgänglig vid: <https://www.regeringen.se/rattsliga-dokument/proposition/2024/12/prop.-20242560>
- [10] CoARA, "Coalition for Advancing Research Assessment". Åtkomstdatum: 29 april 2025. [Online]. Tillgänglig vid: <https://coara.eu/>
- [11] Declaration on Research Assessment (DORA), "San Francisco-deklarationen om forskningsutvärdering", Declaration on Research Assessment (DORA). Åtkomstdatum: 02 maj 2025. [Online]. Tillgänglig vid: <https://sfedora.org/read/read-the-declaration-swedish/>
- [12] Barcelona Declaration, "Commitments", Barcelona Declaration on Open Research Information. Åtkomstdatum: 28 april 2025. [Online]. Tillgänglig vid: <https://barcelona-declaration.org/commitments/>
- [13] H. Van de Sompel *m.fl.*, "FAIR Signposting Profile", Signposting the Scholarly Web. Åtkomstdatum: 08 september 2025. [Online]. Tillgänglig vid: <https://signposting.org/FAIR/>
- [14] K. Payne och C. Verhey, "Schema.org for research data managers: a primer", *International Journal of Big Data Management*, vol. 2, nr 2, s. 95, 2022, doi: [10.1504/IJBDM.2022.128449](https://doi.org/10.1504/IJBDM.2022.128449)
- [15] D. Auclair, J. Carmichael, och J. Thibodeau, "Challenges and roadblocks to robust metadata in the scholarly communications industry", *Insights: the UKSG journal*, vol. 37, nr 1, maj 2024, doi: [10.1629/uksg.642](https://doi.org/10.1629/uksg.642).
- [16] L. Lannom, B. P. Boesch, och S. Sun, "Handle System Overview", Internet Engineering Task Force, Request for Comments RFC 3650, nov. 2003. doi: [10.17487/RFC3650](https://doi.org/10.17487/RFC3650).

- [17] Kungliga biblioteket, "URN:NBN", Kungliga biblioteket – Sveriges nationalbibliotek – kb.se. Åtkomstdatum: 05 mars 2025. [Online]. Tillgänglig vid: <https://www.kb.se/isbn-och-utgivning/urnnbn.html>
- [18] DataCite Metadata Working Group, *DataCite Metadata Schema for the publication and citation of research data and other research outputs v4.6*, 05 december 2024. doi: [10.14454/MZV1-5B55](https://doi.org/10.14454/MZV1-5B55).
- [19] P. Feeney, "Required, recommended, and optional metadata", Crossref. Åtkomstdatum: 25 maj 2025. [Online]. Tillgänglig vid: <https://www.crossref.org/documentation/schema-library/required-recommended-elements/>
- [20] A. Dappert, A. Farquhar, R. Kotarski, och K. Hewlett, "Connecting the persistent identifier ecosystem: Building the technical and human infrastructure for open research", *Data Science Journal*, vol. 16, juni 2017, doi: [10.5334/dsj-2017-028](https://doi.org/10.5334/dsj-2017-028).
- [21] PID Forum Finland, "The National Roadmap for Persistent Identifiers for Finland". april 2024. Åtkomstdatum: 27 maj 2025. [Online]. Tillgänglig vid: <https://urn.fi/URN:NBN:fi-fe2024032512910>
- [22] C. Ferguson *m.fl.*, "D3.1 Survey of Current PID Services Landscape - Revised", Zenodo, okt. 2019. doi: [10.5281/zenodo.3554255](https://doi.org/10.5281/zenodo.3554255).
- [23] S. Soiland-Reyes, C. Goble, och P. Groth, "Evaluating FAIR Digital Object and Linked Data as distributed object systems", *PeerJ Computer Science*, vol. 10, s. e1781, apr. 2024, doi: [10.7717/peerj-cs.1781](https://doi.org/10.7717/peerj-cs.1781).
- [24] S.-C. Andrea, R.-G. Nicolas, L. Van Thed, och C. Rodrigo, "Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: a methodological discussion", *Scientometrics*, vol. 126, nr 8, s. 7149–7165, aug. 2021, doi: [10.1007/s11192-021-04043-5](https://doi.org/10.1007/s11192-021-04043-5).
- [25] DataCite Metadata Working Group, "12. RelatedIdentifier", DataCite Metadata Schema 4.6. Åtkomstdatum: 04 september 2025. [Online]. Tillgänglig vid: <https://datacite-metadata-schema.readthedocs.io/en/4.6/properties/relatedidentifier/>
- [26] P. Manghi, "Challenges in building scholarly knowledge graphs for research assessment in open science", *Quantitative Science Studies*, vol. 5, nr 4, s. 991–1021, nov. 2024, doi: [10.1162/qss_a_00322](https://doi.org/10.1162/qss_a_00322).
- [27] euroCRIS, "Common European Research Information Format (CERIF)", euroCRIS: The International Organisation for Research Information. Åtkomstdatum: 27 augusti 2025. [Online]. Tillgänglig vid: <https://eurocris.org/services/main-features-cerif>
- [28] T. Suominen, C. Tatum, S. Sipponen, W. Rydman, och S. Ross, "Piloting the use of RAiD in Research.fi", *Procedia Computer Science*, vol. 249, s. 224–231, jan. 2024, doi: [10.1016/j.procs.2024.11.068](https://doi.org/10.1016/j.procs.2024.11.068).
- [29] European Open Science Cloud (EOSC), "Research Activity Identifier Service (RAiD)", FAIRCORE4EOSC. Åtkomstdatum: 03 september 2025. [Online]. Tillgänglig vid: <https://faircore4eosc.eu/eosc-core-components/research-activity-identifier-service-raid>
- [30] J. Dederke, B. Hirschmann, och D. Johann, "Der Data Citation Index von Clarivate: Eine wertvolle Ressource für die Forschung und für Bibliotheken?", *B.I.T. online*, vol. 25, nr 1, s. 21–27, 2022, doi: [10.3929/ETHZ-B-000545462](https://doi.org/10.3929/ETHZ-B-000545462).
- [31] Elsevier, "Sunset of Data Monitor", Pure: Pure Help Center for Pure Administrators. Åtkomstdatum: 07 september 2025. [Online]. Tillgänglig vid: <https://helpcenter.pure.elsevier.com/data-sources-and-integrations/sunset-data-monitor>
- [32] Z. Cao, L. Zhang, Y. Huang, och R. Haunschild, "How does the academia refer to open research information data sources? A review study based on OpenAlex and Microsoft Academic series", *Scientometrics*, juni 2025, doi: [10.1007/s11192-025-05347-6](https://doi.org/10.1007/s11192-025-05347-6).
- [33] J. Priem, H. Piwowar, och R. Orr, "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts", 17 juni 2022, *arXiv*: arXiv:2205.01833. doi: [10.48550/arXiv.2205.01833](https://doi.org/10.48550/arXiv.2205.01833).

- [34] A. Mannocci och M. Baglioni, "Exploring scientometrics with the OpenAIRE Graph: introducing the OpenAIRE Beginner's Kit", 19 september 2024, *arXiv*: arXiv:2409.12690. doi: [10.48550/arXiv.2409.12690](https://doi.org/10.48550/arXiv.2409.12690).
- [35] A. Burton *m.fl.*, "The Scholix Framework for interoperability in data-literature information exchange", *D-Lib Magazine*, vol. 23, nr 1/2, jan. 2017, doi: [10.1045/january2017-burton](https://doi.org/10.1045/january2017-burton).
- [36] S. Peroni och D. Shotton, "OpenCitations, an infrastructure organization for open scholarship", *Quantitative Science Studies*, vol. 1, nr 1, s. 428–444, feb. 2020, doi: [10.1162/qss.a.00023](https://doi.org/10.1162/qss.a.00023).
- [37] D. Brickley, M. Burgess, och N. Noy, "Google Dataset Search: Building a search engine for datasets in an open Web ecosystem", i *The World Wide Web Conference*, i WWW '19. New York, NY, USA: Association for Computing Machinery, maj 2019, s. 1365–1375. doi: [10.1145/3308558.3313685](https://doi.org/10.1145/3308558.3313685).
- [38] Undervisnings- och kulturministeriet och CSC – IT-centret för vetenskap, "Forskning.fi". Åtkomstdatum: 08 september 2025. [Online]. Tillgänglig vid: <https://forskning.fi/>
- [39] Sikt – Kunnskapssektorens tjenesteleverandør, "Nasjonalt vitenarkiv (NVA)". Åtkomstdatum: 08 september 2025. [Online]. Tillgänglig vid: <https://nva.sikt.no/>
- [40] Kungliga biblioteket, "Swepub", Kungliga biblioteket – Sveriges nationalbibliotek – kb.se. Åtkomstdatum: 03 mars 2025. [Online]. Tillgänglig vid: <https://kb.se/samverkan-och-utveckling/swepub.html>
- [41] Svensk nationell datatjänst, "Researchdata.se – Din portal till forskningsdata". Åtkomstdatum: 08 september 2025. [Online]. Tillgänglig vid: <https://researchdata.se/>
- [42] K. Vrouwenvelder, N. H. Raia, och A. K. Thomer, "Obstacles to dataset citation using bibliographic management software", *Data Science Journal*, vol. 24, s. 17, maj 2025, doi: [10.5334/dsj-2025-017](https://doi.org/10.5334/dsj-2025-017).
- [43] S. Fricke, E. Enslow, och S. Shipman, "Access to Supplemental Journal Article Materials", *The Serials Librarian*, vol. 80, nr 1–4, s. 85–96, maj 2021, doi: [10.1080/0361526X.2021.1883596](https://doi.org/10.1080/0361526X.2021.1883596).
- [44] G. Macgregor, B. S. Lanchobarrantes, och D. R. Pennington, "Measuring the concept of PID Literacy: User perceptions and understanding of PIDs in support of open scholarly infrastructure", *Open Information Science*, vol. 7, nr 1, mar. 2023, doi: [10.1515/opis-2022-0142](https://doi.org/10.1515/opis-2022-0142).
- [45] L. Candela, D. Mangione, och G. Pavone, "The FAIR Assessment Conundrum: Reflections on Tools and Metrics", *Data Science Journal*, vol. 23, nr 1, maj 2024, doi: [10.5334/dsj-2024-033](https://doi.org/10.5334/dsj-2024-033).
- [46] Vetenskapsrådet, *Öppen tillgång till forskningsdata 2025: en kartläggning, analys och bedömning*. Stockholm: Vetenskapsrådet, 2025. [Online]. Tillgänglig vid: <https://www.vr.se/analys/rapporter/vara-rapporter/2025-08-20-oppen-tillgang-till-forskningsdata--en-kartlaggning-analys-och-bedomning-2025.html>
- [47] Svensk nationell datatjänst, "Beskrivning av DAU (Version: 6)", 25 februari 2025, *Zenodo*. doi: [10.5281/zenodo.14945941](https://doi.org/10.5281/zenodo.14945941).
- [48] Svensk nationell datatjänst, "SND gör alla typer av digitala forskningsdata tillgängliga", Svensk nationell datatjänst. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://snd.se/sv/om-oss>
- [49] NAISS, Nationell akademisk infrastruktur för superdatorer i Sverige, "Om oss", NAISS. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://www.naiss.se/om-oss/>
- [50] SciLifeLab, "Enabling life science", SciLifeLab. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://www.scilifelab.se/about-us/>
- [51] SciLifeLab, "data.scilifelab.se", SciLifeLab Data Platform - hub for data-driven life science research in Sweden. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://data.scilifelab.se/>
- [52] Huminfra, "Huminfra - En nationell infrastruktur för forskning inom humaniora", Huminfra. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://www.huminfra.se/about>

- [53] InfraVis, "About InfraVis", InfraVis. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://infravis.se/about-infravis/>
- [54] Vetenskapsrådet, "Öppen tillgång till forskningsdata", Vetenskapsrådet. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://www.vr.se/uppdrag/oppen-vetenskap/oppen-tillgang-till-forskningsdata.html>
- [55] Kungliga biblioteket, "Öppen tillgång", Kungliga biblioteket. Åtkomstdatum: 19 oktober 2025. [Online]. Tillgänglig vid: <https://www.kb.se/for-bibliotekssektorn/oppen-vetenskap/oppen-tillgang.html>
- [56] Svensk nationell datatjänst, "Katalogen", Researchdata.se. Åtkomstdatum: 10 oktober 2025. [Online]. Tillgänglig vid: <https://researchdata.se/sv/om/katalogen>
- [57] D. Lin *m.fl.*, "The TRUST Principles for digital repositories", *Scientific Data*, vol. 7, nr 1, s. 144, maj 2020, doi: [10.1038/s41597-020-0486-7](https://doi.org/10.1038/s41597-020-0486-7).
- [58] CoreTrustSeal Standards and Certification Board, "CoreTrustSeal Requirements 2023-2025", sep. 2022. doi: [10.5281/zenodo.7051012](https://doi.org/10.5281/zenodo.7051012).
- [59] DIN - Deutsches Institut für Normung, *DIN 31644:2012-04 Information and documentation - Criteria for trustworthy digital archives*, april 2012. doi: [10.31030/1854419](https://doi.org/10.31030/1854419).
- [60] ISO - International Organization for Standardization, "ISO 16363:2025 Space data and information transfer systems — Audit and certification of trustworthy digital repositories", ISO. Åtkomstdatum: 17 september 2025. [Online]. Tillgänglig vid: <https://www.iso.org/standard/87472.html>
- [61] ISO - International Organization for Standardization, "ISO 14721:2025 Space Data System Practices — Reference model for an open archival information system (OAIS)", ISO. Åtkomstdatum: 17 september 2025. [Online]. Tillgänglig vid: <https://www.iso.org/standard/87471.html>
- [62] W. H. Walters, "Data journals: incentivizing data access and documentation within the scholarly communication system", *Insights: the UKSG journal*, vol. 33, s. 18, jan. 2020, doi: [10.1629/uksg.510](https://doi.org/10.1629/uksg.510).
- [63] Vetenskapsrådet, Enheten för data som strategisk resurs, "RUT - Register Utiliser Tool", Registerforskning.se. Åtkomstdatum: 18 september 2025. [Online]. Tillgänglig vid: <https://rut.registerforskning.se/>
- [64] Europeiska kommissionen, "Förordningen om det europeiska hälsodataområdet", Europeiska kommissionen. Åtkomstdatum: 18 september 2025. [Online]. Tillgänglig vid: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_sv
- [65] A. Rauber *m.fl.*, "Precisely and Persistently Identifying and Citing Arbitrary Subsets of Dynamic Data", *Harvard Data Science Review*, vol. 3, nr 4, okt. 2021, doi: [10.1162/99608f92.be565013](https://doi.org/10.1162/99608f92.be565013).
- [66] A. Rauber, M. Parsons, och RDA Data Citation WG, "Precise, Actionable Reference to Dynamic Data: Recommendations of the Working Group on Data Citation (WGDC)", juli 2025, doi: [10.15497/RDA00132](https://doi.org/10.15497/RDA00132).
- [67] C. Bornatici *m.fl.*, "CESSDA Recommendations on Data Citation: Practical Recommendations for Key Stakeholders", *CESSDA ERIC*, mar. 2025, doi: [10.5281/zenodo.15043854](https://doi.org/10.5281/zenodo.15043854).
- [68] M. Martone, "Data Citation Synthesis Group: Joint Declaration of Data Citation Principles", FORCE11, San Diego, CA, 2014. doi: [10.25490/a97f-egyk](https://doi.org/10.25490/a97f-egyk).
- [69] K. Gregory, A. Ninkov, C. Ripp, E. Roblin, I. Peters, och S. Haustein, "Tracing data: A survey investigating disciplinary differences in data citation", *Quantitative Science Studies*, vol. 4, nr 3, s. 622–649, dec. 2023, doi: [10.1162/qss_a_00264](https://doi.org/10.1162/qss_a_00264).
- [70] Svensk nationell datatjänst, "Kartläggning av förlags riktlinjer för delning av forskningsdata", Zenodo, apr. 2025. doi: [10.5281/zenodo.15044399](https://doi.org/10.5281/zenodo.15044399).

- [71] Utbildningsdepartementet, *Lag (2019:504) om ansvar för god forskningssed och prövning av oredlighet i forskning*. [Online]. Tillgänglig vid: <https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-2019504-om-ansvar-for-god-forskningssed-och-sfs-2019-504/>
- [72] Justitiedepartementet L6, *Tryckfrihetsförordning (1949:105)*. [Online]. Tillgänglig vid: <https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/tryckfrihetsforordning-1949105-sfs-1949-105/>
- [73] Kulturdepartementet, *Arkivlag (1990:782)*. [Online]. Tillgänglig vid: <https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/arkivlag-1990782-sfs-1990-782/>
- [74] Riksarkivet, *RA-FS 1999:1 Riksarkivets föreskrifter och allmänna råd om gallring av handlingar i statliga myndigheters forskningsverksamhet*. [Online]. Tillgänglig vid: <https://foreskrifter.riksarkivet.se/api/rafs/pdf/rafs%2FRA-FS%201999-01.pdf>
- [75] SUHF:s expertgrupp för arkiv och informationshantering (SUHF REK 2022:1), "Rekommendation om tillämpning av regelverk angående gallring och bevarande av forskningsinformation", Sveriges universitets- & högskoleförbund, apr. 2022. Åtkomstdatum: 28 augusti 2025. [Online]. Tillgänglig vid: <https://suhf.se/app/uploads/2022/05/REK-2022-1-Rekommendation-om-tillampning-av-regelverk-angaende-gallring-och-bevarande-av-forskningsinformation.docx>
- [76] K. Hoshi Larsson, "En sammanfattning av hur man hittar forskningsdata och var LiU-forskarens forskningsdata finns", Linköping University, Linköping, dec. 2023. doi: [10.3384/report-199505](https://doi.org/10.3384/report-199505).
- [77] L. Andersson, A. Gavel, P. Sjögårde, H. Skyllberg, och K. Widin, "Kartläggning KI:s datapublicering: förstudie och rekommendationer", Karolinska Institutet, Stockholm, okt. 2024. doi: [10.5281/zenodo.13935095](https://doi.org/10.5281/zenodo.13935095).
- [78] M. Almbro och L. Olsson, "Searching for research data – an assessment of data publication practises at Stockholm University", 20 februari 2024. doi: [10.5281/zenodo.10724633](https://doi.org/10.5281/zenodo.10724633).
- [79] J. Philipson, "Prepare to Preserve: the Harvest Combine for Research Data at Stockholm University and the Systems Inventory", i *iPRES 2024 Papers - International Conference on Digital Preservation*, Ghent & Flanders (Belgium): PubPub, sep. 2024. doi: [10.21428/5676bf2d.cb2d42ce](https://doi.org/10.21428/5676bf2d.cb2d42ce).
- [80] L. R. Johnston *m.fl.*, "Seek and you may (not) find: A multi-institutional analysis of where research data are shared", *PLoS ONE*, vol. 19, nr 4, s. e0302426, apr. 2024, doi: [10.1371/journal.pone.0302426](https://doi.org/10.1371/journal.pone.0302426).
- [81] N. Van Wettere, "Affiliation information in DataCite dataset metadata: a Flemish case study", *Data Science Journal*, vol. 20, s. 13, mar. 2021, doi: [10.5334/dsj-2021-013](https://doi.org/10.5334/dsj-2021-013).
- [82] T. Suominen, L. Himanen, L. Hellsten, M. Bennett, och M. Buys, "Piloting the use of PIDGraph data in Research.fi", *Procedia Computer Science*, vol. 249, s. 232–242, jan. 2024, doi: [10.1016/j.procs.2024.11.069](https://doi.org/10.1016/j.procs.2024.11.069).
- [83] A. Bassinet, L. Bracco, A. L'Hôte, E. Jeangirard, P. Lopez, och L. Romary, "Large-scale machine-learning analysis of scientific PDF for monitoring the production and the openness of research data and software in France", *HAL Open Science*, juni 2023, [Online]. Tillgänglig vid: <https://hal.science/hal-04121339v3>
- [84] M. Gabelica, R. Bojčić, och L. Puljak, "Many researchers were not compliant with their published data sharing statement: a mixed-methods study", *Journal of Clinical Epidemiology*, vol. 150, s. 33–41, okt. 2022, doi: [10.1016/j.jclinepi.2022.05.019](https://doi.org/10.1016/j.jclinepi.2022.05.019).
- [85] N. Khan, M. Thelwall, och K. Kousha, "Data sharing and reuse practices: disciplinary differences and improvements needed", *Online Information Review*, vol. 47, nr 6, s. 1036–1064, 2023, doi: [10.1108/OIR-08-2021-0423](https://doi.org/10.1108/OIR-08-2021-0423).

- [86] I. Gerasimov, B. Kc, A. Mehrabian, J. Acker, och M. P. McGuire, "Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Crossref, and DataCite", *Scientometrics*, vol. 129, nr 7, s. 3681–3704, juli 2024, doi: [10.1007/s11192-024-05073-5](https://doi.org/10.1007/s11192-024-05073-5).
- [87] P. Dorta-González, S. M. González-Betancor, och M. I. Dorta-González, "To what extent is researchers' data-sharing motivated by formal mechanisms of recognition and credit?", *Scientometrics*, vol. 126, nr 3, s. 2209–2225, mar. 2021, doi: [10.1007/s11192-021-03869-3](https://doi.org/10.1007/s11192-021-03869-3).
- [88] M. Hovde, "Enkätverktyg [Survey]", Sunet. Åtkomstdatum: 08 oktober 2025. [Online]. Tillgänglig vid: <https://www.sunet.se/services/samarbete/enkatverktyg>
- [89] U. Andersson *m.fl.*, "Data från en enkät om spårbarhet hos datapubliceringar som skapats vid svenska forskningsorganisationer". Chalmers tekniska högskola, 2025. doi: [10.71870/5f35-mc67](https://doi.org/10.71870/5f35-mc67).
- [90] R. Rampin, V. Steeves, och S. DeMott, *Taguette: open-source qualitative data analysis*. (04 februari 2023). Zenodo. doi: [10.5281/zenodo.7606027](https://doi.org/10.5281/zenodo.7606027).
- [91] Vetenskapsrådet, *Öppen tillgång till forskningsdata: en kartläggning, analys och bedömning 2024*. Stockholm: Vetenskapsrådet, 2024. [Elektronisk resurs]. Tillgänglig vid: <https://www.vr.se/analys/rapporter/vara-rapporter/2024-06-26-oppen-tillgang-till-forskningsdata--en-kartlaggning-analys-och-bedomning-2024.html>
- [92] B. Nosek, "Strategy for Culture Change", Center for Open Science. Åtkomstdatum: 14 oktober 2025. [Online]. Tillgänglig vid: <https://www.cos.io/blog/strategy-for-culture-change>
- [93] Center for Open Science, "About", Center for Open Science. Åtkomstdatum: 15 oktober 2025. [Online]. Tillgänglig vid: <https://www.cos.io/about>