



Per Egevad, Sara Kjellberg, Helena Rydberg
Malmö högskola, Bibliotek & IT
per.egevad@mah.se
2014

Analys av forskares publiceringsmönster för utveckling av bibliotekets forskarstöd

Projektet om forskares publiceringsmönster startade på Malmö högskolas bibliotek under våren 2014 och har sin grund i tankar som uppstått i teamen för periodika och publicering. Att utvärdera och uppdatera det periodiska beståndet är en viktig del i arbetet med periodika och vi ville göra en undersökning om huruvida det råder en samstämmighet mellan Malmö högskolas periodiska bestånd och de tidskrifter som faktiskt används av våra forskare. Utvecklingen av samlingarna bör spegla forskarnas praktiker, och genom detta skapa bättre förutsättningar för att stödja dem i deras arbete (Aquilonius *et al.* 2008; Voog *et al.* 2013). Ett sätt att undersöka användningen av tidskrifter är att titta på vilka tidskrifter som forskarna publicerar sina forskningsresultat i. Detta borde ge en fingervisning om vilka tidskrifter som de anser vara betydelsefulla och som biblioteket i sin tur bör ha tillgång till.

För några år sedan gjordes en liknande studie på Odontologiska biblioteket vid Malmö högskola: *Vad kan biblioteket utvinna ur en sökning på den egna publiceringen med hjälp av några bibliometriska metoder? - Ett exempel från Malmö högskola Odontologiska fakultetens publiceringar mellan 2000-2008* (Egevad 2010). I studien visades samband mellan både de tidskrifter forskare publicerade sig i och refererade till och bibliotekets tidskriftsbestånd. Studien visade även på problem med att skala upp analys utifrån tidskrifter forskare refererade till eftersom tillgång och bearbetning av referensuppgifterna måste till stor del göras manuellt. I den här studien undersöks därför om det går att genomföra en liknande studie i större skala endast utifrån tidskrifter som forskare publicerar sig i och vilka resultat detta kan ge.

Syfte

Syftet med projektet är således att undersöka hur väl det periodiska beståndet överensstämmer med forskarnas publicering, samt hur detta kan bidra till utveckling av bibliotekets forskarstöd. Har Malmö högskolas bibliotek tillgång till de tidskrifter där högskolans forskare publicerar sig?

För att kunna undersöka detta har vi utarbetat en metod som kan användas för att ta fram de tidskrifter som forskarna publicerar sig i och jämfört dessa uppgifter med bibliotekets bestånd. För att testa metoden vill vi genomföra en pilotundersökning och analys av publiceringsmönster hos forskare vid fakulteten för Hälsa och Samhälle. Vi menar att denna grupp är lagom stor men ändå tillräckligt varierande för att undersöka.

Metod

I Egevads (2010) studie baserad på forskning i odontologi på Malmö högskola användes en uppsättning bibliometriska metoder för att undersöka publiceringens matchning mot högskolans bestånd utifrån både publicerad tidskrift och refererad tidskrift. Studien baserades därför på såväl artikelns metadata som på dess referenser. Behandlingen av referenslistor är komplicerad och arbetsintensiv, därför valde vi i denna studie att enbart fokusera på artikelmetadata. För att göra en datainsamling och skapa en metod som går att skala upp för en större del av Malmö högskolas forskning behövdes två olika slags metadata från två olika källor. Dels behövdes en insamling av bibliografiska metadata för publikationer skrivna av Malmö högskolas forskare vid fakulteten Hälsa och samhälle, dels en insamling av data om publikationer som biblioteket ger tillgång till via sina system. Urvalet är baserat på vetenskapliga artiklar där forskaren själv angivit att artikeln genomgått peer review.

En viktig utgångspunkt är att skapa en metod som inte kräver specialprogram och särskilda kunskaper utöver grundläggande statistik och Excel. Detta för att analyserna ska kunna återupprepas inom olika ämnesområden och att metoden ska vara enkel att användas av alla inblandade i periodikaarbetet.

Metadata med bibliografisk information om Malmö högskolas forskares publikationer kan hämtas såväl i Swepub som i MUEP (Malmö högskolas institutionella arkiv). Ett första steg var att göra utsökningar av metadata ur dessa system för att se vilket som skulle vara enklast och ge bäst underlag för fortsatt analys. Det visade sig att datan från Swepub krävde mer bearbetning, samt att den inte var lika uppdaterad som MUEP. Dessutom innehåller Swepub inte lika mycket information som MUEP. Swepub-exporten är gjord för export av referenser till böcker i mindre mängder och begränsad till uttag av max ett hundra poster per gång. Detta gör att det är begränsad information som finns lagrad i den exporterade posten samt att uttaget blir mer arbetskrävande.

MUEP är baserat på DSpace vilket är ett system som är utvecklat som mjukvara för öppna institutionella arkiv. Uttag av metadata kan göras genom en tilläggstjänst kallad Extramuep som finns tillgänglig för alla. Där finns möjlighet att få uttag baserade på OAI-PMH XML som till exempel används för att föra in metadatan i Swepub. En sådan fil är svårbehandlad i Excel om man inte har ett sätt att konvertera den. Ett alternativ är att använda sig av metadatan i form av en kommaseparerad fil för att kunna bearbeta den i Excel. Om man vill göra den typen av uttag ur MUEP behöver man dock vara administratör i systemet. En grundidé var att försöka basera uttagen på öppet tillgänglig data. En eftergift fick göras och istället använde vi en kommaseparerad fil som skapas regelbundet för användning i

högskolans administrativa system. Genom att använda den fil som innehåller metadata om alla publikationer inmatade i MUEP kan filtrering istället göras i Excel.

Utöver publikationsmetadata behövs uppgifter om bibliotekets bestånd. Till bibliotekets bestånd räknar vi bibliotekskatalogens tryckta tidskriftsbestånd samt det elektroniska bestånd som biblioteket tillgängliggör via länkserver (SFX). I detta bestånd inkluderas även de open accessresurser som biblioteket valt att aktivera i länkservern. All beståndsdata, såväl tryckt som elektronisk, finns samlat i länkservern.

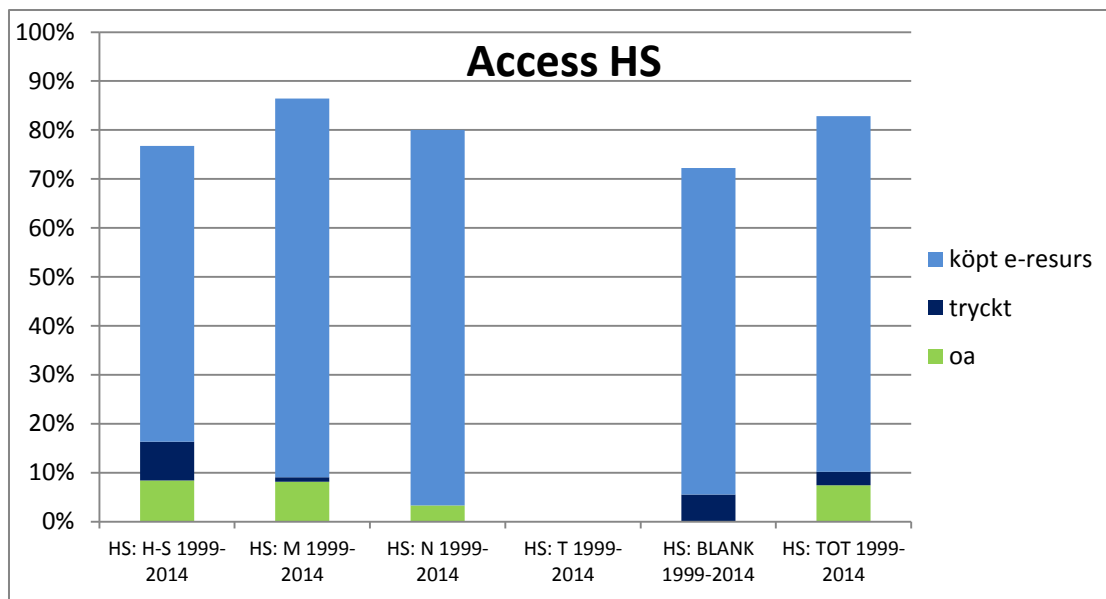
Datinsamlingen har alltså bestått i att hämta metadata från MUEP samt beståndsuppgifter från vår länkserver och dessa har sedan sammanfogats i en analysfil. Det är genom att skapa en analysfil där två olika dataset sammanfogas som en jämförelse mellan de tidskrifter forskarna publicerar sig i och vad vi ger tillgång till kan göras.

Analysfilen består av metadata om artikeln och åtkomstuppgifter om vilka tidskrifter och årgångar biblioteket ger access till. Åtkomstuppgifterna ger svar på följande frågor: 1) är tillgänglig 2) är open access 3) är tryckt 4) om ej 2 eller 3 räknas det som e-access. Vid analysen filtrerar man först fram de artiklar som skall analyseras. Filtringen har då skett utifrån dessa utvalda faktorer som finns i vårt institutionella arkiv: 1) Artikeltyp Peer-review Internationellt eller i Sverige 2) Vetenskapsområde¹ (H-S, M, N, T samt blankt) 3) Område, avdelning inom Malmö högskola. Matchningen mellan artiklar och accessdata sker via ISSN eller tidskriftsnamn. Resultatet redovisas som totalt antal artiklar, antal artiklar som vi ger tillgång till, antalet som är open access eller via tryckta tidskrifter. Utifrån detta beräknas den procentuella fördelningen på artikelnivå eller tidskriftsnivå.

Resultat

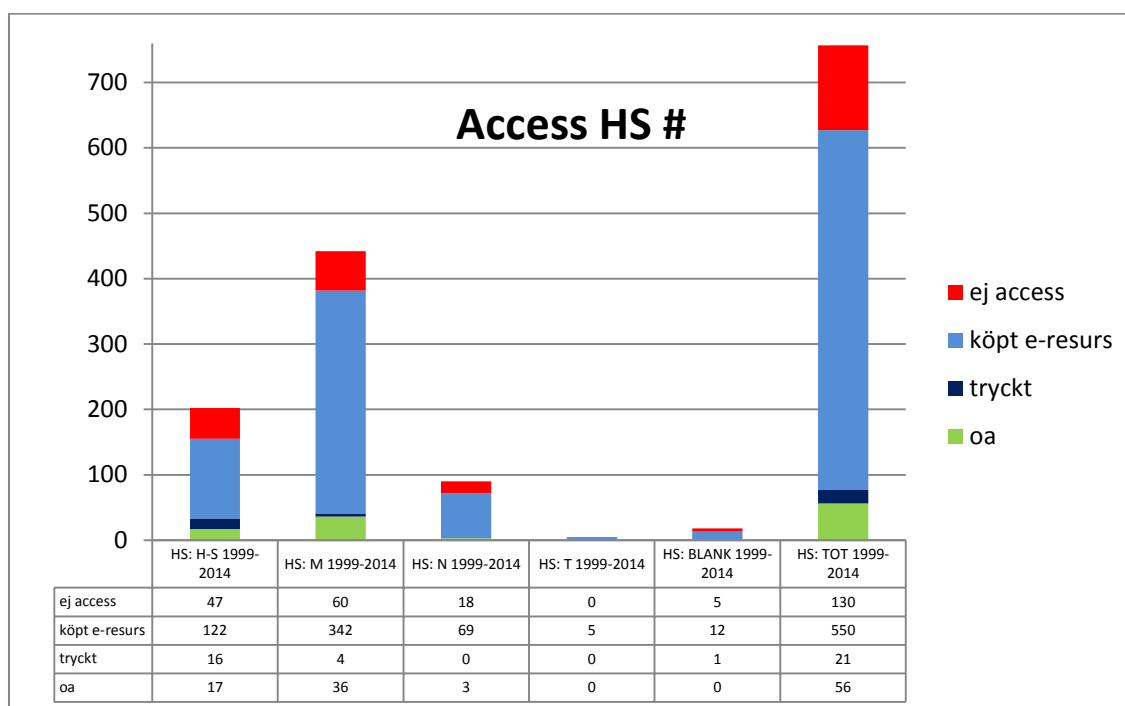
Genom att avgränsa datamängden till enbart publikationer från Hälsa och Samhälle kan vi jämföra hur väl det periodiska beståndet stämmer överens med forskarnas publicering inom denna fakultet. En första beräkning baserad på deras publicering under åren 1999-2014 visar att Malmö högskolas bibliotek har tillgång till en stor del av de tidskrifter som forskarna publicerar sig i. Beräkningarna är gjorda på 757 publiceringar. Forskarna registrerar publikationerna själv i MUEP, och enbart det som de publicerat under sin tid som Malmö högskolaanställda. Sedan 2007 är det obligatoriskt att registrera sina publikationer i MUEP. Detta får betydelse för vilka publikationer vi kan basera analysen på. I arbetet med analysfilen har felaktigheter i katalogposter samt i MUEP upptäckts och dessa har justerats manuellt. Det har också visat sig att Malmö högskolas egna open accessresurser (som vi har tillgång till) inte har varit inlagda i katalogen, och därför inte syns vid uttaget från länkserver. Detta har också justerats manuellt.

¹ Ämnesområdena Humanities/social sciences, medicine, sciences (naturvetenskap) och technology väljer forskaren själv då hen lägger upp sin artikel i MUEP. Blank är när man inte lagt in ngt vetenskapsområde, vilket skedde i det tidiga materialet



Tabell 1 Tillgång till artiklar som forskare från fakulteten Hälsa och Samhälle publicerat, redovisat i procentandel uppdelat enligt ämnesområden H-S, M, N, T, Blank och total.

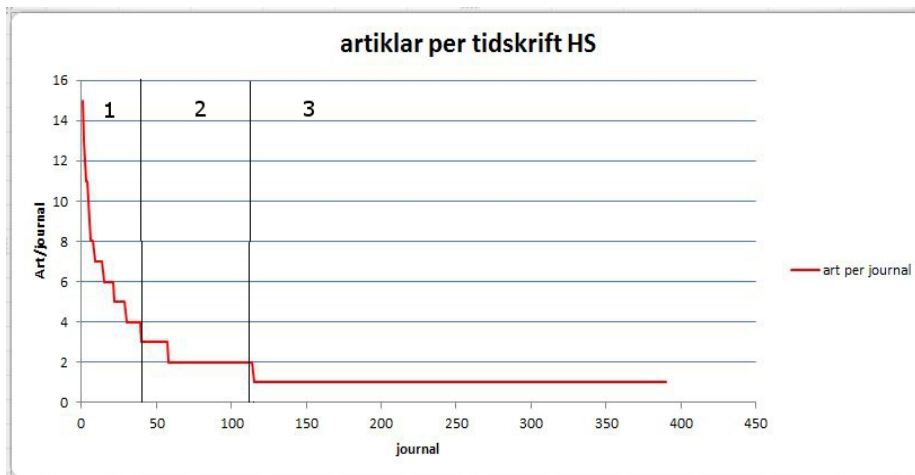
Mer än 80 % av artiklarna finns tillgängliga för Malmö högskolaforskarna antingen som inköpt e-resurs, tryckt tidskrift eller open access². När man tittar på det totala antalet artiklar och andelen som ej ges tillgång till är det 130 stycken som inte finns tillgängliga.



Tabell 2 Tillgång till artiklar som forskare från fakulteten Hälsa och Samhälle publicerat, redovisat i antal artiklar uppdelat enligt ämnesområden H-S, M, N, T, Blank och total.

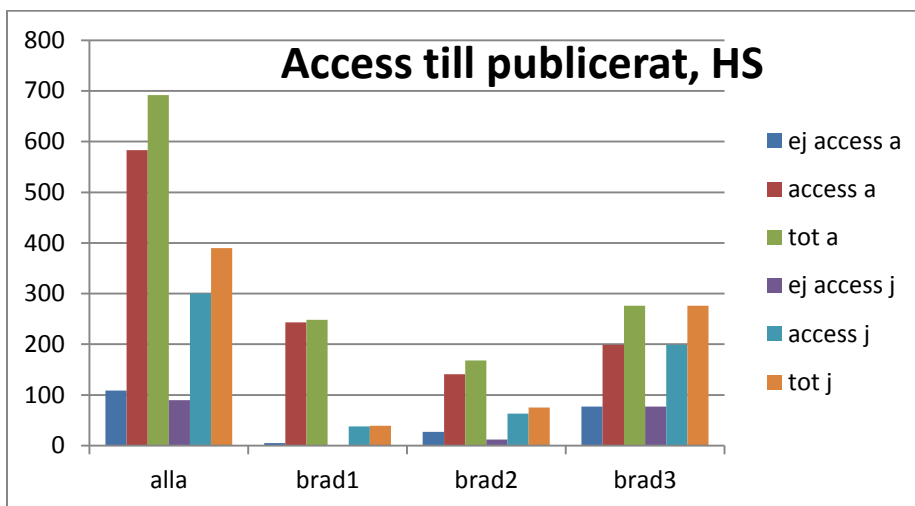
² Andelen open access är inte samma slags beräkning som om man skulle titta på publikationer som är parallellpublicerade i MUEP. Open access här är fråga om vilka källor som är aktiverade i länkservern såsom fritt tillgängliga. Det betyder att vid en sökning via bibliotekets resurser kan man nå publikationen i fråga.

För att analysera tidskriftstillgången har vi vidare använt en förenklad form av analys utifrån Bradfords lag³ vilket innebär att vi har delat upp artiklarna i tre lika delar där den första zonen visar mest publiceringar/tidskrift, den andra medel och den tredje minst. Kurvan nedan visar en förväntad fördelning enligt en sådan beräkning. Ett mindre antal tidskrifter är de mest centrala hos fakultetens forskare och har därför ett större antal publiceringar, medan kurvan sedan planar ut och visar hur forskare publicerat enbart en artikel/tidskrift.



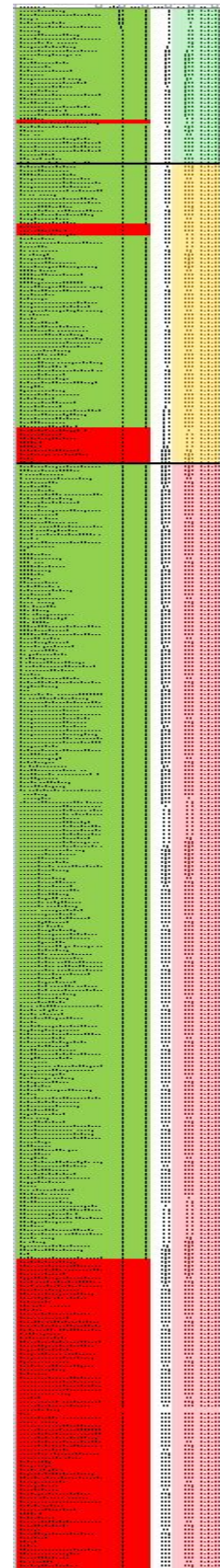
Tabell 3 Antal artiklar per tidskrift publicerade av forskare från Hälsa och Samhälle. De tre olika Bradfordzonerna innehåller ungefär lika många artiklar.

Man kan sedan applicera tillgången till tidskrifter i Bradford-zonerna med tillgången till artiklar i Bradford-zonerna för att se fördelningen i de olika zonerna samt kontrollera om vi har tillgång till de viktigaste tidskrifterna.



Tabell 5 Tillgång enligt indelning i Bradford-zoner för Hälsa och Samhälles publikationer såväl på artikelnivå som tidskriftsnivå.

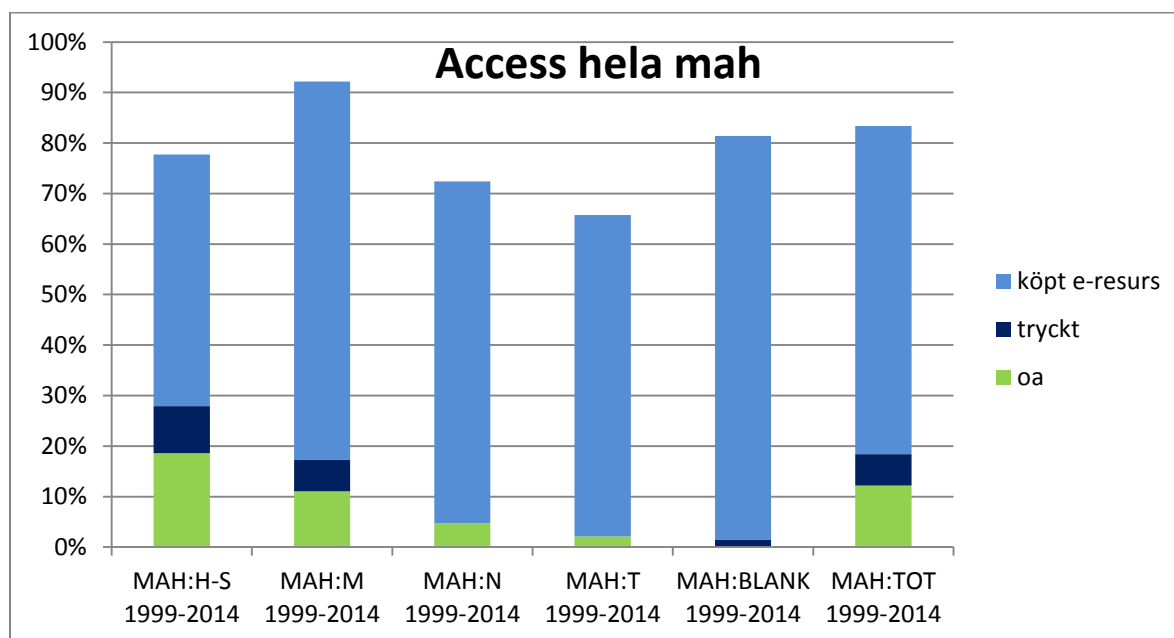
³ Bradfords lag är en förutsägelse hur artiklarna inom ett ämnesområde fördelar sig över tidskrifterna i ämnesområdet. Bradford visar hur artiklarna samlar ihop sig och om man delar upp ett ämnesområde i flera lika delar, artikelmässigt, så ökar antalet artiklar mellan zonerna enligt formeln $1:n:n^2$. (Kärki *et al.* 1998, s. 17)



Tabell 4 Alla tidskrifter ordnade i storleksordning utifrån antalet publicerade artiklar. Grönt=access. Rött=ej access. Bradfordzoner markerade till höger. Zon 1 överst.

När man gör denna indelning blir det tydligt att tillgången till de viktigaste tidskrifterna i Bradford-zon 1 i princip är helt täckt. De större luckorna är i Bradford-zon 3, alltså bland de tidskrifter där forskare enbart publicerat en artikel per tidskrift.

De kombinerade uttagen, från Malmö högskolas institutionella arkiv och länkservrar samlade i en Excel-fil med formler, kan användas som ett verktyg för att göra beräkningar på varje fakultets publicering. Exempelvis kan man ta fram i vilken form tillgången till artiklarna sker, köpt e-resurs, tryckt tidskrift alternativt via open access.



Tabell 6 Total tillgång för hela MAH baserat på publikationer registrerade i MUEP uppdelat på köpt e-resurs, tryckt och open access.

Drygt 10% av artiklarna kommer från open accesstidskrifter. Den tryckta andelen domineras av några svenska tidskrifter från medicin och samhällsvetenskap. Täckningsgraden är 83% tillgång till artiklar som forskare från Malmö högskola publicerat.

Diskussion

I projektet undersöktes hur väl det periodiska beståndet överensstämmer med forskarnas publicering. Syftet var bland annat att se hur detta kan bidra till utveckling av bibliotekets forskarstöd. På detta sätt skapas också en bättre förståelse för Malmö högskolas forskares publiceringsmönster. Analyser av den insamlade datan kan utgöra underlag för såväl utveckling av våra samlingar, kvalitetsarbete med våra system och sist men inte minst mötet med våra forskare utifrån deras vardag.

Ett tydligt resultat av att använda och bearbeta data från vårt institutionella arkiv tillsammans med länkservern är att det kan leda till att identifiera fel i metadatan som kunde åtgärdas. Det tryckta beståndet av tidskrifter inkluderades också i länkservern under projektets gång. Det synliggjordes att ett antal möjliga open accesstidskrifter som Malmö högskola ger ut ej är tillgängliga i DOAJ⁴. Ett utvecklingsprojekt är att bistå tidskriftsredaktionerna med att matcha kraven för open access i DOAJ och registrering av metadata på artikelnivå.

⁴ (Directory of Open Access Journals) DOAJ är ett index på kvalitetsgranskade peer-reviewed open access tidskrifter.

Analyserna gav möjlighet att analysera vårt tidskriftsbestånd utifrån tillgång inom olika vetenskapsområden. Genom att synliggöra publiceringsmönstren och använda exempelvis indelningen i Bradford-zoner kan tidskrifter som är viktiga för forskare publiceringsmässigt bli synliga. Detta kan användas i diskussioner om prenumerationer. Det är ett sätt att synliggöra om det finns glapp i vår samling. I analysen ser man att det är en normal fördelning av tidskrifter som forskarna publicerar sig i. Några enstaka tidskrifter dominerar och därefter följer en lång svans med tidskrifter med enbart en artikel/tidskrift. Genom analyserna kan vi tydligt se tidskrifter med hög publicering men ingen tillgång. En förklaring till varför vi inte prenumererar på en av tidskrifterna i Bradford-zon ett i exemplet från Hälsa och samhälle kan lätt förklaras genom kostnaden för den enskilda tidskriften. Det kan användas som ett underlag i samtal med fakulteterna. Val när det gäller samlingsuppbyggnaden kan då leda till förklaringar om varför fjärrlån är en lösning för vissa enskilda tidskrifter.

En problematik som uppdagades i arbetet var att metadatan fick hämtas internt från vårt institutionella arkiv. Detta då det inte var möjligt att använda öppen data från Swepub. Även datan från vår länkserver finns bara tillgänglig internt. Det betyder att det inte går att genomföra jämförande analyser mellan lärosäten (som en benchmarking) utan samarbete. En annan begränsning är definitionen av vilka tidskrifter forskarna använder. I projektet är utgångspunkten att tidskrifter forskare publicerar sig i också är viktiga för deras informationssökning och kunskapsinhämtning. En analys baserad även på vilka tidskrifter som forskare refererar till skulle ge ett ännu bättre underlag. Det blir även en viss felmarginal i att matchningen av tidskrift baseras på hur den är inlagd som titel och ISSN vilket inte alltid är korrekt utfört. Analysen är baserad enbart på tidskrifter dvs ej konferensbidrag vilket får betydelse för täckningsgraden inom vissa discipliner där publiceringspraktikerna inkluderar konferenser. Ytterligare en problematik finns i förhållande till att vissa open accesskällor inte är tidskriftsbaserade, till exempel arXiv.

En viktig aspekt av analyserna är att definitionen av tillgång baseras på det som biblioteket aktivt visar upp genom systemen. Detta pekar på en viktig diskussion om hur vi gör open access tillgängligt via våra system. En sådan diskussion har nu lett till ett uppdrag att undersöka hur aktivering av open access eventuellt ska gå att likställa med förvärv. En fråga i det uppdraget är om det går att arbeta med open access på samma vis som med det köpta materialet. En slutsats är att metoden fungerar för att ta reda på om vi har det forskarna publicerar sig men att det i nuläget är svårt att basera analyserna på öppen data. Biblioteket får dessutom nytta av arbetet genom att felaktig metadata kan identifieras och rättas till under arbetet.

Vi talar ofta om tillgången till våra samlingar som en del av kärnverksamheten, vilket det också är. Vi menar dock att arbetet med vårt bestånd och forskares användning av tidskrifter eller andra publikationer är ett viktigt bidrag i att synliggöra aspekter av bibliotekets verksamhet som sällan behandlas som tydliga delar av forskarstödet. I forskarstödet inkluderar vi därför alla delar av bibliotekets verksamhet som kan bidra till att underlätta för forskare att utföra forskning.

Litteraturförteckning

Aquilonius, Elisabeth, Gullberg, Maria, Krook, Staffan, Lindholm, Jessica & Tapia, Pablo (2008). *Forskarservice - en utredning om Bibliotek och ITs verksamhet riktad mot högskolans forskning*. Malmö: Malmö högskola, Bibliotek och IT. Tillgänglig: <http://hdl.handle.net/2043/6118>

Egevad, Pär (2010). *Vad kan biblioteket utvinna ur en sökning på den egna publiceringen med hjälp av några bibliometriska metoder? - Ett exempel från Malmö högskola Odontologiska fakultetens publiceringar mellan 2000-2008*. Opublicerat manuskript.

Kärki, Riitta, Kortelainen, Terttu, Eriksson, Kristina & von Ungern-Sternberg, Sara (1998). *Introduktion till bibliometri*. Helsingfors: NORDINFO. (Nordinfo-publikation, 41).

Voog, Hanna, Arnebrant, Kristina, Bank, Matthias, Eltén, Annakim, Johnsson, Maria, Kjellberg, Sara et al (2013). *Tillgänglighet, närhet och synlighet – gemensamma utmaningar för LUB-nätverket för att möta forskares behov av stöd vid Lunds universitet*. Lund: Lunds universitets bibliotek. Tillgänglig: <http://lup.lub.lu.se/record/3404879> [20140623]