

UKÄ:s kompletterande beställarunderlag SwePub 190329

Inledning

Universitetskanslersämbetet (UKÄ) är sedan juni 2018 en av intressenterna i arbetet med utvecklingen av Swepub. Publikationsdata från Swepub fyller en lucka i UKÄ:s statistik över högskoleväsendet. Redan i UKÄ:s årsrapport för universitet och högskolor från 2018 användes data från Swepub och UKÄ avser att ytterligare utveckla användandet, exempelvis genom att inkludera Swepub-data i nyckeltal och bakgrundsinformation i samband med myndighetens granskning av lärosätenas kvalitetssäkringsarbete avseende forskning.

Detta beställarunderlag kompletterar beställarunderlaget som lämnades den 18 januari 2019. Inledningsvis redogörs kort för UKÄ:s omedelbara behov i relation till produktionen av statistik över forskningsoutput från landets lärosäten, samt för en långsiktig vision kring hur Swepub kan nyttjas för att utveckla och förbättra produktion och analys av statistik för högskoleväsendet. Därefter följer en genomgång av UKÄ:s analysbehov i relation till de nationella riktlinjerna för beskrivning av forskningsoutput, men endast de fall där ämbetet bedömer att analysbehovet förändrats sedan det förra beställarunderlaget inkluderas och kommenteras.

Omedelbara behov (1-2 år)

Ett komplett och kvalitetssäkrat datauttag ur Swepub med möjlighet till deduplicering och fraktionering (på lärosäte, forskningsämnesområde etc.). För att detta ska vara möjligt krävs uppgifter av god kvalitet om det totala antalet upphovspersoner samt antalet lokala upphovspersoner per publikation.

Exempel på möjliga analyser utifrån ett sådant datauttag:

- Utvecklingen över tid per publikationstyp, lärosäte och forskningsämnesgrupp
- Ovan i kombination med uppgifter om intäkter per forskningsämnesgrupp och lärosäte från SCB (vartannat år)
- Ovan i kombination med uppgifter om antalet forskande och undervisande anställda per forskningsämnesgrupp

Behovet av komplett och kvalitetssäkrade datauttag kan tillgodoses genom att datummärkta datadumpar från Swepub regelbundet tillgängliggörs via en FTP-lösning, exempelvis en gång per månad. Filerna bör ha ett lätthanterligt format som lämpar sig väl för produktion och analys av statistik (förslagsvis csv) och vara uppdelade årsvis. En

lösning likt ovan finns redan på plats men har enligt Kungliga biblioteket (KB) inte uppdaterats sedan november 2017. Enligt KB har det delvis sin förklaring i att de för närvarande arbetar med att bygga om den bakomliggande lösningen (back-end-systemet) till Swepub. Mot bakgrund av UKÄ:s analys- och produktionsbehov anser ämbetet att KB bör tillhandahålla en FTP-lösning likt ovan även efter det att den befintliga infrastrukturen bakom Swepub ersatts.

Vision (2 år-)

De analyser som kan genomföras utifrån Swepub-data i dagsläget innebär viktiga kunskapstillskott, men är relativt begränsade (se exempel ovan). UKÄ:s vision är att utveckla statistiken över forskningsoutput så att den kan kombineras med andra statistikällor och register. Möjligheterna att inkludera personnummer i Swepub, alternativt att berika Swepub med personnummer genom samkörning med andra register, bör därför utredas.

Att personnummer på något sätt kan kopplas till upphovspersonerna i databasen innebär (utöver en förenklad kvalitetssäkring) att Swepub blir en unik resurs, som avsevärt skulle förbättra möjligheterna att kvantitativt analysera och utvärdera forskningen i Sverige. Detta skulle direkt svara mot de behov av analys och uppföljning av forskning som anges i propositionen *Kunskap i samverkan* (prop. 2106/17:50 s. 66-67). Swepub skulle också kunna användas som datakälla inom UKÄ:s arbete med kvalitetssäkring av lärosätenas verksamhet. UKÄ genomför redan utvärderingar av forskarutbildningar vid landets lärosäten och påbörjar under 2020 ytterligare utvärderingar. För detta ändamål kan Swepub på sikt fungera som en väldigt bra datakälla då den täcker in samtliga publikationer.

Exempel på andra analyser som, med hög intern validitet, skulle kunna göras på längre sikt om personnummer inkluderas i Swepub:

- Antal publikationer ordnade efter kön (per forskningsämnesgrupp, lärosäten etc.)
- Antal publikationer ordnade efter anställningsbenämning/anställningskategori (per forskningsämnesgrupp, lärosäten etc.)
- Antal publikationer ordnade efter åldersgrupper och examensår (per forskningsämnesgrupp, lärosäten etc.)
- Effekter på forskningsoutput av strategiska forskningssatsningar (exempelvis centrumbildningar)
- Analys av hur frekventa forskningssamarbeten är mellan svenska lärosäten.
- Effekter på forskningsoutput av att byta lärosäte för enskilda forskare
- Effekter av karriärutveckling (exempelvis från universitetslektor till professor eller liknande)

Om personnummer inte inkluderas i Swepub bör möjligheterna att berika databasen med personnummer förenklas så långt det är möjligt. Av detta skäl förordar UKÄ att de personliga identifierare (såsom ORCID och lokalt personID) som redan finns i databasen behålls, samt att KB på sikt utreder möjligheterna att skapa unika personliga identifierare

för samtliga upphovspersoner i Swepub. Detta gäller alltså även upphovspersoner som av någon anledning saknar lokalt personID eller ORCID. UKÄ anser vidare att användningen av elementet ”2.2 Födelseår” skulle kunna göras obligatorisk för att ytterligare förenkla beräkning av Swepub genom samkörning med andra register.

Om beräkning med eller inkludering av personnummer inte är möjligt anser UKÄ att variablerna kön och anställningsbenämning/anställningskategori ska registreras vid inrapportering av publikationer till Swepub. Då UKÄ avser att analysera och producera statistik över forskningsoutput utifrån Swepub bör det, för att öka jämförbarheten med andra statistikprodukter från UKÄ, åtminstone finnas möjlighet till en uppdelning efter kön och anställningsbenämning/anställningskategori.¹ Analysmöjligheterna skulle i så fall begränsas i förhållande till UKÄ:s vision men ändå innebära en förbättring.²

Nationella riktlinjer för beskrivning av forskningsoutput - analysbehov

Genomgången nedan tar sin utgångspunkt i Vetenskapsrådets (VR) bedömning av analysbehovet i relation till de nationella riktlinjerna för forskningsoutput, samt från UKÄ:s beställarunderlag inlämnat i januari 2019. UKÄ instämmer i huvudsak med VR:s analysbehov, men har kompletterat behoven där det ansetts önskvärt. UKÄ har valt att endast kommentera de element där analysbehovet förändrats eller utvecklats i förhållande till beställarunderlaget som inlämnades i januari 2019.

Beslut i detta ärende har fattats av avdelningschef Annika Pontén efter föredragning av utredaren Martin Söderhäll.


Annika Pontén
Martin Söderhäll

¹ Om officiell statistik om forskningsoutput vid svenska lärosäten ska produceras, krävs (om det inte finns särskilda skäl mot detta) en uppdelning på kön, se 14 § Förordning (2001:100) om den officiella statistiken.

² Utvecklingen över tid per publikationstyp, lärosäte, forskningsämnesgrupp, kön och anställningsbenämning/anställningskategori kan analyseras. Analyserna kan kombination med uppgifter om intäkter per forskningsämnesgrupp och lärosäte samt med uppgifter om antalet forskande och undervisande anställda per forskningsämnesgrupp.



2.2 Födelseår

Swepub-definition:	Uppgift om födelseår kan läggas till personer.
Användning:	Frivilligt.
Swepub MODS:	13.1.1 Name part
Analysbehov:	Inget analysbehov har uttryckts.
UKÄ 1-2 år	Bör vara obligatoriskt, för att förenkla samkörning med och berikning från andra register.
UKÄ 2 år-	På sikt önskvärt med personnummer.

2.4 PersonID

Swepub-definition:	Uppgift om ett unikt ID som identifierar upphovspersonen.
Användning:	Obligatoriskt. <i>Uppgift om personID är obligatorisk information för den egna organisationens upphovspersoner.</i>
Swepub MODS:	13.1.2 Name identifier
Analysbehov:	Lokalt personID eller ORCID är viktiga för analyser som baserar sig på personidentifikatorer.
UKÄ 1-2 år	UKÄ eniga med VR angående analysbehov. ORCID och lokalt personID bör båda vara obligatoriska. På sikt önskvärt med personnummer (åtminstone behov av kön och anställningsbenämning/kategori.).
UKÄ 2 år-	Personnummer.

2.6 Antal upphovspersoner

Swepub-definition:	Uppgift om antal upphovspersoner (ett numeriskt värde).
Användning:	Obligatoriskt för upphovspersoner som skall medräknas i analys.
Swepub MODS:	14 Creator count
Analysbehov:	I samband med analys och särskilt resursfördelning fraktioneras antalet upphovspersoner. I uppkomna fall, särskilt för artiklar inom fysik, har antalet författare varit såpass stort att alla författare inte kunnat uppges. Därför införs ett fält med summan av antalet upphovspersoner.
UKÄ:	Antal upphovspersoner behövs. Även antalet lokala upphovspersoner

3.22 Ämne

Swepub-definition:	Kontrollerad ämnesklassificering enligt <i>Standard för svensk indelning av forskningsämnen</i> .
Användning:	Obligatoriskt.
Swepub MODS:	8.1 Subject classification according to Swedish research subjects standard
Analysbehov:	Fältet har angetts relevant för analysbehov.
UKÄ:	UKÄ har inga invändningar mot att en publikation klassificeras till fler än ett forskningsämne. Men då måste det vara möjligt att fraktionera på forskningsämne. För statistikbehov behövs 3-siffernivå. UKÄ bör som ansvarig för Standard för svensk indelning av forskningsämnen vara med i diskussionen kring ämnen. UKÄ instämmer i KB:s bedömning att registreringen av forskningsämne är viktig eftersom ett flertal av publikationsposterna i Swepub saknar tillräckligt underlag (abstract, titel etc.) för automatiskt genererad klassificering.