



Kvalitet och publikationsdatabaser

Rekommendationer för att anpassa de svenska lärosätenas publikationsdatabaser till att utgöra underlag för bibliometriska analyser och ekonomisk resursfördelning.

Arbetsgruppen för kvalitetssäkring av publikationsdatabaser på uppdrag av
Forum för bibliotekschefer, SUHF - Sveriges universitets- och högskoleförbund
Slutrapport - augusti 2010

Forum för bibliotekschefer inom SUHF (Sveriges universitets- och högskoleförbund) tillsatte hösten 2009 en arbetsgrupp med uppdraget att ta fram riktlinjer för kvalitetssäkring av publikationsdatabaser vid svenska universitet och högskolor.

Arbetsgruppen konstaterar i denna rapport att de lokala publikationsdatabaserna, genom sin bredare täckning av framför allt de humanistiska och samhällsvetenskapliga forskningsområdena och genom sitt större spektrum av dokumenttyper, kan utvecklas till att ge analys- och medelsfördelningsmöjligheter utöver vad de idag använda internationella databaserna ger. För att förverkliga detta krävs ett kvalitetsarbete inriktat både på åtgärder lokalt vid lärosätena och på justeringar i det nationella överföringsformatet, med vilket data förs till centrala samlingsdatabaser.

Denna rapport innehåller rekommendationer inriktade på ökad kvalitet i lokal täckning och bibliografisk standard, verifikation och nationell överföring.

Arbetsgruppen föreslår att SUHF:

- arbetar för att de lokala databaserna skall utgöra underlag för analys och medelsfördelning på nationell nivå,
- rekommenderar sina medlemmar att anpassa sina publikationstjänster enligt denna rapports rekommendationer,
- initierar en vidareutveckling av formatspecifikationen för SwePub MODS på nationell nivå genom att en nationell samordningsgrupp utses.

Arbetsgruppen för kvalitetssäkring av publikationsdatabaser

Arbetsgruppen tillsattes hösten 2009 av Forum för bibliotekschefer inom SUHF (Sveriges universitets- och högskoleförbund), med uppdraget att ta fram riktlinjer för kvalitetssäkring av publikationsdatabaser vid svenska universitet och högskolor. Gruppen hade sitt första möte i november 2009 och har därefter arbetat med uppdraget under våren 2010.

Arbetsgruppens medlemmar:

Håkan Carlsson, Göteborgs universitet (sammankallande)

Jörgen Eriksson, Lunds universitet, Biblioteksdirektionen

Leif Eriksson, Uppsala universitetsbibliotek

Tore Lund, Chalmers bibliotek

Catharina Rehn, Karolinska institutet

Aina Svensson, Uppsala universitetsbibliotek

Ulf Kronman, Vetenskapsrådet¹

För att få en större kunskapsbredd när det gäller formatfrågor och teknisk kompetens och även bredare erfarenheter av forskningsanalyser och fördelningsmodeller på såväl lokal som nationell nivå tillsatte arbetsgruppen en referensgrupp. Referensgruppen har sammanträtt vid två tillfällen och delat med sig av värdefulla kommentarer och synpunkter.

Referensgruppens medlemmar:

Stefan Andersson, DiVA, Uppsala universitetsbibliotek

Björn Forslöw, Karolinska institutet

Johan Fröberg, Vetenskapsrådet

Magnus Gunnarsson, Vetenskapsrådet

Jessica Lindholm, DSpace, Malmö högskola

Henrik Lindström, LIBRIS (SwePub), Kungl. biblioteket

Anders Malmberg, Uppsala universitet

Kristin Olofsson, LIBRIS (SwePub), Kungl. biblioteket

¹ Ulf Kronman, Vetenskapsrådet, har deltagit i gruppen som teknisk rådgivare i frågor kring datahantering och bibliometriska analyser. Han har inte deltagit i arbetsgruppens rekommendationer vad gäller metoder för medelsfördelning med bibliometriskt underlag.

1. Inledning

1.1. Bakgrund till uppdraget

Det har under senare år blivit allt vanligare att använda bibliometriska analyser som underlag för utvärdering av vetenskaplig produktion och dess kvalitet. För detta används en rad olika modeller och resultatet av sådana analyser kan sedan användas för bland annat ekonomisk fördelning. Den modell som för närvarande används i Sverige på nationell nivå för att fördela statliga forskningsmedel till lärosäten, bygger till viss del på publiceringsstatistik som hämtas från Thomson Reuters citeringsdatabaser, tillgängliga via söktjänsten Web of Science. Web of Science täcker för närvarande drygt 11 000 vetenskapliga tidskrifter och har god täckning för publicering inom medicin, naturvetenskap och teknik. Inom andra vetenskapsområden återfinns de mest högkvalitativa forskningsresultaten i mindre grad i vetenskapliga tidskrifter. Den stora delen av publiceringen inom humaniora och samhällsvetenskaper görs i form av antologier, monografier eller tidskrifter som till största del saknas i Web of Science.

Underlaget från Web of Science fungerar således olika bra för olika ämnesområden och inte alls för att ge en heltäckande bild av t.ex. humanistisk och samhällsvetenskaplig forskning. Risken med att inte inkludera publicering i etablerade källor inom dessa ämnesområden i analyser är att synligheten av Sveriges forskning minskar då man kan befara att svensk publicering anpassar sig till rådande analysystem. Möjligheten att hämta in fullständiga uppgifter från forskarna och samla dem i en nationell publikationsdatabas med syftet att ge en mer fullständig bild av aktuell svensk forskning inom samtliga ämnesområden är därför något som har efterfrågats både av företrädare för lärosätena och av Vetenskapsrådet. Under 2000-talet har publikationsdatabaser gjorts tillgängliga vid alla större lärosäten. I dessa sammanställs lärosätenas vetenskapliga produktion. Sedan november 2009 överförs innehållet i de allra flesta fall till SwePub, en söktjänst för forskningspublikationer från svenska universitet och högskolor som har utvecklats av Enheten för LIBRIS och databaslicenser vid Kungl. biblioteket. Den information som visas i SwePub är därför i allra högsta grad beroende av vilka data som levereras från de lokala databaserna. Om en nationell samlingsdatabas, förutom att fungera som söktjänst, även ska kunna användas som underlag för olika bibliometriska analyser är det således viktigt att de data som levereras till SwePub har tillräckligt hög kvalitet.

SUHF (Sveriges universitets- och högskoleförbund) har sedan en tid tillbaka engagerat sig i frågan om publikationsdatabaser och kvalitetssäkring. SUHF:s styrelse beslutade den 17 juni 2009 att rekommendera sina medlemmar att ha eller vara ansluten till en publikationsdatabas, som kan leverera till SwePub. Styrelsen gav samtidigt Forum för bibliotekschefer i uppdrag att ta fram mer detaljerade riktlinjer för kvalitetssäkring, med Vetenskapsrådet som inbjuden deltagare.² Arbetsgruppen för kvalitetssäkring av publikationsdatabaser fick i första steget i uppdrag av Forum för bibliotekschefer att föreslå hur arbetet med kvalitetssäkring skulle ske och hur detta arbete borde förankras vid

² Rekommendation om tillgång till publikationsdatabas (REK 2009:3): <http://www.suhf.se/web/REK2009-3.aspx>

lärosätena. I nästa steg fick arbetsgruppen fortsatt förtroende att också genomföra uppdraget.

1.2. Inventering och informationsinhämtning

Arbetsgruppen har arbetat enligt den arbetsplan som presenterades för och godkändes av styrgruppen för Forum för bibliotekschefer den 2 december 2009 och som finns tillgänglig på SUHF:s hemsida.³

För att samla så bred kompetens och kunskapsbredd som möjligt har målet som första steg varit att inventera och samla in information och erfarenheter från olika relevanta källor. Information om arbetsgruppens uppdrag har dessutom spridits via biblioteksforum och nationella e-postlistor för bibliometriker för att om möjligt få så mycket input som möjligt till gruppens arbete.

Följande inventeringar har genomförts och resultaten av dessa beskrivs i avsnitt 2 och 3 nedan:

- Nordisk utblick: Kartläggning av situationen från liknande projekt i Norge och Danmark kring arbetet med att utveckla och använda nationella publikationsdatabaser för publiceringsanalyser.
- Bibliometriska analyser idag: Inhämtning av erfarenheter och uppgifter om vilka nuvarande och planerade bibliometriska analyser som görs på lokal nivå av bibliometriker vid svenska lärosäten.
- Underlag för lokal medelstildelning: Inhämtning av information om vilka fördelningsmodeller som för närvarande används vid svenska lärosäten.
- SwePub: Analys och närgranskning av SwePubs nuvarande dataformat, täckning och kvalitetsnivå.

³ http://www.suhf.se/web/Arbetsgrupper_inom_Forum_for_bibliotekschefer.aspx

2. Nordisk utblick

I Norge och Danmark har man infört system för fördelning av forskningsmedel på nationell nivå, där en del baseras på egenregistrerad vetenskaplig publicering, vilket också har efterlysts i Sverige. Finland har också tagit beslut om ett liknande system. I samband med införandet i Norge och Danmark har en diskussion förts, med klara paralleller till den nuvarande svenska situationen, om hur man ska kvalitetssäkra publiceringsdata för att få ett så komplett underlag som möjligt.

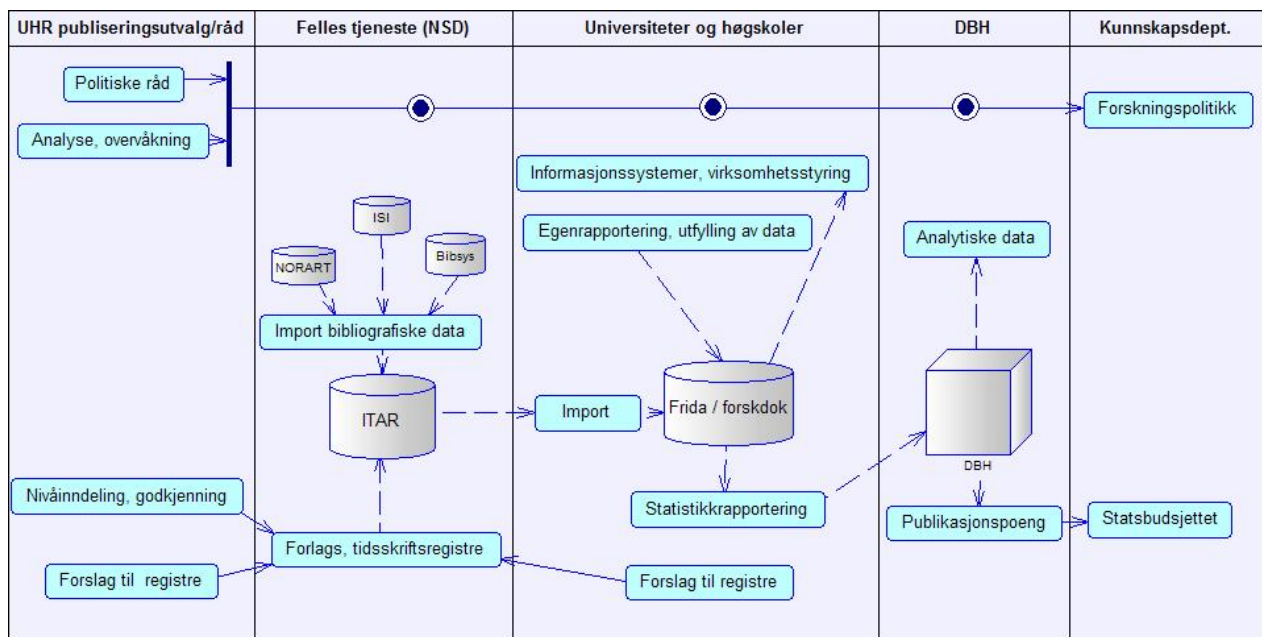
2.1. Norge

Norge har haft ett system för fördelning av forskningsmedel på nationell nivå sedan 2006 som innefattar en del (ca 2 %) av den totala direkta statliga finansieringen som baseras på publicering. Man tar hänsyn till tre publikationstyper; bok, artikel i antologi samt artikel i tidskrift. Även vissa konferensbidrag medräknas. Kriterierna för att komma med i bedömningen är att publikationen presenterar en ny insikt, att den föreligger i en form som gör att resultaten är möjliga att verifiera eller användbara i ny forskning, att den är på ett språk och har en distribution som gör den tillgänglig för forskarsamhället samt att den är peer-reviewgranskad.

För samtliga publikationstyper sker en bedömning av de *kanaler* som publiceringen sker i, dvs. *förlag* för böcker och *tidskrifter* för artiklar. Kanalerna delas in i två nivåer, nivå 1 och nivå 2. Nivå 1 ger en lägre poäng och är upplagd så att den omfattar ca 80% av publiceringarna på de båda nivåerna inom ett visst ämnesområde på världsbasis. Nivå 2, som ger högre poäng, omfattar således ca 20%. Bedömningen av kanalerna görs fortlöpande av nationella ämnesråd (för närvarande 47 st).

För att administrera detta har Universitets- och högskolerådet (UHR) etablerat ett system för forskningsdokumentation. Systemet omfattar en central funktion (förvaltas av Felles tjeneste vid Norsk samfunnsvitenskaplige datatjeneste, NSD) som har till uppgift att leverera bibliografiska data om vetenskaplig publicering samt tillhandahålla ett auktoritetsregister som stöd för universitetens och högskolornas egna system (Frida/Forskdok). Man importerar bibliografiska data från tre källor, Web of Science, BIBSYS samt Norart (Norske og nordiske tidsskriftartikler), till en gemensam databas, ITAR. Från denna kan sedan forskare importera poster till sitt lokala system⁴ och i samband med detta även verifiera posterna. I de lokala systemen sker också egenrapportering av det material som inte finns med i ITAR. Posterna kontrolleras sedan även av speciella "superbrukere" på olika administrativa nivåer av lärosätet, som avdelning, institution, fakultet o.s.v.. En statistikrapportering sker sedan från Frida/Forskdok till Database for statistikk om høgre utdanning (DBH) där sedan publikationspoängen beräknas, som också blir budgetunderlaget för Kunnskapsdepartementet (Fig. 1).

⁴ Med lokalt system avses i detta fall ett låst område (Virtual Private database) i en gemensam Oracle-databas.

Figur 1. Arbetsfördelning och arbetsflöde för den norska fördelningsmodellen⁵

Genom att man förser de lokala systemen med data från den centrala auktoritetsdatabasen ITAR minskar risken för att dubbla registreringar görs. Trots detta upplever normmännen allvarliga problem med dubletter i systemet och i en utvärderingsrapport 2006 påpekas det att man måste göra något åt problemet med dubletter.

2011 planerar man att lansera ett nytt system för att dokumentera vetenskaplig publicering, Norsk vitenskapindeks. Detta kommer att medföra att samtliga forskningsinstitutioner, inklusive hälsosektorn, kommer att få en gemensam plattform för forskningsinformation. Det innebär även att man får en betydande kvalitetshöjning av innehållet i systemet; bl.a. kommer artiklar med flera institutionstillhörigheter inte längre att visas dubbelt utan som en post. Systemet kommer även att innehålla utökad användarstöd i syfte att höja datakvaliteten.

2.2. Danmark

Danmark är på väg att införa ett system som till stora delar liknar det norska vad gäller kriterier för bedömning av publicering. Systemet ska börja gälla 2010 men kommer att fasas in successivt.

När det gäller systemlandskapet skiljer sig det danska systemet från det norska på ett flertal punkter. Danmark har en lång tradition av lokala system för publicering på högskolorna och på senare år har två system utkristalliserat sig som dominerande, ORBIT och Pure. Det har medfört att man i stället för att registrera i en central databas som i Norge, litar till att

⁵ <http://dbh.nsd.uib.no/vitpub/dokumentasjon.jsp>, NIFU STEP, Rapport 33/2008

bibliografiska data ska föras över från de lokala systemen till en central databas, Den Danske Forskningsdatabas (DDF), vilket påminner mer om den svenska situationen med livskraftiga lokala databaser och en nationell samlingsdatabas.

För att kvalitetssäkra de bibliografiska uppgifterna använder man dels nationella hjälpdatabaser med auktoritetsdata på liknande sätt som i Norge, och dels nationella standarder när det gäller metadata. De standarder som används är dels OAI-PMH som protokoll för själva överföringen samt DDF-MXD (Danish Research Database – Metadata Exchange Format for Documents) som gemensam standard för det nationella överföringsformatet.⁶ DDF-MXD omfattar ett antal standardiserade vokabulärer bl.a. för dokumenttyp, typ av innehåll samt vetenskaplig nivå (referentbedömt, populärvetenskapligt etc).

Valideringen i det danska systemet sker automatiskt via DDF-MXD i samband med överföringen från de lokala systemen till DDF på ett liknande sätt som i SwePub. Dock räcker inte alltid den automatiska valideringen till och man pekar på att det bör finnas ansvariga personer som kontrollerar metadata både på lokal och på central nivå.

Deduplicering är ett återkommande problem. Det finns idag ingen mekanism för att upptäcka dubletter i DDF. Det kan komma in dubletter i DDF både från en enskild lokal publikationsdatabas eller från flera olika system i samband med samförfattande. Ett utvecklingsarbete måste till för att eliminera dubletterna, t.ex. genom att man från de lokala databaserna enkelt kan söka i den centrala basen före registrering. Man bedömer dock att en manuell genomgång av posterna i den centrala basen måste göras för att helt kvalitetssäkra data innan de används för beräkning av forskningsanslaget.

På sikt diskuteras emellertid att man måste utveckla en ny dansk forskningsdatabas liknande den i Norge, men på grund av att man relativt snabbt måste kunna leverera underlag till det nya fördelningssystemet är denna fråga skjuten på framtiden.

⁶ http://mx.forskningsdatabasen.dk/mxd/1.1.0/DDF_MXD_v1.1.0.pdf

3. Läget idag i Sverige

3.1. Dagens infrastruktur

Sveriges infrastruktur när det gäller publicering och redovisning av forskningspublikationer liknar på många sätt den som finns i Danmark genom att det även här finns en tradition av publiceringssystem vid de lokala lärosätena. De första lokala systemen utvecklades och startades omkring år 2000 med syftet att tillgängliggöra publikationer digitalt i fulltext, i första hand avhandlingar och examensarbeten. Genom lärosätenas regleringsbrev 2003, där Utbildningsdepartementet efterfrågar redovisning av forskningspublicering vid svenska lärosäten, uppstod även behovet av lokala databaser för redovisning av all vetenskaplig publicering. Detta har under senare år inneburit att systemen har utvecklats så att de flesta databaser idag är integrerade publiceringsdatabaser där syftet är såväl digital fulltextpublicering som bibliografisk redovisning av publikationer. Vid omkring hälften av lärosätena har beslut tagits om obligatorisk registrering av forskningspublikationer i den lokala databasen. En kartläggning från 2008 visade att 17 lärosäten då hade sådana beslut.⁷

I tabell 1 redovisas en lista över Sveriges universitet och högskolor som är medlemmar i SUHF, och de publikationsdatabaser som används för publiceringsredovisning (maj 2010). Av SUHF:s 41 medlemmar är det 35 lärosäten som har eller är anslutna till en publikationsdatabas. Sex lärosäten saknar sådana databaser. De olika publiceringssystem som används för bibliografisk registrering och redovisning är DiVA (25), DSpace (2), Scigloo (2), Pure (1) samt lokalt egenutvecklade system (5). Av de 35 lärosäten som idag har eller är anslutna till en publikationsdatabas är det 28 som i dagsläget levererar till SwePub. De 7 lärosäten som har lokala databaser men ännu inte levererar till SwePub är antingen helt nyanslutna och har ännu inte kommit igång med registrering lokalt eller också arbetar de med anpassning av det lokala systemet för att möjliggöra leverans.

Tabell 1. Tabell över SUHF:s 41 medlemmar och de databaser som används för redovisning av lärosätets publicering (bibliografisk redovisning) samt vilka lärosäten som levererar data till SwePub (maj 2010). Tabellen visar att 35 av lärosätena har eller är anslutna till en publikationsdatabas och att 28 av dessa levererar till SwePub. 6 lärosäten saknar databaser.

Lärosäte	System för redovisning av lärosätets publicering (maj 2010)	Levererar till SwePub (maj 2010)	Kommentarer
Blekinge tekniska högskola	Egenutvecklat	Ja	
Chalmers tekniska högskola	Scigloo	Ja	
Dans- och cirkushögskolan	-	-	
Dramatiska institutet	-	-	
Ersta Sköndal högskola	DiVA	Nej	Nyanslutna till DiVA våren 2010
Försvvarshögskolan	DiVA	Ja	

⁷ Kartläggning av högskolans elektroniska publicering och rekommendationer för nationell samverkan: http://www.suhf.se/web/E-publicering_jan-08.as

Gymnastik- och idrottshögskolan	DiVA	Ja	
Göteborgs universitet	Scigloo (+ DSpace för fulltextpublicering)	Ja	
Handelshögskolan i Stockholm	DiVA	Nej	Planerar lokal lansering av DiVA hösten 2010
Högskolan Dalarna	Egenutvecklat	Ja	
Högskolan i Borås	DSpace	Ja	
Högskolan i Gävle	DiVA	Ja	
Högskolan i Halmstad	DiVA	Ja	
Högskolan i Jönköping	DiVA	Ja	
Högskolan i Skövde	DiVA	Ja	
Högskolan Väst	DiVA	Ja	
Högskolan Kristianstad	DiVA	Ja	
Högskolan på Gotland	DiVA	Ja	
Karlstads universitet	DiVA	Ja	
Karolinska institutet	Egenutvecklat ⁸	-	Förstudie om publikationsdatabas (som också kan leverera till Swepub) pågår, beräknas vara klar hösten 2010
Konstfack	-	-	
Kungliga Konsthögskolan	-	-	
Kungliga Musikhögskolan i Stockholm	DiVA	Nej	Nyanslutna till DiVA våren 2010
Kungliga Tekniska högskolan	DiVA	Ja	
Linköpings universitet	DiVA	Ja	
Linnéuniversitetet	DiVA	Ja	
Luleå tekniska universitet	Pure	Ja	
Lunds universitet	Egenutvecklat	Ja	
Malmö högskola	DSpace	Ja	
Mittuniversitetet	DiVA	Ja	
Mälardalens högskola	DiVA	Ja	
Operahögskolan i Stockholm	-	-	
Röda Korsets högskola	DiVA	Nej	Nyanslutna till DiVA våren 2010
Sophiahemmet högskola	DiVA	Nej	Nyanslutna till DiVA våren 2010
Stockholms universitet	DiVA	Ja	
Sveriges lantbruksuniversitet	Egenutvecklat	Nej	Leverans till SwePub beräknas

⁸ Karolinska institutets bibliometridatabas innehåller information om uppskattningsvis 90% av lärosätets peer-reviewgranskade forskningspublikationer och baseras uteslutande på data från Thomson Reuters citeringsdatabaser. Lärosätet har i dagsläget ingen publikationsdatabas där kompletterande material kan läggas in.

	(+ EPrints för fulltextpublicering)		vara möjligt slutet av 2010
Södertörns högskola	DiVA	Ja	
Teaterhögskolan i Stockholm	-	-	
Umeå universitet	DiVA	Ja	
Uppsala universitet	DiVA	Ja	
Örebro universitet	DiVA	Ja	

3.2. Nuvarande nationellt dataformat och täckningen i SwePub

Under arbetet med SVEP-rekommendationerna⁹ och sedan i samband med utvecklingen av söktjänsten SwePub har ett svensk nationellt dataformat¹⁰ utarbetats. Detta utnyttjas nu vid överföringen av publikationer från lärosätena till den nationella SwePub-databasen. Som grund till överföringsformatet har XML-formatet *Metadata Object Description Schema*, MODS, tagits. De poster som hämtas in till SwePub valideras mot formatet och systemet kontrollerar att fält som ingår i formatets miniminivå har innehåll innan det importerar en publikationspost. Följande fält är för närvarande angivna i överföringsformatet som obligatoriska:

- Titel
- Upphovsman
- Ansvarig organisation
- Lärosäteskod (ansvarig för leverans av datapost)
- Publikationstyp
- Innehållsanmärkning enligt den kontrollerade ämnesklassifikationen SVEP
- Utgivningsuppgift (publiceringsår)
- Identifierator (länk) till lokal post
- Språk
- Värdepublikation (för artikel/forskningsöversikt/recension/kapitel/konferensbidrag)

I praktiken avviker SwePub:s valideringsfilter något från specifikationen genom att det släpper igenom poster utan innehållsmärkning enligt ämnesklassifikationen SVEP, eftersom en del lärosäten inte levererar den informationen över huvud taget.

För att göra en grov skattning av täckningsgraden för de forskningspublikationer som finns registrerade i de lokala databaserna och som därmed redovisas i SwePub, gjordes en rundfrågning i januari 2010 bland de lärosäten som levererar till SwePub. Med täckningsgrad menas här andel forskningspublikationer av lärosätets uppskattade totala produktion. Resultatet gav svar från 21 lärosäten och följande uppgifter om grad av täckning av lärosätets publicering:

⁹ http://www.kb.se/Dokument/Om/projekt/SVEP/rekommendationer_universitet.pdf

¹⁰ SwePub Formatspecifikation: http://www.kb.se/dokument/Libris/teknisk_information/swepub_format.pdf

- 75 % av lärosätena rapporterar relativt god täckning av lärosätets publicering från år 2004 och framåt.
- Mer än hälften skattar att databasen täcker mer än 80 % av lärosätets vetenskapliga publicering.
- 75 % av lärosätena har någon form av kvalitetskontroll kopplad till registreringen av poster i databasen.

Som en del i inventeringen gjordes också en kvantitativ analys av hur stort antal publikationer, som i nuläget levereras in till Swepub. Som det framgår i tabell 2 gjorde 27 lärosäten sina publikationer tillgängliga för höstning till Swepub.

Tabell 2. Analys av materialet i SwePub. Analysen är gjord på data från 2010-02-10 och reflekterar inte nödvändigtvis databasens utseende just nu.

Lärosäte	Antal publikationer i Swepub	Första år med mer än 70% av 2009-års antal publikationer	Antal publikationer i Swepub 2009	Andel poster i Swepub 2009 med svep-kategorin	Andel poster i Swepub 2009 med abstrakt	Andel poster i Swepub 2009 med författarfyllningar	Andel poster i Swepub 2009 med lokala organisationsidentiteter	Andel poster i Swepub 2009 med inom-universitetsdubletter enligt Libris	Antal referensskade artiklar i Swepub 2009	Andel referensskade artiklar i Swepub 2009 med ISSN	Antal böcker och bokkapitel i Swepub 2009	Andel böcker och bokkapitel med förlag i Swepub 2009
Blekinge	3014	2002	221	99%	95%	0%	0%	0,5%	64	95%	16	0%
Chalmers	21283	2004	2146	100%	53%	30%	100%	0,7%	811	34%	65	20%
Dalarna	2328	2003	243	96%	44%	26%	0%	0,4%	76	0%	29	100%
Försvar	12	2009	5	20%	20%	100%	100%	0,0%	1	100%	2	100%
Gymnastik	419	2002	21	86%	62%	100%	100%	0,0%	14	100%	2	100%
Göteborg	42489	2004	5671	100%	52%	28%	100%	1,3%	2681	68%	721	15%
Borås	2076	2007	531	98%	32%	0%	0%	0,8%	111	95%	94	88%
Gotland	94	2009	24	96%	21%	71%	100%	0,0%	4	100%	12	25%
Halmstad	1014	2004	138	97%	84%	95%	100%	1,4%	55	91%	17	100%
Gävle	2476	2004	217	54%	46%	63%	100%	0,9%	66	88%	33	91%
Kalmar*	196	2004	21	81%	76%	90%	100%	0,0%	15	100%	2	100%
Skövde	1297	2004	171	96%	89%	100%	100%	0,6%	70	99%	13	100%
Jönköping	6236	2004	547	48%	32%	63%	100%	1,1%	181	93%	78	88%
Kristianstad	836	2007	205	94%	54%	95%	100%	0,0%	64	98%	36	94%
Väst	501	2006	84	98%	33%	73%	100%	0,0%	36	33%	21	100%
Karlstad	428	2008	107	95%	61%	99%	100%	0,9%	7	100%	26	100%
KTH	5486	2000	487	55%	74%	96%	100%	0,6%	95	91%	2	100%
Linköping	32589	2002	2995	98%	75%	66%	100%	1,7%	1522	99%	267	90%
Luleå	18220	2003	1287	0%	69%	100%	0%	0,9%	0		99	95%
Lund	67553	2002	5291	100%	76%	0%	0%	1,9%	3322	97%	275	93%
Malmö	1898	2007	335	90%	50%	0%	0%	2,7%	114	22%	47	85%
Mälardalen	2629	2004	198	69%	63%	81%	100%	2,0%	66	83%	26	96%
Mittuniv	6182	2002	571	98%	60%	45%	100%	1,6%	183	98%	108	93%
Örebro	3504	2006	464	87%	52%	82%	100%	2,8%	163	87%	77	94%
Södertörn	180	2000	16	25%	31%	100%	100%	0,0%	0		13	100%
Stockholm	19931	2007	3909	55%	43%	69%	100%	1,0%	1718	88%	797	89%
Umeå	18091	2004	1801	34%	67%	70%	100%	2,5%	1106	93%	204	85%
Uppsala	15869	2007	3030	74%	58%	87%	100%	0,0%	1552	96%	427	91%
Växjö*	2723	2006	446	83%	46%	64%	100%	1,6%	94	84%	124	94%

*) Har efter analysens genomförande gått samman till Linnéuniversitetet

Lunds universitet har god täckning från 2002 och framåt och står för det största antalet publikationer. De flesta lärosätena har god täckning från mitten av 00-talet. Andelen publikationer som har abstrakt (sammanfattning) är hög, medan andelen som innefattar de obligatoriska SVEP-ämnesorden är förvånande låg. Lokala ID levereras av de allra flesta, men några lärosäten gör av tekniska skäl ännu inte detta. Andelen lokala dubletter, andelen tidskriftsartiklar med angivet ISSN och andelen böcker och kapitel med förlag kommenteras senare i rapporten som områden av vikt för bibliometriska analyser.

3.3. Bibliometriska analyser

För att få en bild av vilka olika bibliometriska analyser som görs på lokal nivå vid svenska lärosäten gjordes under januari 2010 en inventering genom ett utskick på en nationell bibliometri-e-postlista (Metrics). Svar erhöles från Chalmers, Göteborgs universitet, Karlstads universitet, Karolinska institutet, KTH, Linköpings universitet, Lunds universitet, Sveriges lantbruksuniversitet, Stockholms universitet, Umeå universitet och Uppsala universitet.

Svaren visade att det finns ett mycket stort antal olika bibliometriska mått som används av lärosätena (eller som skulle användas om man hade tillgång till rätt typ av data). De analyser som görs syftar i huvudsak till att få en överblick av det egna lärosätets forskning och att göra det möjligt att jämföra denna med forskning från andra lärosäten. I allt högre utsträckning görs olika analyser för att få en uppfattning om forskningens kvalitet. För att kunna analysera även utsnitt av lärosätets forskning är det viktigt att för varje publikation ha en tydlig koppling till den individ och institution som producerat artikeln samt att ha en uppfattning om publikationernas ämnesområde. Man är naturligtvis även intresserad av mängden publikationer och hur stor del av lärosätets publikationer som kan återfinnas i t.ex. Thomson Reuters citeringsdatabaser, då dessa uppgifter är ett inslag vid nationell resursfördelning. De flesta lärosäten använder uppgifter ur lokala publikationsdatabaser för flertalet av sina analyser, ensamma eller i kombination med data från Thomson Reuters citeringsdatabaser.

Databasposternas format är sällan något stort problem; mycket viktigare är att databasen är fullständig (att allt material matats in) och att de inmatade uppgifterna är fullständiga, korrekta och validerade. Flertalet av de lärosäten som kontaktats gör analyser som kräver citeringsdata, vilka vanligen hämtas ur Web of Science och i enstaka fall från Elseviers databas Scopus.¹¹

De lärosäten som besvarade vår förfrågan lämnade en mängd kommentarer som belyste situationen på lärosätena. Flertalet har problem med att den bibliografiska kvaliteten på egeninmatade poster är dålig, och att täckningsgraden är otillräcklig. Många trycker också på vikten av dubblettrensning för att analysresultaten ska bli tillförlitliga och en av de tillfrågade poängterar särskilt att detta bör ske även i ett centralt system eftersom samma publikation kan förväntas återkomma i flera av de lokala publikationsdatabaserna.

Ett förslag man ger för att komma till bukt med problemen är att kunna utbyta information med externa bibliografiska databaser (t.ex. Web of Science, Scopus, Libris och PubMed¹²) genom att noggrant ange unika ID:n. Några lärosäten nämner dessutom att forskarna själva inte ska tillåtas ändra i poster om dessa importerats från något externt system. Kopplingar till externa databaser skulle också kunna användas för berikning av posterna med ämnesord från etablerade databaser (t.ex. MeSH-termer¹³) vilket skulle kunna förbättra sökning och

¹¹ SCOPUS är en databas från förlaget Elsevier som innehåller vetenskaplig litteratur och webbresurser på abstrakt- och citeringsnivå. <http://www.scopus.com>

¹² PubMed/Medline är världens största databas inom biomedicin och produceras vid National Library of Medicine, USA

¹³ MeSH - Medical Subject Headings är en kontrollerad vokabulär för sökning, indexering och katalogisering som produceras vid National Library of Medicine, USA. Den är mycket spridd internationellt och översatt till flera språk, däribland svenska.

öppna för standardiserade analysfrågor på ämne. När det gäller ämnesindelningen anser ett par av de svarande att de nuvarande SVEP-kategorierna har låg precision och inte är användbara för att söka ut och jämföra publikationer mellan lärosäten. Det vore önskvärt med en ämnesstruktur som är mer konsekvent genomförd, särskilt om detta skulle göra det möjligt att göra sökningar utifrån ämneskategorierna i den modell som används av regeringen/Vetenskapsrådet.

Kopplingen mellan lärosätets författare och publikationer samt deras organisatoriska tillhörighet får en hel del utrymme i svaren och nämns vara av stor vikt. Egeninmatade uppgifter om författare, t.ex. namn och organisatorisk tillhörighet, är ibland bristande och någon nämner att man i efterhand får kontrollera det forskarna har angivit mot adressfälten i publikationen för att kunna lita på resultaten. Flera nämner att författare ibland hör till mer än en organisatorisk enhet inom lärosätet och att dubbeladressering måste kunna få förekomma, såväl för författare som för publikation. En av de tillfrågade föreslår också att man i SwePub bör etablera en standard för hur adressformatet ska registreras så att det blir lättare att göra jämförelser mellan fakulteter, institutioner och universitet/högskolor. Slutligen använder vissa lärosäten två olika sätt att koppla en publikation till en institution, dels genom var forskaren var när artikeln publicerades (vilket motsvarar adressfältet på publikationen), dels var forskaren är aktiv just nu (vilket kan erhållas ur personkatalogen).

Många lärosäten nämner behovet av att lokalt kunna återskapa de analyser som ligger till grund för de nationella resursfördelningsmodellerna i såväl Sverige som i Norge/Danmark. Man försöker återskapa Waringberäkningarna¹⁴ från regeringens modell och vill kunna genomföra analyser analoga med den "norska modellen", där en centralt fastställd svensk tidskrifts- och förlagslista skulle göra det lättare att göra jämförelser mellan universiteten/högskolorna.

Tabell 3. Vanliga typer av bibliometriska analyser vid svenska lärosäten

Antal publikationer	<p>Analysen som tar fram publikationer per person/institution/universitet m.fl. görs vid samtliga lärosäten. Ofta är resultatet uppdelat i olika publikationstyper enligt t.ex. refereegranskad artikel, bokkapitel, icke vetenskaplig artikel och liknande.</p> <p>Det görs även bedömningar av täckningsgraden för lärosätets publikationer i externa databaser. Resultatet kan t.ex. visas som andelen av enhetens refereegranskade artiklar som ingår i Web of Science/Scopus.</p>
Var forskningspublikationerna produceras och hur de sampublicerats	<p>Publikationernas koppling till individ, institution/avdelning och universitet utnyttjas i flera fall för att identifiera den enhet som ska analyseras och för att redovisa delresultat.</p> <p>I vissa fall tar man hänsyn till författar-/organisationsordningen på publikationerna.</p> <p>Kartläggningar görs av vilka inom lärosätet som sampublicerar, och av lärosätets sampublicering med andra organisationer. Exempelanalyser kan vara vilka forskare/lärosäten som sampublicerat och inom vilka ämnen, eller andelen</p>

¹⁴ "Waring" är en beteckning på en del av den metodik som används för att kompensera för olikheter i publiceringskulturer mellan olika vetenskapsområden i regeringens modell för fördelning av basmedel till de svenska lärosätena.

http://www.vr.se/download/18.72e6b52e1211cd0bba8800010145/bibliometrisk_indikator.pdf

	sampublikationer med andra institutioner/lärosäten/länder/företag.
Identifiera kvalitet	<p>Bibliometriska citeringsindikatorer är vanliga – oftast framtagna via Thomson Reuters citeringsdatabaser (på egen hand eller via tredje part) men i vissa fall används även Scopus-data.</p> <p>Exempel som ges är:</p> <ul style="list-style-type: none"> • C - antal citeringar • Cf - antalet citeringar normerat mot världsmedelvärdet för publikationer av samma typ, från samma år och i samma ämnesområde • Top n% - antalet citeringar jämfört med den n:te percentilen för publikationer av samma typ, från samma år och i samma ämnesområde • CPP/JCS - antalet citeringar normerat mot världsmedelvärdet för publikationer av samma typ, från samma år och inom samma tidskrift • JIF - Thomson Reuters Journal Impact Factor • JCf - Cf för en individuell tidskrift, en fältnormerad variant av JIF • H-index för enskilda personer (kräver citeringsdata för hela publikationstiden, dvs ofta tillbaka till 60-70-talet) <p>Denna typ av indikatorer används bland annat för att identifiera excellens, t.ex. lokalisera lämpliga kandidater inom lärosätet inför internationell utlysning av forskningsmedel.</p>
Ämneskartläggningar	<p>Exempel som ges är kartläggning av ett nytt forskningsområde (t.ex. vilka som är de mest citerade forskarna, mappning av sampubliceringar m.m.) eller av vilka (personer och institutioner) inom lärosätet som arbetar inom samma fält (inför bl.a. besök från andra lärosäten eller en eventuell omorganisation).</p> <p>Analyser görs också för att identifiera ämnesinriktning på ett urval publikationer, t.ex. alla publikationer från en institution, baserat på ord ur titel, sammanfattning och/eller kontrollerade ämnesord.</p>
Jämförelseanalyser	Många lärosäten studerar hur deras forskning förhåller sig till forskning från andra universitet/högskolor, såväl kvalitets- som ämnesmässigt. För dessa analyser tar man fram avancerade adressöksträngar och det efterfrågats kvalitetskontrollerade adresser för både svenska och för vissa utvalda utländska universitet.
Resursfördelningspecifika analyser	<p>Utöver ovanstående analyser är det relativt vanligt med analyser som direkt har sitt ursprung i de resursfördelningsmodeller som idag används i Norden:</p> <ul style="list-style-type: none"> • Analyser enligt 'norska modellen' med listor över publiceringskanaler, antingen med de norska/danska listorna eller egenutvecklade listor. • Kontroll av vilka publikationer som räknas med för lärosätet i regeringens resursfördelning. • I flera fall har man gjort försök att återskapa Waringfördelningen för det egna lärosätet.

3.4. Fördelningsmodeller

På liknande sätt inventerades vilka fördelningsmodeller som idag används vid svenska lärosäten. Sammanställningen har gjorts utifrån uppgifter från Blekinge tekniska högskola, Göteborgs universitet, Karlstads universitet, Karolinska institutet, KTH, Luleå tekniska

universitet, Lunds universitet, Malmö högskola, Sveriges lantbruksuniversitet, Umeå universitet, Uppsala Universitet, Växjö universitet och Örebro universitet.

Många lärosäten använder redan idag bibliometriska mått för fördelning av forskningsmedel, antingen på lärosätetsnivå eller på fakultetsnivå. Fördelningen sker oftast så att man fördelar en viss del av anslagen efter prestationer, men ibland delar man ut en fast summa pengar per viss typ av prestation. Som datakälla används antingen lokal publikationsdatabas eller uppgifter från Web of Science alternativt en kombination av båda.

De använda fördelningsmodellerna kan grovt indelas enligt följande:

- Norsk modell fullt ut (publikationsräkning med kvalitetsnivågruppering och kontroll av tillåtna källor mot officiell lista, mer information om den norska modellen finns i avsnitt 2.1 ovan)
- Förenklad norsk modell (utan kvalitetsnivågruppering och matchning mot källistor)
- Rent lokala modeller av publikationsräkning, med eller utan viktning
- Modeller som använder data från Web of Science (antal publikationer, antal citeringar, fältnormerade citeringsmått, impactfaktorer)
- Modeller som (utöver traditionella Web of Science-mått) använder Waringmått av samma typ som används i den nationella fördelningsindikatorn.

Flera lärosäten uttrycker ett önskemål att de lokala fördelningssystemen skall harmoniera med det nationella, samtidigt som man påpekar att detta är svårt att uppnå så länge det senare bara bygger på Web of Science-data. På vissa lärosäten har man på grund av de ämnesmässiga skillnaderna accepterat olika fördelningssystem inom olika fakulteter.

Analysen för resursfördelning skiljer sig inte till sin karaktär från andra bibliometriska analyser, men kraven på att publikationsdatabasen skall vara fullständig och korrekt blir högre om fördelningssystemet skall kunna behålla sin trovärdighet och legitimitet.

Beträffande *fullständigheten* så visar erfarenheten att täckningen ökar när kvantitetsrelaterade fördelningssystem införs; problemet blir då snarare att undvika överinmatning i form av dubletter eller av ännu inte formellt publicerat material (preprints, accepterade artiklar, provisoriskt e-publicerade artiklar).

Kraven på *bibliografisk nivå* finns även här, och eftersom man ofta kopplar ihop data från publikationsdatabasen med externa databaser (t.ex. Web of Science, Scopus, norska käll-listan) är det särskilt viktigt att pålitlig matchning mot externa databaser kan göras.

Kopplingen mellan adress, författare och publikation blir central – utan institutionsanknytning, inga pengar. I vissa fall blir också författarordningen ekonomiskt viktig. I de fall där man vill identifiera de som vid analystillfället är aktiva vid varje institution (med avsikt att fördela medel efter deras framtida potential, baserat på forskarnas tidigare resultat) blir det också viktigt att via personaldatabaser kunna koppla till författarnas aktuella adresser.

Ett speciellt problem vid medelsfördelning är den vikt som olika *kategoriseringar* kan få, t.ex. publikationstyp, granskningsnivå, nivå/målgrupp och geografisk spridning. Sådana kategoriseringar är delvis subjektiva och kan vara svåra både att fylla i och att kontrollera.

Flera lärosäten har pekat ut *konferensbidrag* som den besvärligaste publikationstypen i sammanhanget. Hur skall man kunna bedöma "hur publicerat" och "hur fackgranskat" bidraget är? Har det publicerats i en tryckt konferenshandling eller bara på nätet, eller finns bidraget bara på en CD eller ett USB-minne som delats ut till deltagarna? Om det uppges som "fackgranskat" – är det hela bidraget eller bara en sammanfattning som fackgranskats? Och så vidare.

Ämnesindelning verkar inte ingå som en parameter i något lokalt fördelningssystem, men kan möjligen komma att bli aktuellt för att stödja vissa former av (nationella) fördelningssystem.

4. Krav för kvalitetssäkring - tre nivåer

Utifrån inventeringsresultatet har vi inom arbetsgruppen formulerat tre kravnivåer för kvalitetssäkring av publikationsdatabaser. En nivå har som mål att ställa tillräckliga krav på publikationsdatabaserna för att de ska kunna användas som underlag för olika typer av bibliometriska analyser vid de svenska lärosätena. En andra nivå gäller rekommendationer om vad som krävs för att även göra databaserna lämpliga som underlag till medelsfördelning. Ett mål med denna uppdelning är att göra analysen och rapporten användbar för lärosätenas lokala kvalitetsarbete oberoende av den nationella diskussionen. Vi har också valt att i en tredje nivå lyfta ur krav och problemställningar som specifikt rör den nationella delen.

4.1. Krav för bibliometriska analyser

Redan idag görs ett antal mycket kvalificerade bibliometriska analyser vid de svenska lärosätena utifrån lärosätenas publikationsdatabaser. Baserat på inventeringen växer följande kravbild fram för publikationsdatabaser som ska fungera som underlag för sådana analyser.

Databasen ska innehålla

- lärosätets egen forskningspublicering med god täckning
- data som är uppställda på ett sådant sätt att det går att jämföra olika lärosätens publicering
- möjligheten att lätt avgöra vilka publikationer i externa databaser (t.ex. Web of Science) som motsvaras av databasens publikationsregistreringar
- klara kopplingar mellan publikationsregistreringarnas författare och specifika forskare och organisationer vid lärosätet
- uppgift om publikationens ämne/vetenskapsområde
- data om tidskrift (ISSN) och förlag med hög kvalitet
- bibliografiska uppgifter som i övrigt är fullständiga, korrekta och validerade.
- en publikationsmängd utan dubblettposter

4.2. Krav för medelsfördelning

I kraven på publikationsdatabasen i relation till medelsfördelning återkommer samtliga de krav som finns för allmänna bibliometriska analyser. För att publikationsdatabaserna även ska anses lämpliga som underlag för medelsfördelning bör man också kräva att slutanvändarna (forskarna) och lärosätenas ledningar ska kunna överblicka hur deras produktion visas upp och bedöms. Publikationsunderlaget bör explicit ha validerats av forskarna eller kanske lämpligast av institutionens prefekt.

4.3. Krav för överföring till nationell nivå

När data ska användas på ett effektivt sätt nationellt bör de också aggregeras och dedupliceras centralt. Den nationella åtkomsten av data kan både fungera som underlag för medelsfördelning nationellt och för jämförande analyser lärosätena emellan. Detta kräver ett nationellt överföringsformat som kan överföra de data som krävs för analyser och medelsfördelning enligt ovan samt att lärosätena kan leverera publiceringsdata enligt detta format.

5. Problemområden och rekommendationer

Arbetsgruppen har utifrån inventeringen och kraven i avsnitt 4 identifierat problemområden som gör informationen i dagens publikationsdatabaser svårtillgänglig för de analyser man önskar göra på nationell och/eller lokal nivå. För varje problemområde har möjliga åtgärder föreslagits.

5.1. Problemområden i relation till krav för bibliometriska analyser

5.1.1. God täckning av lärosätenas publicering

Ett tankesätt som delvis skiljer informations- och spridningsarbetet från analysarbetet är att kraven på rådata skiftar från önskemål om rika och informativa bibliografiska data (gärna med abstrakt och ämnesord) till önskemål om en komplett databas innehållande lärosätets samtliga publikationer med krav på obligatoriska fält och komplett ifyllda poster. Ett problemområde växer fram ur behovet att se till att så många som möjligt av lärosätets verkliga publikationer också återfinns i databasen. Bibliometrisk analys kräver god täckning av lärosätets publicering.

Lärosätena skiljer sig åt i fråga om vilken andel av lärosätets publikationer man anser sig ha i sin publikationsdatabas. Av de som svarat på enkäten om täckningsgrad och kvalitetskontroll bedömer 2 av 16 lärosäten att publikationsdatabasen täcker mindre än 50 % av lärosätets publikationer, medan 8 att den täcker mer än 90 %. Man anser i många fall att maximal täckning i databasen uppnåtts först från år 2007 (8 av 19).

Det framgår ur enkätmaterialen att det mest effektiva sättet att uppnå högre täckning i en lokal databas är att skapa lokal användning för dess innehåll. Detta kan till exempel göras genom användande av publikationslistor på personnivå i informationssyfte eller implementering av lokala fördelningsystem baserade på de lokala databaserna.

Samtidigt som man har god täckning, kräver man också att publikationer inte överregistrerats, vilket skapar dubletter i databasen. I de inventeringar som gruppen har gjort av de publikationsanalyser som görs vid svenska lärosäten framgår tydligt att ett av de problem många lärosäten har är förekomsten av dubletter inom den lokala databasen. Till denna problematik tillkommer det fenomen att lärosätena i flera fall kommer att mata in information om sampublicationer och därigenom skapa ett slags "legitima dubletter" vilka behöver hanteras nationellt. Dessa problem förekommer även i Norge och Danmark.

Därför rekommenderar arbetsgruppen:

- att lärosätena arbetar för god täckning i publikationsdatabasen
- att man nationellt väljer att rekommendera att alla eller en delmängd av de nationellt fastslagna dokumenttyperna speciellt kan komma i fråga för analyser. Samma dokumenttyper bör även ligga till grund för angivandet av när databasens täckning motsvarar analysbehovet.
- att de lokala databaserna öppet redovisar det år man anser databasen har tillräckligt god täckning för att kunna ligga till underlag för bibliometrisk analys.
- att överredovisning i form av lokala dubblettposter undviks på olika sätt (t.ex. genom att implementera dubblettkontroll vid inmatning av publikationer, skapa arbetsflöden för att ta bort dubletter eller erbjuda tillfälliga lokala dokumenttyper för bristfälliga publikationsposter såsom "preprint", som annars lätt leder till svårupptäckta dubletter)

5.1.2. Publikationsinformationens bibliografiska kvalitet

En avsevärd andel av posterna i universitetens publikationsdatabaser är manuellt registrerade av forskare, administratörer och bibliotekarier vid lärosätena. Den manuella registreringen ger möjlighet till att göra kompletta analyser på publikationer som inte förtecknas i externa databaser, men skapar också tillfällen till registreringsfel, stavfel, variationer av utformning av innehåll och varierande tolkningar av vilka uppgifter som bör finnas i respektive datafält, något som idag nämns som ett problem av flertalet lärosäten. Bristerna i datakvalitet och enhetlighet blir än mer påtagliga när posterna från de lokala databaserna förs samman till ett centralt system och speciellt bekymmersamma i samband med deduplicering och sammanslagning av poster som är registrerade vid olika lärosäten.

Förutom de skillnader i bibliografiska data som skapas av den manuella hanteringen på lärosätena skapar även de olika datasystemen vid lärosätena olika förutsättningar för bibliografisk kvalitet. Till exempel har inte vissa publikationshanteringssystem i dagens läge någon funktion för att skapa en koppling mellan författare och adresser, vilket är nödvändigt för att rättvist dela upp sampublicerade publikationer på de ingående lärosätena.

Vid många lärosäten utgör ordningsföljden på författarna till en publikation en viktig uppgift vid statistiska analyser och man kan inte utesluta att frågan för vissa ämnesområden behöver adresseras även nationellt. I SwePub MODS är direktivet att författarna ska anges i den ordning de står i publikationen, men det finns idag inget stöd i SwePub MODS för att bibehålla den ordningen när data läggs in i SwePub. Dessutom finns lokala system där författarordningen på samtliga författare som angivits på originalet inte bevaras och det egna lärosätets författares placering därför inte går att identifiera.

Av de lärosäten som svarat på vår enkätfråga om kvalitetskontroll svarade 16 att de har någon form av kvalitetskontroll av posterna medan 5 saknar kontroll eller kontrollerar enbart vissa kategorier som avhandlingar eller poster med uppladdade filer.

Därför rekommenderar arbetsgruppen:

- att externa data utnyttjas så långt det är möjligt för att undvika felinmatning lokalt. Detta kan till exempel göras genom att forskarna under inläggningen erbjuder färdiginmatade data för komplettering.
- att varje post granskas innan den publiceras eller görs allmänt tillgänglig i det publika systemet. En nationell diskussion om vad som bör ingå i granskningen välkomnas, men bibliografisk kvalitet, deduplicering (borttagning av dubletter) och för vissa dokumenttyper verifikation mot den verkliga publikationen bör ingå.
- att det tydligt framgår vid inmatning huruvida en post innehåller tillräcklig information för att kunna vara underlag för analys respektive medelsfördelning.

5.2. Problemområden i relation till krav för medelsfördelning

I samband med att data från lärosätenas publikationsdatabaser används som underlag för medelsfördelning (lokalt på lärosätena eller centralt) ökar kraven på kontroll av att uppgifterna är korrekta. Dessutom skiljer sig de krav som ställs på en databas för informations spridning, som var det ursprungliga syftet för många av publikationsdatabaserna, till viss del från de krav som ställs på en databas som ska användas för medelsfördelning. Många forskare vill att man bara analyserar deras "bästa" publikationer som underlag för medelsfördelning, men det ligger i lärosätenas intresse att visa upp så många av sina publikationer som möjligt.

Om databasen ska användas för medelsfördelning, så är det viktigt att slutanvändarna (forskarna) känner att de kan överblicka hur deras produktion visas upp och bedöms. Det gör det också enklare för dem att lägga in material i databasen på ett sätt så att det uppnår den önskade "användbarhetsgraden", t.ex. för informations spridning respektive för analys. Dessutom finns ett intresse från såväl analytikerhåll (ur förankringssynpunkt) som ledningshåll att ledningen för lärosätet informeras om hur lärosätet analyseras.

Därför rekommenderar arbetsgruppen:

- att arbetsflöden byggs upp lokalt där varje forskare årligen godkänner sina publikationer innan rapportering sker till medelsfördelning.
- att arbetsflöden byggs upp lokalt där varje prefekt eller motsvarande årligen godkänner sin institutions publikationer innan rapportering sker till medelsfördelning.
- att det framgår publikt i databasen huruvida en publikation godkänts på respektive nivå genom att SwePub MODS ger utrymme för detta.

5.3. Problemområden i relation till krav för överföring till nationell nivå

Redan under utvecklingsarbetet med SwePub visade det sig att det fanns ett stort önskemål från olika intressenter att ha någon form av nationell samlingsdatabas för

forskningspublikationer, initialt för informationssökning. Det har även visat sig i samband med arbetsgruppens inventeringar att det vid lärosätena finns starka önskemål om en nationell databas lämpad för bibliometriska analyser. Om nationell medelsfördelning baserat på publikationsdatabaserna kommer i fråga, så kommer detta senare önskemål att växa. SwePubs lämplighet för analysverksamhet har med anledning av detta diskuterats inom arbetsgruppen och en del problem har lyfts fram.

Om egeninmatade data ska användas nationellt, i SwePub eller annorstädes, så behöver alltid någon form av överföring ske av data från en lokal domän till en nationell samlingsdatabas. Detta steg finns även i nationellt administrerade lokala system, som i Norge. För att en sådan dataöverföring ska vara möjlig krävs ett nationellt överföringsformat. Genom att fokusera på just överföringsformatet så låser man inte fast sig vid en lösning som utgår från den nuvarande SwePub-databasen. Så länge det nationella överföringsformatet lämpar sig för både Libris aktuella databehov, för SwePub och för ett eventuellt separat analyssystem, så kan det stödja båda systemen om diskussionen av någon anledning i framtiden leder fram till krav på en logisk delning. Detta avsnitt inriktar sig på att ta upp problemområden i det nationella överföringsformatet i relation till de analyskrav som finns på nationell nivå.

5.3.1. Implementering av dagens nationella överföringsformat

Av de SUHF-anslutna lärosätena levererar idag 28 lärosäten data till SwePub-systemet. De som just nu saknas är de konstnärliga högskolorna i Stockholm samt Ersta Sköndal högskola, Handelshögskolan i Stockholm, Karolinska institutet, Röda Korsets högskola, Sophiahemmet högskola och Sveriges lantbruksuniversitet. Som det framgår i tabell 1 planerar flera av dessa lärosäten att implementera gränssnitt för överföring till SwePub MODS redan under året. Av SUHF:s styrelsebeslut 2009-06-17 framgår det att SUHF rekommenderar lärosätena att sträva i denna riktning. Detta anser vi som en nödvändig utveckling i arbetet mot ett underlag för nationell medelsfördelning.

Vi ser också att det även återstår en del att göra för de lärosäten som redan nu levererar till SwePub för att deras data helt och hållet kan svara upp mot de krav som formatet ställer. En del obligatoriska fält saknas fortfarande, vilket innebär att alla publikationer inte förs över till SwePub eller att man tvingats släppa på vissa krav i formatet. Det bör klart framgå att överföringsformatet är det som definierar vilka publikationer som kommer i fråga för medelsfördelning.

Därför rekommenderar arbetsgruppen:

- att SUHF fortsätter att peka ut de lokala databaserna och möjligheten att hämta ut posterna därifrån enligt ett nationellt överföringsformat som SUHF:s förslag till datakälla för fördelning av medel. Detta för att skapa incitament för de lärosäten som inte är anslutna till en publikationsdatabas eller kan leverera enligt SwePub MODS att skapa möjligheter till detta.
- att SUHF anammar dagens nationella överföringsformat (SwePub MODS) och dess obligatoriska fält som en grund för det fortsatta arbetet, och att publikationer med

ofullständiga data i fortsättningen inte kan räkna med att komma i fråga för nationell medelsfördelning.

5.3.2. Utveckling av nuvarande nationella format

I teorin räcker SwePubs MODS-specifikationen långt också för statistiska analysbehov, även om den från början främst utformades för informationsspridning. Trots detta har arbetsgruppen identifierat ett mindre antal förslag till förändringar i specifikationen, vilka skulle göra det möjligt att kunna använda databaserna även för nationella analyser och medelsfördelning.

En av de mest avgörande skillnaderna mellan hur SwePub MODS-specifikationen är utformad idag och hur den bör se ut för att även motsvara kraven för analyser är att statistiska analyser ställer strängare krav på koppling mellan upphovsmän och organisationer, alltså författarnas organisatoriska affilieringar. SwePub MODS anger att "adressangivelser för det egna lärosätets författare *bör* anges". I ett system för statistiska analyser behöver detta göras om till ett obligatoriskt krav och inmatningen av adressuppgifter behöver också vara kontrollerad, så att adressuppgifterna automatiskt kan kopplas till lärosätena.

Ytterligare en brist i dagens SwePub-MODS-format är att det tillåter inmatning av författarnas namn i en relativt fri form och dessutom inte har någon specifikation för hantering av dubbel- och mellannamn. Bristen på enhetlig utformning av författarnamn blir ett avsevärt problem när dubblettposter från olika lärosäten ska föras samman med bibehållna uppgifter om författare och deras respektive lärosäte.

Ett annat mer praktiskt problem med de data som finns i SwePub-systemet är att de härrör från olika publikationsdatabaser vid olika lärosäten, vilket innebär att både utformningen av data och klassifikationen av datainnehållet skiljer sig åt. En användning av universitetens publikationsdata för centrala statistiska analyser ställer i vissa delar högre krav på enhetlighet hos datainnehållet än användning för lokala analyser och informationsspridning eftersom poster från olika publikationsdatabaser måste kunna jämföras med varandra. Dokumenttyper och innehållstyper enligt SVEP och SwePub MODS bör användas enhetligt över riket.

Flera av de problem som finns i de lokala publikationsdatabaserna återuppstår (om än i något förändrad form) när dessa ska slås samman till en gemensam databas. Adressuppgifter behöver utformas enhetligt så att adresserna kan hänföras entydigt till lärosäten, författarnamn behöver vara utformade så att uppgifterna kan slås samman vid deduplicering av publikationer (eftersom det ännu inte finns något nationellt författar-ID som posterna kan föras med lokalt), författarordningen bör säkras på ett betryggande sätt, ISSN-nummer behöver vara i sådant format att de automatiskt kan matchas mot tidskriftslistor, och uppgifter om förlag behöver kunna matchas mot förlagslistor. Publikationsposterna behöver också i största möjliga mån innehålla identifierare av olika slag (t.ex. Libris-ID, Thomson ID) som underlättar deduplicering inom SwePub och mot motsvarande publikationsposter från andra databaser.

Det är känt att de kvalitetsmått som används varierar kraftigt mellan olika ämnesområden och att man i analyser ofta behöver ta hänsyn till publikationernas respektive forskningsfält. I de fall man har en publikationsdatabas med information från flera källor saknas en gemensam ämnesindelning att normera mot, vilket flera lärosäten upplevt som ett problem. I synnerhet eftersom man tillämpat olika ämnesindelning vid olika lärosäten vilket avsevärt försvårar jämförelseanalyser. Vi föreslår därför att man även i fortsättningen har ämnesindelning enligt nationella definitioner som ett obligatoriskt fält i det nationella överföringsformatet.

Därför rekommenderar arbetsgruppen:

- att man uppdaterar SwePub-MODS-specifikationen enligt följande:
 - Angivelse av organisationstillhörighet (företrädevis enligt lokala auktoritetsregister) görs obligatoriskt på personnivå för de personer som är knutna till det registrerande lärosätet.
 - Externa id (referenser till motsvarande publikation i externa system genom ISBN, DOI, ISSN, URI, Thomson ID, Scopus ID, PubMed ID) används så långt det är möjligt och görs till ett steg i granskningen.
 - Identifieringen av författare underlättas genom att personliga ID görs obligatoriska för universitetets författare och att ett enhetligt system införs för hur namn skrivs och delas upp i MODS-elementen. Detta gäller främst dubbelnamn i för- eller efternamn.
 - Författarordningen och andra uppräknings anges explicit i XML-formatet
 - Det blir obligatoriskt att lokala identifikationssträngar görs tillgängliga nationellt för att posten ska anses lämplig för medelsfördelning.
 - Granskningsgraden (forskare/prefekt) kommuniceras i MODS-formatet.
- att man på nationell nivå skapar ett register över tidskrifter och förlag, som fritt kan användas som hjälpdatabas för att skapa pålitliga ID för publikationskanalerna lokalt och nationellt.
- att SVEP/SCB:s ämnesord används obligatoriskt och i sinom tid ersätts av SCB:s nya EU-anpassade klassifikationsschema, när detta blir tillgängligt.
- att de lokala systemen anpassas så att de motsvarar de nationella kraven, så att överföringen mellan forskaren och den nationella nivån inte hindras på något sätt av lokala arbetsflöden eller system.
- att man initierar en användargrupp för de som levererar enligt det nationella överföringsformatet där man kan diskutera och besluta om t.ex. användningen av publikationstyper eller hur man kan integrera de nationella listorna över tidskrifter och förlag i de egna systemen.
- att man ser över behovet av en certifiering av publikationsdatabaser, där man både godkänner och stödjer lärosätenas fortsatta utveckling av publikationsdatabaser.

5.4. Specifikt lokala analysbehov

Det finns som nämnts tidigare, mycket gemensamt mellan de lokala analysbehoven och de nationella. Flertalet rekommendationer för det nationella systemet ovan bör således

tillämpas även i lokala publikationsdatabaser, i synnerhet om man har för avsikt att använda dem som underlag för lokal resursfördelning. Dock berörs en lokal databas inte av rekommendationen gällande sammanslagning av poster från olika lärosäten (avsnitt 5.3).

För ett lokalt lärosäte är det å andra sidan viktigt att veta inte bara att publikationen tillhör det egna lärosätet, utan vem/vilka av publikationens författare som tillhör lärosätet och inom vilken fakultet/institution han/hon var aktiv när publikationen kom till. Denna information måste således också ingå i en lokal publikationsdatabas. Många lärosäten har upplevt problem med multipla namnformer på såväl personer som organisatoriska enheter och har, för att bemöta detta, inrättat kopplingar mot personalkataloger och andra auktoritetsregister som man kunnat använda för att applicera unika ID-nummer på varje analyserbar enhet (t.ex. författare, institution, fakultet).

Därför rekommenderar arbetsgruppen:

- att alla lokal data ska innehålla anknytning till person och till varje persons organisation, där publikationen producerades. Detta ska göras genom anknytning till lokala auktoriteter/ID:n från t.ex. den lokala personaldatabasen eller liknande.

6. Kommentarer från teknisk remiss

Som ett sista steg i projektets arbete sändes rapporten ut för teknisk remiss till samtliga bibliotekschefer vid SUHF-lärosäten, särskilt adresserat till personer som arbetar med publikationsdatabaser och bibliometri. Målet var att fånga upp tekniska problem i underlaget innan rapporten överlämnades till uppdragsgivaren. Svar kom in från tolv lärosäten med nyttiga förslag.

På det hela var remissinstanserna mycket positiva och man välkomnade arbetsgruppens uppdrag och förslag. Flera nämnde att de redan genomfört många av rekommendationerna på eget initiativ och att de resterande var välfunna. Många kommentarer berörde mer tekniskt detaljerade frågor kring kvalitetssäkring och implementering i systemen, vilket tydligt visar på behovet av ökad nationell samordning i dessa frågor. Dessa var de punkter som flera tog upp:

- En viktig del i SUHF:s kommande arbete bör vara att tillsätta en teknisk samordningsgrupp som mer i detalj kan arbeta vidare med de rekommendationer och riktlinjer som föreslås i rapporten. Några viktiga frågor för en sådan grupp bör vara riktlinjer för organisatorisk tillhörighet, nationell samordning av auktoritetsregister för tidskrifter och förlag samt i framtiden även för personer och organisationer. Andra viktiga frågor är samordning och riktlinjer för dokumenttyper, ämnesord och bibliografiskt granskningsarbete.
- Arbetet med att genomföra rekommendationerna kan bli resurskrävande. Detta gäller främst för lärosäten som inte har lokala databaser just nu eller för vilka det krävs omfattande ändringar. Man pekar också på att det krävs extra personella insatser för de lärosäten som inte tidigare infört arbetsflöden för bibliografisk granskning. Det påpekades också att det finns möjliga konflikter mellan de lokala tjänsternas olika uppdrag och behov och de högre krav som en central tjänst, användbar för analys och förmedling, kommer att ställa.

Vi är som grupp tacksam för kommentarerna och har uppdaterat dokumentet därefter.

7. Slutsatser

Arbetsgruppen konstaterar att de lokala publikationsdatabaserna, genom sin bredare täckning av framför allt de humanistiska och samhällsvetenskapliga forskningsområdena och sitt större spektrum av dokumenttyper, kan ge analys- och medelsfördelningsmöjligheter utöver vad de idag använda internationella databaserna ger. För att förverkliga detta krävs både ett lokalt kvalitetsarbete med publikationsdatabaserna och, för nationell användning, en ökad standardisering och samordning av data. Vi har därför valt att skilja de rekommendationer som rör de lokala systemen från de som rör de nationella behoven.

Lokalt ser vi behov av att både arbeta med databasernas täckning (se till att alla publikationer registreras) och med posternas bibliografiska kvalitet (se till att registreringarna blir korrekta). I arbetet med täckningen trycker vi på behovet av lokala användningsområden för att öka rapporteringstalen. Vi välkomnar även en specificering av vilka dokumenttyper som ska prioriteras för framtida nationella fördelningar, att databaserna redovisar från vilket år de är pålitliga för analyser och att lokala dubletter och andra typer av överrapportering undviks.

I arbetet med den bibliografiska kvaliteten i databaserna framhåller vi två viktiga behov. Externa data (data från externa databaser) som man fritt får lägga in i lokala databaser bör utnyttjas så långt som möjligt. Detta gör att felinmatningar i egenrapportering undviks. Det kan dessutom ge möjlighet till ytterligare täckning och avlastar forskarna. Vidare bör varje post granskas innan den publiceras för att verifiera den bibliografiska kvaliteten, undvika dubblettinmatning och möjliggöra kontroll mot publikationen. Vi tycker också att det är viktigt att det framgår av inmatningen av informationen om publikationen av någon anledning inte kan komma i fråga för nationell fördelning.

Om publikationsdata ska användas för medelsfördelning, så anser arbetsgruppen att man dessutom ska införa ett moment av verifikation, eftersom publikationsdata kan läggas in i databasen på många olika sätt och av olika aktörer. Detta bör göras genom att både forskare och prefekter årligen verifierar publikationslistor och intygar att innehållet är korrekt för deras person respektive institution.

När data förs samman på nationell nivå uppstår dels behov av ett nationellt överföringsformat, dels behov att lösa specifika problem som bara framkommer då data från olika källor förs samman. Oberoende av hur publikationsregistrering görs i framtiden, så finns det behov av ett gemensamt format. Idag utgörs detta av SwePub MODS, som vi också anser vara ett lämpligt format att utgå från. För att göra detta anser arbetsgruppen att samtliga lärosäten ska ha tillgång till en publikationsdatabas som kan leverera enligt SwePub MODS-specifikationen, vilket också bör vara ett krav för att en publikation kan komma i fråga för analyser och medelsfördelning. Vi föreslår därför också ett antal förbättringar som behöver tillföras i formatet för att göra det användbart för pålitliga analyser. Vi föreslår att nationella hjälpdatabaser skapas för att ge möjlighet till fullständig inmatning av uppgift om ISSN och förlag, vilka är viktiga för analysmetoder som baseras på publikationskanal.

Sammanfattningsvis föreslår arbetsgruppen att SUHF:

- arbetar för att de lokala databaserna skall utgöra underlag för analys och medelsfördelning på nationell nivå,
- rekommenderar sina medlemmar att anpassa sina publikationstjänster enligt denna rapports rekommendationer,
- initierar en vidareutveckling av formatspecifikationen för SwePub MODS på nationell nivå genom att en nationell samordningsgrupp utses.