

Professor Jan Fagerberg, Universitetet i Oslo

jan.fagerberg@tik.uio.no

20.04.2010

Om måling av forskning

Siden konkurranse og prestasjonsmåling i mange år har stått sentralt i forskningssystemet, har måter å foreta slik måling på naturlig nok tiltrukket seg en del oppmerksomhet og studiet av hvordan dette best kan gjøres har utviklet seg til et eget fagfelt. Den tradisjonelle måten å gjøre dette er naturligvis peer review, fagfelle-vurdering, som har spilt og nok kommer til å fortsette å spille en svært viktig rolle i forskningssystemet. Ordningen er imidlertid ressurs- og tidkrevende, særlig ved vurdering av hele miljøer, og resultatet kan være følsomt for valget av "peers", spesielt på felter hvor det hersker faglig uenighet (og det gjør det jo ofte). Det har derfor vært lagt en god del arbeid i forskningssystemet i Norge og andre land for å finne fram til mer objektive vurderingsformer som kan supplere "peer review" metoden eller være innspill til disse. To forhold som har fått mye oppmerksomhet er *omfanget* av faglig produksjon og den *innflytelsen* som denne har. Faglig produksjon kan ta ulike former, som for eksempel i bildende kunst, musikk og arkitektur, og innflytelsen kan også skje på forskjellige måter. Det er tvilsomt om man konstruere ett sett kriterier som kan fange opp så store forskjeller. Men i mesteparten av forskningssystemet er den dominerende produksjonsmåten skriftlige arbeider og jeg skal i det følgende begrense meg til å diskutere metoder for å måle omfanget og innflytelsen av disse.

Skriftlige arbeider kan ta flere former. I forskningssystemet er de viktigste artikler i tidsskrifter, bøker (monografier) og artikler i redigerte bøker. Bruken av disse publikasjonene varierer mye mellom ulike forskningsfelter. Inspirert av praksisen i naturvitenskapen har mye av oppmerksomheten vært fokusert på den første av disse publikasjonsformene. Det er imidlertid uten videre klart at dette i mange tilfeller vil gi et skjevt bilde, og at det er ønskelig å åpne for et bredere tilgang av publikasjonstyper.

For praktiske formål har det også vært eksperimentert en god del med å rangere publikasjoner etter den antatte kvalitet som publikasjonskanalen har. Et mye brukt mål er den såkalte "impact factor", dvs. det antall siteringer som artiklene i et tidsskrift gjennomsnittelig får gjennom en tidsperiode, som ISI Thompson kalkulerer for de tidsskriftene som er registret i deres system. Garfield, som startet ISI og konstruerte dette målet, var imidlertid nøye med å påpeke at "journal impact factor" er et mål for tidsskrifters innflytelse, ikke kvaliteten eller innflytelsen til bidragene i disse. Det skyldes, som Garfield (1998) påpekte, at variasjonen i kvalitet og/eller innflytelse innen tidsskrifter er svært stor. Selv de beste tidsskriftene publiserer av og til artikler som få eller ingen finner interessante. Tilsvarende kan små og ukjente tidsskrifter, eller bøker, inneholde bidrag som får svært stor

innflytelse. Man kan derfor ikke bruke publikasjonskanalens meritter til å trekke slutninger om kvaliteten eller innflytelsen på noe som er publisert der.

En modifisert versjon av dette er, eventuelt med utgangspunkt i "journal impact factor", å søke å klassifisere publikasjonskanalene i et mindre antall "kvalitetsklasser". En slik klasseinndeling kan også baseres på - eller justeres gjennom - vurderinger fra fagfolk. Et slikt system praktiseres i Norge i dag (Sivertsen 2008). Men den prinsipielle kritikken mot å bruke karakteristika ved publikasjonskanaler til å måle kvaliteten ved forskning har like stor gyldighet i dette tilfellet. Seglen (2009) har påvist at inndelingen i to klasser, slik den i dag anvendes i det norske systemet, ikke har noen prediksjonskraft med hensyn på hvor innflytelsesrike bidragene viser seg å være. Et ytterligere problem er at dette systemet, slik det praktiseres i Norge, åpner for at enkeltpersoner eller miljøer med innflytelse i forskningssystemet kan få "ekstrapremiering" for publisering i kanaler de har et nært forhold til. Slike forhold bidrar til å underminere legitimiteten til ordningen. Det beste ville derfor antagelig være å følge Seglens anbefaling om å avskaffe denne klasseinndelingen og la alle publikasjonskanaler telle like mye så lenge visse formelle kriterier er oppfylt.

Den virkelige testen for et forskningsbidrag er ikke hvor det blir publisert, men hvordan det blir mottatt av andre forskere. Dette kan måles gjennom siteringer. Riktignok er det noen mulige fallgruber, forskere har for eksempel en tendens til å sitere seg selv, og det kan hevdes at slike siteringer har mindre "verdi" (noe som ikke nødvendigvis er tilfelle selvsagt). Men dette vil det enkelt være mulig å justere for (og det vil uansett ikke være av særlig betydning når det gjelder høyt siterte arbeider). En annen innvending er at siteringshyppigheten varierer mellom forskningsfelt, noe som er korrekt, men som det også kan justeres for om man ønsker. Videre hevdes at det at mange siteringer ikke nødvendigvis signaliserer høy kvalitet fordi man kan få mange siteringer hvis man sier noe helt feil (som andre så søker å motbevise). Dette kan selvsagt ikke utelukkes, men det er antagelig et uhyre sjeldent fenomen, som neppe har stor betydning i praksis (og feiltagelser trenger heller ikke være dårlig forskning). Konklusjonen er at den vurderingen som andre forskere gir av et forskningsbidrag verdi gjennom siteringer, antagelig er den beste kvantitative indikatoren vi i dag har for å vurdere (den samfunnsmessige) verdien av forskning. Forskning som ikke interesserer noen er av svært begrenset samfunnsmessig verdi uansett hvor kunnskapsrik forskeren ellers måtte være.

Siden siteringer skjer i ettertid betyr det at det vil være et visst tidsspenn - som godt kan være fem år eller mer - mellom når publikasjonen skjer og innflytelsen kan måles. Dette er knapt noe argument mot å ta denne dimensjonen i betraktning. Forskning er en kumulativ aktivitet. For forskere så vel som miljøer vil det være umulig å gi en god vurdering av deres produktivitet uten legge et lengre tidsperspektiv til grunn. Endringer fra et år til et annet er neppe noe å legge særlig vekt på uansett.

Det hevdes ofte at siteringsanalyse gjør det umulig å undersøke innflytelsen av andre publikasjoner enn tidsskrifter. Men det er ikke riktig. Også bøker, kapitler etc. siteres i rikt

monn i tidsskifter. Referanser i tidsskrifter kan ses som resultat av en seleksjonsprosess hvor et bredt utvalg publikasjonsformer inngår. Derimot inneholder datakilder som Web of Science (Thomson) og Scopus (Elsevier) i liten grad referanser i bøker. Men det finnes andre datakilder som har en bredere profil i så måte. Den viktigste er Google Scholar som i tillegg til referanser i tidsskrifter og (en del) bøker også inneholder workingpapers, rapporter m.v. Denne datakilden er i motsetning til de kommersielle databasene fritt tilgjengelig.

I Australia er utviklet et eget (gratis)program, "Publish or Perish", som forskere kan bruke til å konstruere en detaljert siteringsprofil på seg selv (eller andre) ved hjelp Google Scholar. Innovasjonssenteret "Circle" ved Lund's universitet (Nordens største innovasjonsforskningsmiljø) har gjennom flere år analysert medarbeidernes siteringer ved hjelp av både Web of Science(Thomson) og Google Scholar og publisert dette i årsrapporten. Resultatene viser at det er langt flere treff med Google Scholar (ca fire ganger så mange). Men rangeringen av forskernes produktivitet er nokså lik uavhengig av hvilken datakilde som blir brukt. Kanskje er det slik at de forskerne som siteres mest i vitenskapelige tidsskrifter også er de som har mest gjennomslag i andre sammenhenger? I så fall betyr det kanskje ikke så mye hvilken database man bruker og hva slags siteringer man inkluderer. Dessuten er jo siteringer i såkalt "grå" litteratur også tegn på at forskningen blir brukt.

Konklusjonen er at dersom man skal måle forskning, så er det ikke tilstrekkelig å se utelukkende på produksjonen. Det man først og fremst må legge vekt på er hvordan forskningen blir verdsatt i forskersamfunnet (og eventuelt også i samfunnet mer generelt). Siteringer er det beste målet for dette. Slik måling må skje i et lengre tidsperspektiv (flere år). Det systemet som er bygget opp i Norge med selvregistrering av publikasjoner (Sivertsen 2008) er en viktig ressurs å bygge videre på. Knyttet til dette burde det være mulig å utvikle et bredt sett indikatorer (basert på produksjon, sitering, bruk/spredning) som hver enkelt forsker eller forskningsmiljø enkelt kunne tilordne disse publikasjonene. Et eksempel på hvordan dette eventuelt kan gjøres er det såkalte REPEC-systemet, som drives av en uavhengig stiftelse initiert av forskere og bibliotekarer innenfor økonomi-faget (se box).

Hvilken nytte kan man evt. ha av dette? Det kan for det første være nyttig for den enkelte forsker/forskningsgruppe som en dokumentasjon på deres meritter i forskning til bruk for eksempel i søknadsprosesser og overlegninger om fordeling av ressurser internt. Det kan også brukes av ledelsen i den organisasjonen de inngår i til å identifisere felter hvor organisasjonen står sterkt og hvem som står bak dette. Slike data er også egnet til å sammenligne forskningsmiljøer og organisasjonsledd innen samme felt på tvers for eksempel universiteter og institutter. De kan også, brukt med en viss forsiktighet, være nyttig i bredere sammenligninger mellom de ulike leddene i forskningssystemet.

BOX. REPEC-systemet (<http://ideas.repec.org/>)

REPEC er et IT-basert registreringssystem for forskning i "economics and related disciplines". Det drives av en uavhengig stiftelse og er gratis for de som deltar. Systemet bygger på samarbeid med tidsskriftutgivere, forlag og organisasjoner som har egne publikasjonsserier (working papers og lignende). Forskerne må selv registrere seg. Pr. i dag har systemet over 20 000 registrerte brukere i verden som helhet. Som ledd i arbeidet registrerer systemet publikasjoner, nedlastinger og siteringer i de serier som er registrert i systemet. Slike data danner grunnlag for rangering av individuelle forskere, de organisasjonsleddene de er knyttet til og publikasjonsserier.

REPEC opererer med en rekke indikatorer for hver enkelt forskers innsats. Disse bygger på

- Antall registrerte publikasjoner
- Siteringer
- H-indeks (h publikasjoner som er sitert h ganger eller mer)
- Antall forfattere som siterer deg
- Antall sider publisert i tidsskrifter
- Abstract views
- Nedlastinger

I rangeringen veies disse indikatorene, som bygger på data for så vel produksjon som bruk av forskning, sammen til en felles indeks. Enhver bruker av systemet kan imidlertid lage sin egen indeks med så mange eller få indikatorer hun eller han finner relevant.

Referanser

- Garfield, E. (1998) The Impact Factor and Using It Correctly, *Der Unfallchirurg*, 48(2) p.413, June.
- Seglen, P. O. (2009) "Er tidsskrift-renneme og artikkeltelling adekvate mål for vitenskapelig kvalitet og kvantitet", I *Det Norske Videnskaps-Akademi: Hvordan måle vitenskap? Søkelys på bibliometriske metoder*, Oslo, s. 39-70
- Sivertsen, G. (2008) *Norsk Vitenskapsindeks, Rapport 33/2008*, NIFUSTEP, Oslo