

Forskningskvalitet: kan det måles – og skal bibliotekerne gøre det?

Indtryk fra DF's årsmøde 2007

Af Piet Seiden

Det kan ikke komme bag på nogen, at vi nu også i universitetsverdenen skal måles og vejes. I vores tiltagende komplekse og utålmodige verden, er der et politisk ønske om at kunne overskue resultaterne af samfundets aktiviteter, at se på nytteværdien af dem.

Piet Seiden
Danmarks Biblioteksskole
ps@db.dk



Vi ser tendensen afspejlet i smiley-ordninger i fødevarebranchen, matematiktestning i folkeskolen, resultatkontrakter for offentlige institutioner og nu også et initiativ, som skal bringe forskningens resultater ind på en endimensional akse - ønsket om forskningsindikatorer. Det er tankevækkende, at hvor vi for ikke så mange år siden så frem til at computerne, ekspertsystemerne, skulle hjælpe til at optimere beslutningsprocesserne, så har vi idag tilsyneladende forladt denne mere teknokratiske tankegang og er slået over på en mere populistisk kurs, hvor simple endimensionale mål, som har en større overlevelseshance i massemediernes vridemaskine, har vundet frem som det objektive beslutningsgrundlag.

Derfor var det meget interessant at deltage i DF's årsmøde 2007, som havde fokuseret på temaet: Forskningskvalitet: kan det måles - og skal bibliotekerne gøre det? Jeg var ikke den eneste, som havde fundet det interessant, idet der var over 150 deltagere i mødet på Munkebjerg Hotel ved Vejle. Som keynote speaker havde man inviteret Kate Barker fra University of Manchester, som igennem en årrække har forsket i effekten af netop forskningsevaluering. Indlægget opsummerede erfaringer gjort i Storbritannien gennem 20 år i form af det såkaldte RAE-system (Research Assessment Exercise). Der er her tale om, at man har defineret 67 fagområder og for hvert af disse nedsat et bedømmelsesudvalg på

6-10 fageksperter, som skal give en kvalitativ vurdering af forskningen på de enkelte institutioner, baseret på forskningens resultater. Herved fik man så også en diskussion af, hvad der skal tælle som forskning, en debat som nok er helt nødvendig. I UK resulterede den i vedtagelse af en bred forskningsdefinition, som også inkluderer design og produktudvikling. Baseret på indberetningerne fra universiteterne, konstrueres til hver RAE en empirisk formel, som findes ved efterfølgende at "regne baglæns". Herefter klassificeres institutionen i en af de 5 klasser, der opereres med.

Universiteterne har haft blandede erfaringer med systemet. De positive træk har været, at det har introduceret et incitement til at være ansvarlig, stimuleret interessen for forskningen og frembragt den nødvendige infrastruktur, samt bidraget til udvikling af forskningsstrategier, hvilket samlet har styrket forskningen. Negative træk har været, at undervisningens status er aftaget, og at aktiviteter som ikke nødvendigvis giver et publiceringsoutput, som fx samarbejder med industri eller andre institutioner, ikke kan betale sig at satse på. Desuden har man set et "transfer marked" for forskere med et stort antal publikationer, som man forsøger at tiltrække for at booste sin institutions rating. Dette og andre uhensigtsmæssigheder, har medført, at man fra 2008 for en lang række fag går over til et metrik-baseret profileringssystem, baseret på tal for

publikationsaktivitet, eksternt finansieringsgrad og det man kunne kalde synlighedsgrad; deltagelse i redaktionelt arbejde, pristildelinger, konsulent- og kommissionsarbejde. I nogle fagområder såsom humaniora og kunst, matematik og samfundsfag vil man dog bevare det direkte peer-review. Der er dog kritiske røster fremme, idet systemet vurderes at ville koste over £100mio. at implementere og derfor af nogle anses som spild af ressourcer.

Vi blev altså klar over, at det var en kompliceret sag, at måle forskningens kvalitet – og hvorfor skal man egentlig også det? Direktør Lars Goldschmidt fra foreningen af rådgivende ingeniører forsøgte at sætte ord på dette spørgsmål under overskriften "forskningspolitik, -kvalitet og -formidling", baseret på det perspektiv han har opbygget gennem sine erfaringer fra forskningsmiljøer, forvaltning, lobbyvirksomhed og erhvervsliv. Overordnet handler det om, at vi i Danmark skal fastholde vores mulighed for værdibaseret produktion i en verden, hvor det er globaliseringen, som sætter spillereglerne. Ifølge LG, var konkurrenceparameteren ikke forskningens direkte erhvervsanvendelighed, altså viden som kan blive til produkter på kortest mulig tid, men derimod opretholdelsen af forskningsmiljøer, som så kan indgå i det LG benævnte "tusk-handlen med viden". Adgang til viden uanset oprindelse er altafgørende, da det ofte er billigere at skaffe sig adgang til viden frem for at skulle

Overordnet handler det om, at vi i Danmark skal fastholde vores mulighed for værdibaseret produktion i en verden, hvor det er globaliseringen, som sætter spillereglerne.

producere den selv, hvilket må siges at understrege betydningen af en stærk forskningsbibliotekssektor. Set i det lys, kan en kvalitetsindikator hjælpe med til at synliggøre, hvor der skal sættes forskningsmæssigt. Men de eksisterende kriterier for forskningskvalitet er ikke uproblematisk. Især samfundsfag og humaniora får problemer, hvis man anvender samme metodologi, som man gør inden for medicin, teknik og naturvidenskab. Derfor må en indikator bestå af flere facetter og Danmarks Forskningspolitiske Råd, som LG også repræsenterer, har anbefalet en model med et 6-punktskriterie; fire kvalitetsindikatorer (publicering, citationer, eksterne bevillinger, internationale samarbejder) og to relevansindikatorer (erhvervsrelevans, samfundsrelevans). Evalueringen skulle ske som selvevaluering, følges op med eksterne evalueringer som kontrol/supplement.

Således var der allerede mange begreber i luften, da Annette Dørge Jessen fra Forsknings- og Innovationsstyrelsens (FIST) nye evaluerings- og analyseenhed fik lejlighed til at redegøre for det danske initiativ på området. Det drejede sig primært om udviklingen af en model til forskningsevaluering, en bibliometrisk baseret forskningsindikator. I første omgang er det planen, at dele af basisbevillingerne til universiteterne fremover skal gives ud fra kvalitetsmål. Den bibliometriske indikator, som tager sit udgangspunkt i det tilsvarende initiativ i Norge (den norske model), skal tænkes som en national indikator, som også kan bruges lokalt. Indikatoren søger at tage højde for forskelle i publiceringspraksis mellem de forskellige hovedområder, idet den norske model opererer med dels en opsplitning af publikationernes type (artikel, monografi, osv.) og dels kvaliteten af denne, gennem en inddeling af publikationskanalen i to klasser. Der arbejdes for tiden med at tilpasse denne model til danske forhold. Et tilbagevendende punkt var understregningen af behovet for pålidelige data for publikationsaktiviteten, idet der er et erkendt problem her.

Derefter blev der åbnet op for en debat om, hvorvidt den norske model var velegnet til danske forhold. Erland Kolding Nielsen var hurtig til at påpege, at især historie, musikvidenskab og filologi kunne komme slemt i klemme, idet man fx ikke havde mulighed for at beskrive arbejdsindsatsen omkring sådan noget som kildeudgaver i denne model. Kate Barker havde også tidligere påpeget, at man i Australien havde set en eksplosion i antallet af publikationer efter indførelsen af en bibliometrisk indikator, idet mange artikler var blevet splittet op i mindre enheder, så de talte

flere gange. Det blev en god debat, som vidnede om et ønske fra FIST om at have en dialogbaseret tilgang, hvilket der blev taget godt imod.

Andendagsballet blev åbnet af Carsten Riis, dekan ved teologisk fakultet på Aarhus Universitet, som havde valgt at fokusere på problematikken omkring kvalitetsfordelte basismidler. Universiteterne er under forandring og ikke mindst på Århus universitet mærker man det stærkt efter fusioneringen med en række sektorforskningsinstitutioner, hvilket har medført, at en del rådgivningsopgaver er rykket ind i universitetsregi. Finansieringen har i mange år været båret af en historisk betinget basisbevilling og i nyere tid er STÅ-målene så kommet til som en regulering afhængig af studenteraktivitet. Nu kommer der så også et initiativ, som handler om inddragelse af en forskningskvalitetsindikator i bevillingsmodellen. Ydermere skal det eksisterende system omkring STÅ ændres til en uddannelseskvalitetsindikator og der suppleres med en vidensspredningsindikator. Generelt er universiteterne positivt indstillet på at blive målt og vejet i en konkurrencesituation, blot målene er baseret på ordentlige metoder og der tages hensyn til fagenes forskellighed mht. offentlig synlighed, publiceringstradition mm. Men, som han understregede gang på gang, så er der ikke tale om den højere retfærdighed. Som det også blev diskuteret tidligere, er en publiceringsindikator ud fra den norske model diskutabel på en række punkter, som fx hvornår er det forskning og hvornår er det forskningsformidling? I en række humanistiske fag, er det ikke muligt reelt at skelne her. Begrebet peer-review er også meget forskelligt praktiseret, skal man vægte det? Skal man opgøre antal publikationer i forhold til antal VIP-ansatte, hvoraf nogle kan være på deltid, eller i forhold til bevillingens størrelse?

Herefter fik vi et biblioteksperspektiv på problemstillingerne, idet overbibliotekar Niels-Henrik Gylstorff fra AUB diskuterede sine tanker om bibliotekets rolle i målingen af kvaliteten af moderinstitutionens forskning. Som det også fremgik af tidligere indlæg, så er datakvaliteten vigtig, hvis man skal kunne udsige noget om institutionens forskning. Traditionelt har forskerne aldrig været gode til at indrapportere deres forskning og bibliotekerne på de enkelte universiteter har i mange tilfælde været involveret i en række initiativer, som har haft til hensigt at forbedre forskningsregistreringen. Derfor ligger der en naturlig opgave for bibliotekerne i at indgå i en valideringsproces, så man sikrer sig, at data er

korrekte. Omfanget af bibliotekets engagement skal nok afgøres lokalt, men det er ikke nogen helt trivial problemstilling at skaffe gode publikationsdata til en kvalitetsindikator.

Men hvad kan man egentlig udsige baseret på publicering og citationer? Jeppe Nicolaisen fra Danmarks Biblioteksskole gav en yderligere uddybning af nogle af de problemer, der knytter sig til at lave forskningsanalyse baseret på publikationstal eller citationsanalyse. Indlægget formede sig som en kritisk analyse af især citationsanalyse, hvor han plæderede for, at man nemt kan komme til at fejlfortolke resultaterne, med mindre man inddrager en forståelse for de sociologiske omstændigheder, som arbejdet er blevet til under.

For at følge diskussionen omkring datagrundlaget helt til dørs, havde professor Benny Laurup det sidste indlæg med titlen "Forskningskvalitetsmål og målemetoders pålidelighed". Sammen med sine medarbejdere har BL igennem nogen tid arbejdet med at analysere validiteten af et lave diskrimination på basis af citationer. Højenergifysik, som er BL's disciplin, er karakteriseret ved en meget homogen publiceringstradition i relativt få tidsskrifter og er meget godt dækket over en længere periode i en enkelt database, SPIRES. Baseret på et større udtræk fra denne publikationsdatabase, har BL og medarbejdere undersøgt nogle forskellige kvalitetsmål, fx antallet af publikationer og citationslistens middelværdi. Ud fra disse kvalitetsmål søgte man at indplacere forskere i ti forskellige klasser, hvorefter man udregner sandsynligheden for, at en given artikel ville blive placeret i en af disse grupper, dvs. hvor sikker en klassifikation kan man foretage af en forsker baseret på et givet kvalitetsmål?

Det viste sig at være svært. Alene baseret på antallet af publikationer var det ikke muligt i det givne datasæt at klassificere en given forfatter med hensyn til antal citationer, hvilket må give stof til eftertanke. Middelværdien af antal citationer viste sig at være bedre til at klassificere efter og medianen til citationsantal endnu bedre. Imidlertid kunne denne klassifikation for mange af klassernes vedkommende kun foretages med en rimelig sikkerhed, såfremt der var tale om en forsker, som havde forfattet mere end 50 artikler. Konklusionen på analysen var, at indtil videre er vi nok bedst tjent med at læse artiklerne, hvis vi vil have en bedømmelse af forskerens kvalitet.

Sammenfattende må det siges at have været et overordentlig interessant årsmøde i DF, som virkelig kom godt rundt om det faglige tema og fik sat mange tanker i gang.

Kate Barker havde også tidligere påpeget, at man i Australien havde set en eksplosion i antallet af publikationer efter indførelsen af en bibliometrisk indikator, idet mange artikler var blevet splittet op i mindre enheder, så de talte flere gange.