

Datastatus vid leveranserna till Swepub

Sammanställning av enkätsvaren från de organisationer som levererar data till Swepub

Tuija Drake, Enheten för metadata och systemstöd, Avdelningen för informationssystem, Kungliga biblioteket.

Inledning

I oktober 2018 skickade Kungliga biblioteket en [enkät](#) till de organisationer som levererar data till Swepub. Syftet med enkäten var att kartlägga organisationernas egna uppfattningar om datastatus i förhållande till arbetet med kvalitetssäkringen av publikationsdata i de lokala publikationsdatabaserna samt dataleveranserna till Swepub, den nationella publikationsdatabasen av forskningspublicering i Sverige.

Av de 45 organisationer som levererar data till Swepub svarade 33 (ca 75%). Inga av frågorna var obligatoriska att svara på och därmed finns det skillnader på antalet svar per fråga. Utöver förbestämda svarsalternativ fanns det även möjlighet att ange andra alternativ eller kommentarer.

Enkätsvaren kommer att användas inom vidareutveckling av Swepub för förbättrade validering och berikning av datakvalitet samt visning av datastatus per levererande organisation.

Kvalitetssäkring

Årtalet då publikationsposterna har mest uttömmande granskats i de lokala publikationsdatabaserna är viktig information för uppvisning av datastatus. Nästintill alla organisationer som levererar data till Swepub har kvalitetssäkrat sina poster senast från och med 2012. Årtalet 2012 har pekats som det år då data borde vara kvalitetssäkrade för att kunna användas vid bibliometriska analyser.

Publikationstyper

Den allra vanligaste publikationstypen (avhandlingar exkluderade) i de lokala publikationsdatabaserna är tidskriftsartiklar. Hälften anger konferensbidrag som vanligt förekommande, därefter kommer bokkapitel. En tredjedel anger även rapporter som vanligt förekommande i de lokala publikationsdatabaserna.

Vad för publikationstyper är mest förekommande i den lokala publikationsdatabasen (avhandlingar exkluderade)?

33 svar

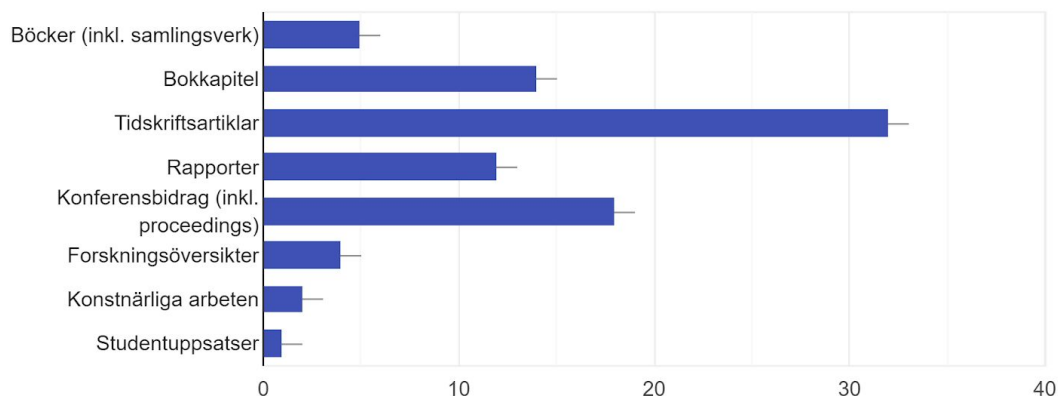


Bild 1. Mest förekommande publikationstyper i de lokala publikationsdatabaserna (exl. avhandlingar)

Nästan hälften av de svarande organisationerna anger sig ha en lokal prioriteringsordning för kvalitetssäkring av publikationer utifrån publikationstypen. Enligt svaren är tidskriftsartiklar inte bara mest förekommande utan också mest kvalitetssäkrade. Lite under hälften anger även avhandlingar, böcker (samlingsverk inkluderade) och bokkapitel att vara mest kvalitetssäkrade. Däremot är konferensbidrag och rapporter inte lika kvalitetssäkrade i förhållande till hur mycket de förekommer i de lokala publikationsdatabaserna. Några anger att alla publikationstyper är lika kvalitetssäkrade utöver de som anges vara övriga publikationer.

Vilka publikationstyper är mest kvalitetssäkrade?

32 svar

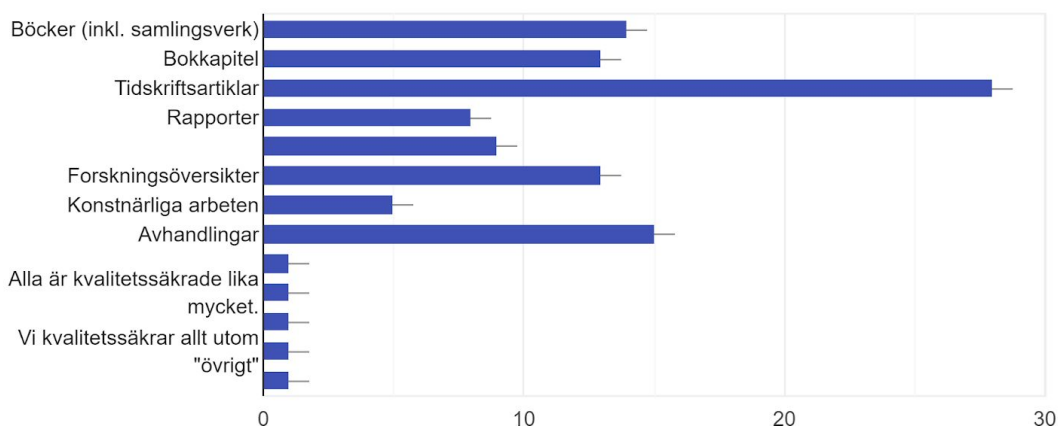


Bild 2. Mest kvalitetssäkrade publikationstyper i de lokala publikationsdatabaserna

Ämnesklassificering

Enligt [Nationella riktlinjer för beskrivning av vetenskaplig output](#) ska ämnesklassificeringen anges minst på nivå 2 utifrån Statistiska centralbyråns (SCB) och Universitetskanslersämbetets (UKÄ) [Standard för svensk indelning av forskningsämnen 2011](#), förutom för konst och medicin där ämnesklassificeringen ska ligga på nivå 3. Ingen av de svarande organisationerna använder sig av prioriteringsordning för kvalitetssäkring av publikationerna utifrån ämnesområdet. Ca 41% av svarande anger ämnesklassificering vanligtvis på nivå 3 (5-siffernivå) och ca 34% på nivå 2 (3-siffernivå). Ca 16% anger det på olika nivåer beroende på ämnesområdet. Ca 9% anger ämnesklassificeringen på nivå 1 (1-siffernivå).

På vilken nivå anges SCB/UKÄ-ämnesklassificeringen vanligtvis enligt lokal praxis?

32 svar

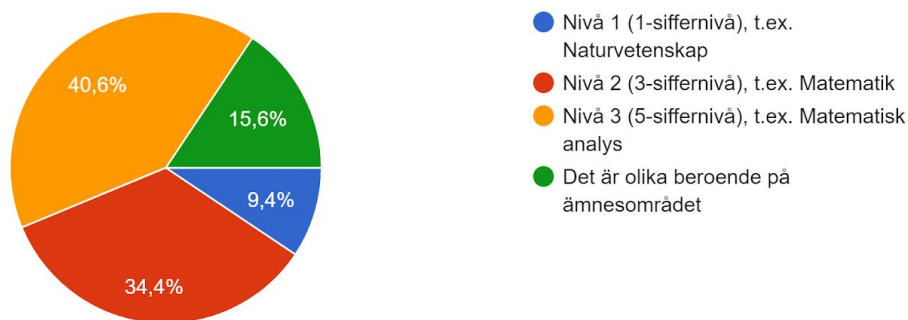


Bild 3. Nivå på ämnesklassificering enligt SCB/UKÄ-standarden som vanligtvis anges i de lokala publikationsdatabaserna

Det som försvårar ämnesklassificeringen är framför allt saknaden av lämpliga ämneskategorier men även bristande ämneskunskap, speciellt inom medicin, naturvetenskap och teknik. Vissa av de svarande organisationerna använder sig av olika hjälpmedel för att klassificera publikationer, till exempel automatgenererad ämnesklassificering.

Ämnesklassificeringar som saknas i SCB/UKÄ-standarden gäller framför allt tvärvetenskapliga ämnen, exempelvis *turismvetenskap* och *estetik*. I vissa fall finns passande kategori men ur en annan aspekt, exempelvis *hälsovetenskap* som klassificeras under medicin men forskas även utifrån andra ämnesaspekter. Standarden upplevs ojämn då andra ämneskategorier är finfördelade medan andra har kategoriserats grovt. Konst är inte välrepresenterad och kategorierna blandar forskning på konstnärlig och vetenskaplig grund. De organisationer som är koncentrerade på ett specifikt ämne upplever det svårt att hitta lämpliga kategorier om den lokala forskningen tillämpar olika ämnesområden utifrån sitt perspektiv, exempelvis *idrott* och *krigsvetenskap*. Mer tillämpande instruktioner efterfrågas i

linje med "här ingår...", till exempel för att ytterligare beskriva skillnaden mellan *juridik* samt *juridik och samhälle*.

Metadatakrav

Det som upplevs svårast att kvalitetssäkra utifrån [de krav som Vetenskapsrådet](#) ställer för att data ska kunna inkluderas vid bibliometriska analyser är framför allt affilieringar (ca 50%). Delvis är detta på grund av att affilieringen inte alltid framgår ur publikationen, närmast i bokkapitel och böcker, delvis att det är tidskrävande att ange affiliering även till alla externa. Dessutom kan externa affilieringar ofta anges enbart som fritext på grund av saknaden av en generell organisationslista. Speciellt svårt anses affilieringen av externa upphovspersoner vara för de publikationer som inte täcks av Scopus och Web of Science. I vissa fall är det inte ens tekniskt möjligt att ange ytterligare affilieringar.

Kravet på ämnesklassificeringen på nivå 2 anges som andra svårast att uppfylla, på grund av de ovan nämnda orsakerna. Även kravet på DOI anses vara svårt eftersom DOI inte alltid finns utdelat för alla publikationer eller information om detta saknas, främst för organisationernas egna publikationer och konferensbidrag. Anmärkningsvärt är att kravet på ISSN anses inte svårt att uppfylla trots att alla seriella resurser inte har blivit tilldelade ISSN.

Utifrån Vetenskapsrådets beställarunderlag, vad är svårast för er att uppfylla?

27 svar

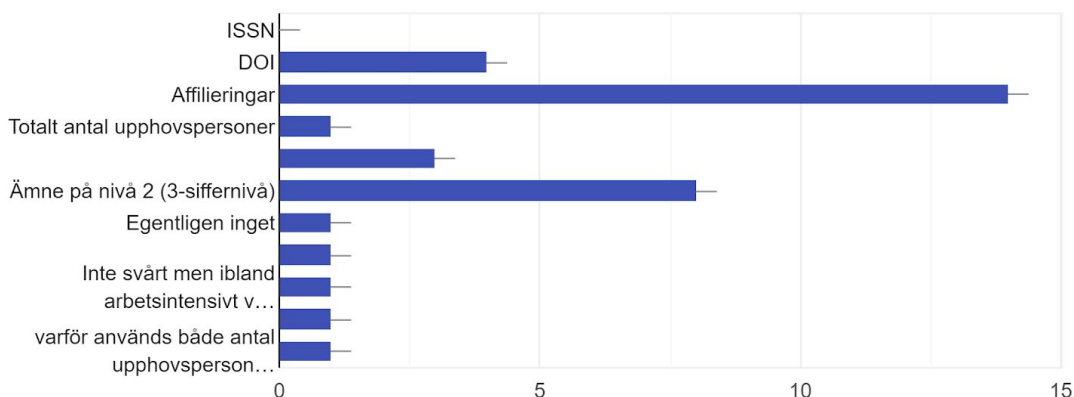


Bild 4. De svåraste kraven enligt Vetenskapsrådets beställarunderlag

Anmärkningar

Det som poängterades utifrån den lokala kvalitetssäkringen var att i många fall är äldre poster inte kvalitetssäkrade eller ämnesklassificerade på nivå 2 och att ändrade praxis och organisationsstrukturer påverkar kvalitet på posterna. Den typ av problematik gäller främst

institutionsnamn, ämneskategorisering och publikationstyper. Överlag dock är det bra koll på posterna.

Särskilt problematiskt att kvalitetssäkra uppges vara identifikatorer, externa affilieringar och konferensbidrag. Att dra linjen mellan övrigt vetenskapligt och populärvetenskapligt material upplevs också svårt. Saknaden av standarder för uppmärkningen av öppen tillgång anses vara en försvårande faktor för registrering och därmed uppföljning av öppen tillgång.

Ibland går inte heller publikationen ihop med verkligheten. Enligt de nationella riktlinjerna ska registreringen utgå från publikationen, men publikationen kan skilja sig från faktiska förhållanden, särskilt vanligt förekommande vid affiliering. Det kan också saknas en lämplig kategori för den typ av publikation. Som problematiskt upplevs också att data från originalposten inte alltid finns med i den post som har slagits ihop vid dubbelthantering i Swepub sök.

Önskemål

Önskemålen gällande kvalitetssäkringen i de lokala publikationsdatabaserna samt Swepub handlade framför allt om tid och resurser att utföra arbete med datakvalitetshöjande åtgärder. Dels önskas tid och resurser för att manuellt kontrollera data som har lagts in i de lokala publikationsdatabaserna, även bakåt i tiden, dels för att kunna rätta till poster som har markerats som ofullständiga av Swepub.

Likaså handlar en stor del av önskemålen om tydlig kommunikation och riktlinjer samt verktyg från Swepub. Automatisk ämnesklassificering förväntas vara till hjälp med de problem som upplevs med ämnesklassificeringen men på sikt önskas revidering av SCB/UKÄ-ämnesklassificeringen så att den motsvarar svenska forskningsförhållanden bättre.

Möjligheten att exportera listor på publikationsposter som har markerats som ofullständiga av Swepub uppskattas. Det önskas dock tydligare rapportering kring ofullständiga data samt tydliga riktlinjer om kraven på metadata, till exempel undrar forskande myndigheter om alla krav på metadata ska även gälla dem, inte bara lärosätena. Att kunna validera publikationsposter direkt samt kontinuerliga uppdateringar vore i och för sig ännu bättre. Det önskas även högre support-beredskap hos Swepub. SwepubID nämns också i önskelistan.

Ett kontrollerat organisationsregister skulle underlätta inmatning av externa affilieringar likaså som tidskriftslista skulle underlätta inmatning av publiceringskanaler. Kontrollerade fält för identifikatorer skulle underlätta direkt validering lokalt. Överhuvudtaget önskas att det som gäller nationellt finns fullt ut implementerat i de lokala systemen. Det finns även några önskemål om praktiskt användning av Swepub-data och förtydligande av syftet med de lokala publikationsdatabaserna i förhållande till analyser och CRIS-system (Current Research Information Systems). En sista önskan är ett gemensamt nationellt system för registrering av publikationer.

Dataleveranserna

Det finns osäkerhet kring frågorna rörande dataleveranser. För de organisationer som deltar i DiVA-konsortiet sköter DiVA centralt dataleveranserna till Swepub. De som administrerar de lokala publikationsdatabaserna har lite kännedom hur dataleveranserna ser ut, exempelvis kunde inte alla svara på frågan om raderade poster anges enligt OAI-PMH-protokollet för att stödja selektiv höstning.

I dagsläget utesluter ingen av de svarande organisationerna data från leveranserna. Däremot är det några organisationer som inte för tillfället levererar data på grund av serverbyte eller gör en fullständig dataleverans bara ett par gånger om året. För den här typen av uppehåll önskas möjlighet att markera och kommunicera detta utåt till slutanvändare samt stöd från Swepub att sätta upp dataleveranser.

Önskemål gällande dataleveranser rör framför allt välfungerande och uppdaterad Swepub med tydliga felrapporter och användarvänligt gränssnitt för administrering av dataleveranser för att hjälpa organisationer med datakvalitetshöjande arbete. Informationen om dataleveranser och kraven på metadata samt om processen för hantering av data i Swepub bör vara lättillgänglig och tydlig. Automatisk ämnesklassificering står också på önskelistan. DiVA-organisationer anger välfungerande samarbete mellan DiVA och Swepub som viktigt.

Slutsats

De organisationer som levererar data till Swepub arbetar aktivt med kvalitetssäkringen av data i sina lokala publikationsdatabaser. Det som påverkar datakvalitet handlar om saknaden av informationsunderlaget, exempelvis affilieringar och identifikatorer, saknaden av kategorier, exempelvis vissa ämneskategorier eller publikationstyper, tekniska förutsättningar samt stöd från Swepub i form av tydliga krav, riktlinjer, tekniskt support och verktyg. Den allra största påverkan på datakvalitet har dock tid och resurser, framför allat gällande registrering och kvalitetssäkring av de publikationer som inte finns med i kommersiella databaser som Scopus och Web of Science. Om höjning av datakvaliteten anses vara viktigt av ledningarna vid de organisationer som deltar i Swepub och intressenterna för Swepub, behöver resurser utses även till detta arbete utöver vidareutvecklingen av Swepub.