

Bibliometrija: Leidenski manifest o vrednotenju raziskovalnega dela

- **Diana Hicks,**
- **Paul Wouters,**
- **Ludo Waltman,**
- **Sarah de Rijcke**
- **Ismael Rafols**

Nature **520**, 429–431 (23. april 2015)

Prevod/Translation: Mojca Breznik, Univerza v Mariboru (9. julij 2019)

Vir/Source: <https://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>

Diana Hicks, Paul Wouters in kolegi pozivajo k uporabi desetih načel pri vrednotenju raziskovalnega dela.

Pri ocenjevanju znanosti imajo podatki vedno večjo vlogo. Ocenjevanje raziskovalnega dela, ki je bilo nekoč prilagojeno in so ga izvajali strokovnjaki, je zdaj rutinsko ter temelji na metrikah¹. Kar predstavlja težavo, je to, da so trenutno glavno vodilo pri vrednotenju podatki namesto presoja. Metrike se uporabljajo vse pogosteje. Običajno so dobronamerne, vendar ne razpolagajo vedno z ustreznimi podatki in se pogosto napačno uporabljajo. Z orodji, ki so bila zasnovana za izboljšanje sistema, lahko sistemu celo škodujemo, saj vrednotenje vedno pogosteje izvajajo organizacije, ki nimajo ustreznega znanja ali priporočil v povezavi z dobrimi praksami ter interpretacijo.

Pred letom 2000 so strokovnjaki za strokovne analize uporabljali indeks znanstvenih citatov (Science Citation Index) na CD-ROM-u, ki so ga pripravili na Inštitutu za znanstveno informiranje (ISI). Leta 2002 je Thomson Reuters vzpostavil integrirano spletno platformo, s čimer je postala podatkovna zbirka Web of Science dostopna širšemu občinstvu. Ustvarjeni sta bili konkurenčni podatkovni bazi citatov: Elsevier's Scopus (izdana leta 2004) in Google Scholar (beta različica izdana leta 2004). Uvedena so bila spletna orodja za lažjo primerjavo obsega in vpliva raziskovalnega dela posameznih institucij, kot so InCites (uporablja Web of Science) in SciVal (uporablja Scopus) ter programska oprema za analizo posameznih profilov citiranja z uporabo iskalnika Google Scholar (Publish or Perish, izdan leta 2007).

Leta 2005 je Jorge Hirsch, fizik s Kalifornijske univerze v San Diegu, predlagal *h*-indeks, ki je populariziral štetje citatov za posamezne raziskovalce. Zanimanje za faktor vpliva strokovnih revij je vztrajno raslo vse od leta 1995 (glejte '[Impact-factor obsession](#)').

V zadnjem času pridobivajo metrike, ki so povezane z družbeno rabo in spletnimi komentarji, čedalje večji pomen — F1000Prime je bila vzpostavljena leta 2002, Mendeley leta 2008 in [Altmetric.com](#) (s podporo Macmillan Science and Education, ki ima v lasti Nature Publishing Group) leta 2011.

Kot strokovnjaki za scientometrijo, družboslovci in strokovno podporno osebje za raziskovalno dejavnost smo z vedno večjo zaskrbljenostjo opazovali splošno razširjeno

napačno uporabo kazalnikov pri vrednotenju znanstvene dejavnosti. V nadaljevanju je navedenih zgolj nekaj od številnih primerov. Univerze po vsem svetu so postale obsedene s svojim položajem na svetovnih lestvicah (kot sta Šanghajska lestvica in seznam revije *Times Higher Education*), čeprav tovrstni sezname temeljijo na po našem mnenju nepravilnih podatkih in poljubnih kazalnikih.

Ponekod se pri zaposlovanju kandidatov zahtevajo vrednosti h -indeksa. Številne univerze se pri odločitvah o napredovanju opirajo na mejne vrednosti h -indeksa ter na število objavljenih člankov v revijah z visokim faktorjem vpliva. Raziskovalci se ponašajo s temi ocenami v svojih življenjepisih, zlasti na področju biomedicine. Mentorji po vsem svetu nagovarjajo doktorske študente, naj svoje delo objavijo v revijah z visokim faktorjem vpliva in pridobijo zunanje financiranje pred pripravo.

Nekatere univerze v Skandinaviji in na Kitajskem dodeljujejo sredstva ali dodatke za raziskovalno delo na podlagi števil: na primer z računanjem individualnih ocen vpliva za dodelitev »sredstev na podlagi uspešnosti« ali z izplačilom dodatka raziskovalcem za objavo v reviji s faktorjem vpliva, ki je višji od 15 (ref. 2).

V mnogih primerih raziskovalci in ocenjevalci še vedno presojujejo uravnoteženo. Kljub temu pa je postala zloraba raziskovalnih metrik preveč razširjena, da bi jo bilo mogoče spregledati.

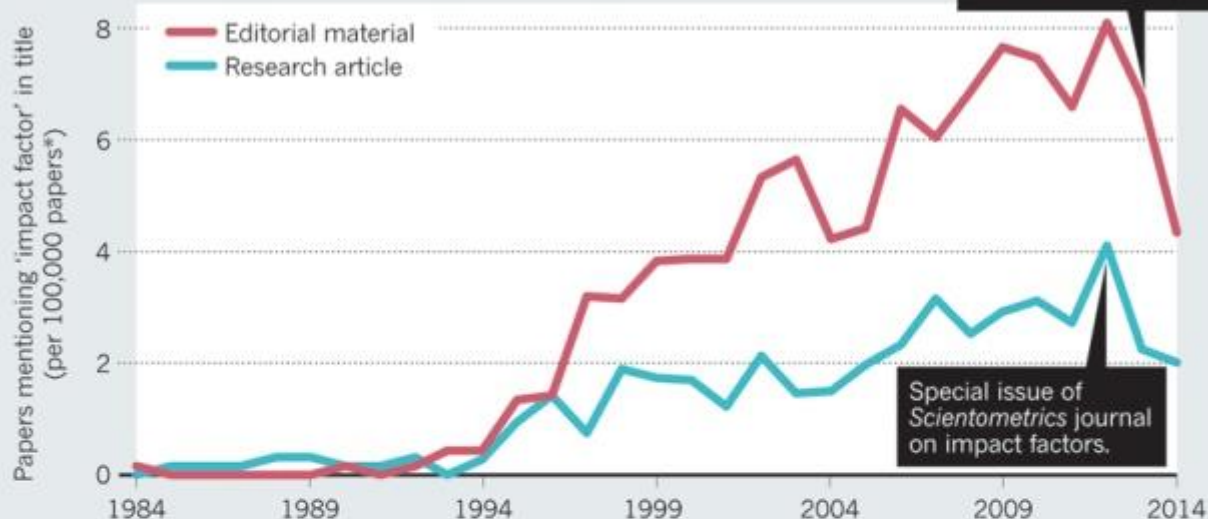
Zato predstavljamo Leidenski manifest, ki je dobil ime po konferenci, na kateri se je izoblikoval (glejte <http://sti2014.cwts.nl>). Njegovih deset načel ni nekaj novega za strokovnjake s področja scientometrije, vendar pa jih noben od nas ne bi mogel natančno povzeti v celoti, saj še niso bila kodificirana. Tisti, ki osvetljujejo razmere na tem področju, kot je Eugene Garfield (ustanovitelj inštituta ISI), vedno znova navajajo nekatera od teh načel^{3,4}. Vendar pa njihova dognanja ne dosežejo osebja na univerzi, ki ob poročanju ocenjevalcev ni strokovno seznanjeno z relevantno metodologijo. Znanstveniki med iskanjem literature, s pomočjo katere bi lahko podali vrednotenje, naletijo na gradivo, ki je razkrojeno po raznoraznih strokovnih revijah, do katerih nimajo vedno dostopa.

Ponujamo ta izbor dobrih praks s področja vrednotenja raziskovalnega dela, ki temelji na metrikah, da se lahko nanj oprejo tako raziskovalci v odnosu do ocenjevalcev kot tudi ocenjevalci v odnosu do kazalnikov.

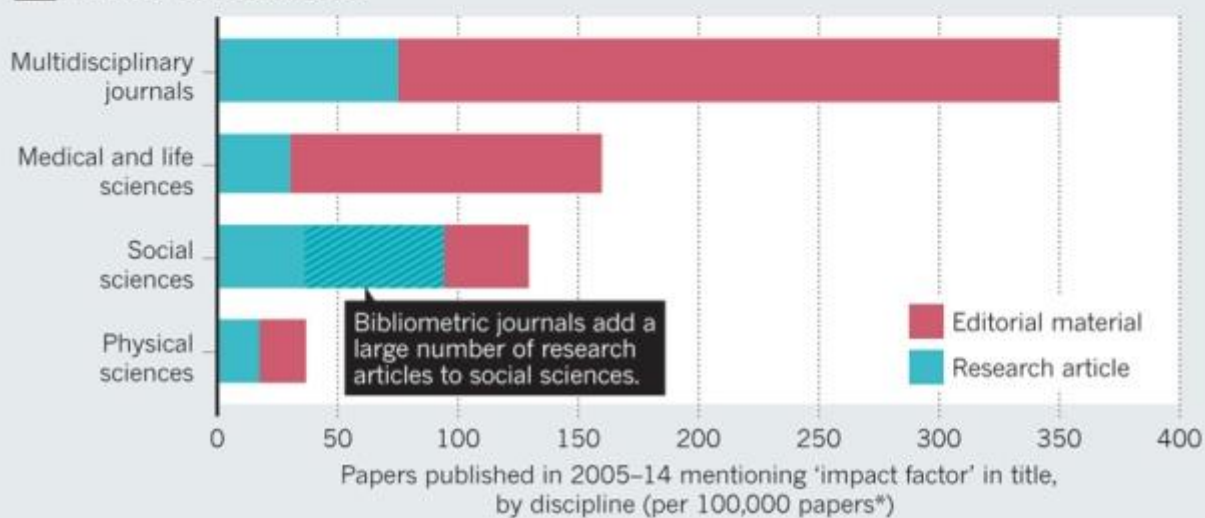
IMPACT-FACTOR OBSESSION

Soaring interest in one crude measure — the average citation counts of items published in a journal in the past two years — illustrates the crisis in research evaluation.

1 ARTICLES MENTIONING 'IMPACT FACTOR' IN TITLE



2 WHO IS MOST OBSESSED?



*Indexed in the Web of Science. †DORA, San Francisco Declaration on Research Assessment.

Vir podatkov: Thomson Reuters Web of Science; Analiza: D.H., L.W.

Deset načel

1) Kvantitativno vrednotenje naj podpira kvalitativno vrednotenje raziskovalnih institucij in skupin raziskovalcev s strani strokovnjakov. Kvantitativno vrednotenje lahko omili enostranskost pri strokovnih pregledih ter spodbudi nadaljnje razprave. S tem se podkrepijo strokovni pregledi, saj je o kolegih težko presojati brez kopice relevantnih informacij. Vendar pa ocenjevalci ne smejo podleči skušnjavi, da bi svoje odločanje prepustili številkam. Kazalniki ne smejo nadomestiti utemeljene presoje. Vsak mora ohraniti odgovornost za svoja vrednotenja.

2) Raziskovalna uspešnost naj se ocenjuje v luči raziskovalnih ciljev institucije, skupine ali raziskovalca. Programske cilje je treba opredeliti na samem začetku, kazalniki za merjenje uspešnosti pa se morajo jasno nanašati na te cilje. Pri izbiri kazalnikov in načinov njihove uporabe naj se upošteva širši družbeno-ekonomski in kulturni kontekst. Znanstveniki sledijo različnim raziskovalnim ciljem. Raziskovalno delo, ki premika meje akademskega znanja, se razlikuje od raziskovalnega dela, ki se osredotoča na iskanje rešitev za družbene probleme. Ocena lahko temelji na dosežkih, ki so pomembni za politiko, industrijo ali javnost, in ne toliko na pojmovanju odličnosti v akademskem svetu. Noben model vrednotenja ne ustreza popolnoma vsem kontekstom.

3) Varovati je potrebno odličnost v lokalno pomembnih raziskavah. V mnogih delih sveta se znanstvena odličnost enači z objavami v angleškem jeziku. Španska zakonodaja na primer navaja željo po objavljanju španskih strokovnjakov v revijah z visokim faktorjem vpliva. Faktor vpliva se izračuna za revije, ki so indeksirane v ameriškem in še vedno v večji meri na angleškem jeziku temelječem servisu Web of Science. Te pristranskosti so zlasti problematične v družboslovju in humanistiki, kjer je raziskovalno delo tesneje povezano z regionalnim in nacionalnim okoljem. Tudi številna druga področja imajo nacionalno ali regionalno dimenzijo – na primer epidemiologija virusa HIV v podsaharski Afriki.

Pluralizem in družbeni pomen se pogosto potlačita, da bi bili strokovni članki dovolj zanimivi za uvrstitev v revije z visokim faktorjem vpliva: revije, ki objavljajo v angleškem jeziku. Španski družboslovci, ki so pogosto citirani v servisu Web of Science, so pri svojem delu uporabili abstraktne modele ali pa preučujejo ameriške podatke. Specifičnost družboslovnih tem se v člankih z visokim faktorjem vpliva, ki so zapisani v španskem jeziku, izgubi. Gre za teme, kot so lokalno delovno pravo, zdravstveno varstvo za starejše ali zaposlovanje priseljencev⁵. Vzpostavitev metrik, temelječih na literaturi, ki ni zapisana v angleškem jeziku, bi pripomogla k prepoznavanju in nagrajevanju odličnosti lokalno pomembnega raziskovalnega dela.

4) Podatke in analitične procese je potrebno hraniti na odprt, pregleden in enostaven način. Oblikovanje podatkovnih zbirk, potrebnih za ocenjevanje, naj sledi jasno opredeljenim pravilom, ki se postavijo pred zaključkom raziskav. To je bila nekaj desetletij običajna praksa med akademskimi in komercialnimi skupinami, ki so zasnovale metodologijo bibliometrijskega vrednotenja. Te skupine so se sklicevale na protokole, objavljene v strokovno pregledani literaturi. Tovrstna preglednost je omogočila nadzor. Tako je na primer leta 2010 javna razprava o tehničnih lastnostih pomembnega kazalnika, ki ga je uporabljala ena od naših skupin (Center za znanstvene in tehnološke študije na Univerzi v Leidnu na Nizozemskem) vodila v ponovno izračunavanje tega kazalnika⁶. Tudi za nove udeležence bi morali veljati enaki standardi in nihče ne bi smel pristati na neuničljiv ocenjevalni stroj.

Kazalnike naj odlikuje preprostost, saj se s tem poveča preglednost. Preproste metrike pa lahko popačijo sliko (glejte 7. načelo). Ocenjevalci si morajo prizadevati za uravnoteženost – preproste kazalnike, ki upoštevajo zapletenost raziskovalnega procesa.

5) Tisti, ki se jih ocenjuje, morajo imeti pravico preveriti podatke in analize podatkov. Da bi zagotovili kakovost podatkov, bi morali vsi raziskovalci, ki so vključeni v bibliometrijske študije, imeti možnost preveriti, ali so bili njihovi dosežki pravilno prepoznani. Vsi, ki vodijo in upravljajo ocenjevalne postopke bi morali zagotoviti pravilnost

podatkov s samopreverjanjem ali revizijo tretje strani. Univerze bi lahko to vpeljale v svoje informativne sisteme za raziskovalno dejavnost, postati pa bi morale tudi vodilno načelo pri izbiranju ponudnikov tovrstnih sistemov. Zbiranje in obdelovanje točnih, visokokakovostnih podatkov zahteva čas in denar. Temu je treba nameniti ustrezna sredstva.

6) Upoštevati je potrebno razlike med vedami pri načinu objav in praksah

citiranja. Najbolje je določiti sklop morebitnih kazalnikov ter omogočiti, da strokovnjaki posameznih ved izberejo med njimi. Nekaj let nazaj je skupina zgodovinarjev iz Evrope prejela relativno nizko oceno pri nacionalnem ovrednotenju na podlagi strokovnega pregleda, ker so pisali knjige namesto člankov v revijah, ki so indeksirane s strani Web of Science. Zgodovinarji so imeli to smolo, da so bili del oddelka za psihologijo. Zgodovinarjem in družboslovcem se v nabor publikacij štejejo knjige in literatura v nacionalnem jeziku, računalniškim strokovnjakom pa prispevki na konferencah.

Stopnje citiranosti se med vedami razlikujejo: pri najvišje uvrščenih revijah na področju matematike se faktorji vpliva gibljejo okrog 3, pri najvišje uvrščenih revijah na področju celične biologije pa okrog 30. Zahtevani so normalizirani kazalniki, najzanesljivejša metoda normalizacije pa temelji na percentilih: vsak strokovni članek se ovrednoti na podlagi percentila, h kateremu sodi glede na porazdelitev citiranosti znotraj vede (npr. najboljši 1 %, 10 % ali 20 %). Ena pogosto citirana objava nekoliko izboljša uvrstitev univerze na lestvici, ki temelji na percentilnih kazalnikih, na lestvici, ki temelji na povprečnih vrednostih citiranosti⁷, pa lahko pomeni premik iz sredine v sam vrh lestvice.

7) Ocena individualnih raziskovalcev mora temeljiti na kvalitativni oceni njihovega raziskovalnega dela.

S starostjo raziskovalca se viša tudi njegov *h*-indeks, četudi ne objavlja novih strokovnih člankov. *H*-indeks se razlikuje glede na vedo: naravoslovci dosegajo vrednost 200; fiziki 100 in družboslovci 20–30 (ref. 8). Njegova vrednost je odvisna od baze podatkov, tako je na primer *h*-indeks raziskovalcev s področja informatike okrog 10 v Web of Science, v Google Scholar⁹ pa 20–30. Branje in vrednotenje raziskovalčevega dela je veliko ustrežnejše kot opiranje na eno številko. Tudi pri primerjavi velikega števila raziskovalcev je najbolje upoštevati več podatkov o strokovnem delu posameznika, izkušnjah, aktivnostih in vplivnosti.

8) Izogibati se je potrebno neustreznim konkretizacijam in lažnim

preciziranjem. Znanstveni in tehnološki kazalniki so pogosto konceptualno nejasni in negotovi ter zahtevajo močne predpostavke, ki niso univerzalno sprejete. Dolgo časa se je na primer razpravljalo o pomenu štetja citatov. Zato se v primerih dobre prakse uporablja več kazalnikov za prikaz bolj zanesljive in pluralistične slike. Če je mogoče negotovost in napake količinsko opredeliti, na primer z uporabo paličnih grafiknov za prikaz razpona napake, bi morali tovrstni podatki spremljati objavljene vrednosti kazalnikov. Če to ni mogoče, bi se morali oblikovalci kazalnikov izogniti vsaj lažnemu preciziranju. Faktor vpliva revije je na primer objavljen na tri decimalna mesta natančno, da se izognemo izenačenim rezultatom. Ob upoštevanju konceptualne nejasnosti in naključne spremenljivosti števila citiranj pa razlikovanje med revijami na podlagi zelo majhnih razlik med faktorji vpliva ni smiselno. Izogibati se je treba lažnemu preciziranju: šteje le eno decimalno mesto.

9) Upoštevati je potrebno sistemske učinke ocenjevanja in različnih kazalnikov. S

spodbudami, ki jih ustvarijo, kazalniki spreminjajo sistem. Tovrstne učinke moramo

predvideti. To pomeni, da ima sklop kazalnikov vedno prednost, saj bi en sam kazalnik privabil manipulacije in preusmeritev cilja (pri čemer merjenje postane cilj). Tako je na primer v 90. letih prejšnjega stoletja Avstralija financirala univerzitetno raziskovalno delo z uporabo formule, ki v večji meri temelji na številu strokovnih člankov, objavljenih s strani inštitutov. Univerze so lahko izračunale »vrednost« članka v strokovno pregledani reviji; leta 2000 je to bilo 800 avstralskih dolarjev (okrog 480 ameriških dolarjev leta 2000) v okviru financiranja raziskovalnega dela. Avstralski raziskovalci so pričakovano objavljali vedno večje število člankov, vendar v manj pogosto citiranih revijah, s čimer je kakovost člankov padla¹⁰.

10) Kazalnike raziskovalnega dela je potrebno redno pregledovati in

posodabljati. Raziskovalni nameni in ocenjevalni cilji se spreminjajo, sam raziskovalni sistem pa se razvija sočasno. Nekdaj uporabne metrike postajajo neustrezne, zato se izoblikujejo nove. Sisteme kazalnikov je treba pregledati in po potrebi prilagoditi. Avstralija je prepoznala učinke svoje preproste formule in leta 2010 predstavila zahtevnejšo pobudo *Excellence in Research for Australia*, ki poudarja kakovost.

Nadaljnji koraki

Ob upoštevanju teh desetih načel lahko vrednotenje raziskovalnega dela pomembno prispeva k razvoju znanosti in njenih povezav z družbo. Metrično vrednotenje raziskovalnega dela lahko zagotovi ključne podatke, ki bi jih bilo težko zbrati ali razumeti z individualno strokovno oceno. Ne smemo pa dovoliti, da bi se kvantitativni podatki prelevili iz sredstva v cilj.

Najboljše odločitve se sprejemajo z združevanjem zanesljive statistike in občutka za cilj in naravo raziskovalnega dela, ki se ocenjuje. Potrebni so tako kvantitativni kot kvalitativni dokazi, ki so vsak po svoje objektivni. Sprejemanje odločitev o znanosti mora temeljiti na visokokakovostnih procesih, ki jih podpirajo najkakovostnejši podatki.

Reference

1. Wouters, P. in *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (eds Cronin, B. & Sugimoto, C.) 47–66 (MIT Press, 2014).
2. Shao, J. & Shen, H. *Learned Publ.* **24**, 95–97 (2011).
3. Seglen, P. O. *Br. Med. J.* **314**, 498–502 (1997).
4. Garfield, E. *J. Am. Med. Assoc.* **295**, 90–93 (2006).
5. López Piñeiro, C. & Hicks, D. *Res. Eval.* **24**, 78–89 (2015).
6. van Raan, A. F. J., van Leeuwen, T. N., Visser, M. S., van Eck, N. J. & Waltman, L. J. *Informetrics* **4**, 431–435 (2010).
7. Waltman, L. et al. *J. Am. Soc. Inf. Sci. Technol.* **63**, 2419–2432 (2012).
8. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).
9. Bar-Ilan, J. *Scientometrics* **74**, 257–271 (2008).
10. Butler, L. *Res. Policy* **32**, 143–155 (2003).