

LEIDENO MANIFESTAS

dėl mokslo rezultatų vertinimo principų ir rodiklių taikymo

Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, Ismael Rafols 2015.

Bibliometrics: The Leiden Manifesto for research metrics ¹

iš *The Nature*, 2015, kovo 23, t. 520, p. 429–431, <https://doi.org/10.1038/520429a>

Diana Hicks, Paulas Woutersas, Ludo Waltmanas, Sarah de Rijcke ir Ismaelis Rafolsas ragina vadovautis dešimtimi principų vertinant mokslinius tyrimus

Mokslo vadybos srityje vis dažniau naudojami kiekybiniai duomenys. Mokslinių tyrimų vertinimas, anksčiau atliekamas tik pačių mokslininkų, dabar tapo su įvairiais rodikliais susijusiu formalumu [1]. Bėda ta, kad šiuolaikinis vertinimas dažnai grindžiamas ne sveiku protu, o skaičiais. Mokslo vertinimui skirtų rodiklių vis daugiau, ir nors jie randasi iš gerų ketinimų, ne visais atvejais pagrįstai parinkti, ir dažniausiai netinkamai taikomi. Rizikuojame pakenkti sistemai priemonėmis, sukurtomis ją gerinti, nes vertinimą vis dažniau atlieka organizacijos, neturinčios nei žinių apie gerąją praktiką, nei rekomendacijų dėl jos interpretavimo.

Iki 2000 m. buvo naudojamas Mokslinės informacijos instituto (angl. *Institute for Scientific Information*, trump. ISI) kompaktinėje plokštelėje išleistas mokslo citavimo indeksas (angl. *Science Citation Index*), kurį ekspertai taikė atlikdami specializuotas analizes. 2002 m. „Thomson Reuters“ pristatė integruotą saityno platformą, tokiu būdu plačiai atverdama prieigą prie „Web of Science“ duomenų bazės. Buvo sukurti konkuruojantys citavimo indeksai: „Elsevier“ pristatė „Scopus“ (2004 m.), o „Google“ – „Scholar“ (beta versija išleista 2004 m.). Parengtos saityno priemonės, leidžiančios nesunkiai palyginti organizacijų mokslinių tyrimų produktyvumą ir cituojamumą, kaip antai „InCities“ (naudojant „Web of Science“) bei „SciVal“ (naudojant „Scopus“), sukurta programinė įranga, kuria, remiantis „Google Scholar“ („Publish or Perish“, išleista 2007 m.), analizuojami individualūs citavimo duomenys.

2005 m. San Diege esančio Kalifornijos universiteto fizikas Jorge'as Hirschas, siekdamas populiarinti atskirų mokslininkų citavimo duomenų rinkimą, pasiūlė taikyti *h* indeksą. Nuo 1995 m. buvo vis labiau domimasi žurnalų cituojamumo rodikliu (žr. „Impact-factor obsession“).

Pastaruoju metu padaugėjo rodiklių, siejamų su turinio naudojimu ir komentavimu: 2002 m. sukurta „F1000Prime“ priemonė, 2008 m. – „Mendeley“, o 2011 m. – „Altmetric.com“ (remiama „Macmillan Science and Education“, kuriai priklauso „Nature Publishing Group“).

Mes – mokslometrijos specialistai, socialinių mokslų tyrėjai ir mokslo administratoriai – vis dažniau matome, kad mokslinių tyrimų rezultatai vertinami taikant netinkamus rodiklius. Pateikiame tik pora pavyzdžių iš daugelio: visų šalių universitetai iš paskutiniųjų stengiasi gerinti savo vietą pasauliniuose reitinguose (pavyzdžiui, Šanchajaus reitinge arba „Times“ aukštojo mokslo sąrašė) net kai, mūsų manymu, tokie sąrašai sudaromi remiantis netiksliais duomenimis ir abejotinais rodikliais.

Kai kurie darbdaviai remiasi kandidatų *h* indekso verte. Net keliuose universitetuose sprendimai dėl pareigų paaukštinimo priimami atsižvelgiant į ribines *h* indekso vertes bei straipsnių, paskelbtų didelį cituojamumo rodiklį turinčiuose žurnaluose, skaičių. Mokslininkai gyvenimo aprašymuose naudojami proga pasigirti tokiais rezultatais, ypač biomedicinos srityje. Praktiškai visur doktorantų vadovai reikalauja, kad doktorantai skelbtų straipsnius didelį cituojamumo rodiklį turinčiuose žurnaluose, o išorinį finansavimą gautų dar jų neparengę.

¹ LEIDEN MANIFESTO FOR RESEARCH METRICS <http://www.leidenmanifesto.org/>

Kai kuriuose Skandinavijos ir Kinijos universitetuose finansavimas moksliniams tyrimams arba premijos už juos skiriami remiantis kiekybiniu rodikliu, pavyzdžiui, suskaičiuavus autorių publikacijų cituojamumą, pagal kurį paskirstomas „skatinimas už veiklos rezultatus“ arba skiriant tyrėjams premijas už publikacijas žurnaluose, kurių cituojamumo rodiklis didesnis nei 15 [2].

Daugeliu atvejų tyrėjai ir jų vertintojai priima suderintą sprendimą. Tačiau kiekybiniais rodikliais pernelyg piktnaudžiaujama, kad būtų galima to nepaisyti.

Todėl skelbiame Leideno manifestą, taip pavadintą, nes buvo parengtas Leideno konferencijos metu (žr. <http://sti2014.cwts.nl>). Nors jame išdėstyti dešimt principų nėra nauji mokslometrijos specialistams, nė vienas nebūtume galėję jų tiksliai nurodyti, nes iki šiol dėl jų nebuvo sutarta. Įžymūs šios srities mokslininkai, pavyzdžiui, Eugene Garfield'as (ISI įkūrėjas), jau yra paskelbę kai kuriuos iš šių principų [3, 4]. Tačiau jų nebūna šalia, kai vertintojai atsiskaito universitetų vadovams, kurie nėra šios metodologijos specialistai.

Mokslininkai, ieškantys literatūros šaltinių, kuriais remdamiesi galėtų ginčytis dėl taikomo vertinimo, randa pasklidusios informacijos jiems menkai žinomuose ir nelengvai prieinamuose žurnaluose.

Siūlome šią gerąją rodikliais grįstą mokslinių tyrimų vertinimo praktiką, kuri mokslininkams padės patikrinti vertintojų atsakingumą, o vertintojams patikrinti, ar rodikliai tinkami.

Dešimt principų

1. Kokybinį ekspertinį vertinimą papildyti kiekybiniu vertinimu. Kiekybiniai rodikliai padės išvengti galimo ekspertų šališkumo ir palengvins svarstymus. Rodiklių paskirtis – sustiprinti ekspertinį vertinimą, kuris tampa nelengva užduotimi, kai tenka vertinti kolegas, todėl būtina naudoti įvairių tinkamą informaciją. Tačiau vertintojai privalo vengti priimti sprendimus, remdamiesi tik skaičiais. Kompetentingo sprendimo negali pakeisti rodikliai. Kiekvienas vertintojas yra atsakingas už savo sprendimą.

2. Veiklos rezultatus vertinti atsižvelgiant į mokslo institucijos, grupės arba tyrėjo misiją. Mokslo veiklos tikslai turėtų būti nurodyti pačioje pradžioje, o veiklos rezultatams vertinti skirti rodikliai – aiškiai susieti su šiais tikslais. Renkantis rodiklius ir jų taikymo būdus, patariama atsižvelgti į platesnį socialinį ir ekonominį kontekstą. Mokslininkai sprendžia įvairialypes užduotis. Tyrimas, kurio tikslas – plėsti akademinę žinių ribas, skiriasi nuo tyrimo, kuriuo siekiama spręsti socialines problemas. Ekspertinis vertinimas gali būti atliekamas remiantis ne tik akademinėmis idėjomis, kuriomis tobulinamas mokslas, bet ir turėtų būti atsižvelgiama į politikai, pramonei ar visuomenei svarbius pasiekimus. Nėra vieno vertinimo modelio, kuris būtų tinkamas visais atvejais.

3. Siekti kokybės atliekant vietinės reikšmės tyrimus. Daugelyje pasaulio vietovių mokslinių tyrimų kokybės ženklu laikomas jų rezultatų publikavimas anglų kalba. Pavyzdžiui, Ispanijos teisės aktuose nurodytas pageidavimas, kad šios valstybės mokslininkai skelbtų savo darbus didelį cituojamumą rodiklį turinčiuose žurnaluose. Skaičiuojamas dažniausiai anglų kalba leidžiamų žurnalų, indeksuojamų JAV veikiančios „Web of Science“, cituojamumo rodiklis. Toks šališkumas sukelia itin daug problemų socialinių ir humanitarinių mokslų srityse, kurių tyrimai dažniausiai būna regioniniai arba nacionaliniai. Daugelyje kitų sričių taip pat gali būti aktuali nacionalinė arba regioninė problema, pavyzdžiui, ŽIV epidemiologija Užsachario Afrikoje.

Tokio pliuralizmo ir aktualumo visuomenei linkstama nepaisyti, kai norima parengti straipsnius, kurie galėtų tikti didelį cituojamumą rodiklį turintiems anglų kalba leidžiamiems žurnalams. Pvz., Ispanijos sociologai, kurių darbai „Web of Science“ bazėse turi aukštą cituojamumą rodiklį, dirbo su abstrakčiais modeliais arba tyrinėjo JAV duomenis. Praradimu laikoma tai, kad didelį potencialą turintys sociologai neskelbia publikacijų gimtąja kalba tokiomis Ispanijai svarbiomis temomis, kaip nacionalinė darbo teisė, pagyvenusių žmonių sveikatos priežiūra arba imigrantų užimtumas. Rodikliai, skirti aukšto lygio, bet ne anglų kalba leidžiamiems žurnalams vertinti, padėtų nustatyti ir deramai pripažinti mokslininkus, atliekančius regionams reikšmingus tyrimus.

4. Užtikrinti atvirus, skaidrius ir paprastus duomenų rinkimo ir analizės procesus. Kuriant vertinimui skirtas duomenų bazes būtina laikytis aiškiai apibrėžtų taisyklių, nustatytų dar nepasibaigus tyrimams, kurie bus vertinami. Tokia praktika buvo paplitusi ir keletą dešimtmečių buvo taikoma akademinėse bendruomenėse ir komercinėse grupėse, kūrusiose mokslometrijos vertinimo metodikas. Šios grupės savo protokolus skelbė recenzuojamuose žurnaluose. Toks skaidrumas užtikrina galimybę nuodugniai įvertinti taikomus rodiklius. Pavyzdžiui, 2010 m. Leideno universiteto (Nyderlandai) Mokslo ir inovacijų studijų centre (CWTS) dėl kilusios viešos diskusijos dėl vienos grupės taikomo svarbaus rodiklio techninių ypatybių teko šį rodiklį perskaičiuoti [6]. Naujos komercinės organizacijos, siekiančios dalyvauti vertinimo procesuose, privalo laikytis tokių pačių standartų ir niekam neturėtų tikti neaiškus bei neskaidrus mechaniškas vertinimas. Skaidrumą didinantis paprastumas – kiekvieno rodiklio privalumas. Tačiau perdėm supaprastinti mokslo vertinimui skirti rodikliai gali iškreipti rezultatus (žr. 7 principą). Vertintojai privalo siekti pusiausvyros – paprastus rodiklius suderinti su sudėtingu mokslinio tyrimo procesu.

5. Leisti vertinamiesiems patikrinti duomenų ir metodikos teisingumą. Norint užtikrinti duomenų kokybę, visiems vertinimo dalyviams turėtų būti sudaryta galimybė patikrinti, ar visi jų darbai apskaityti ir įvertinti teisingai. Kiekvienas administratorius, valdantis vertinimo procesą, turėtų duomenų tikslumą užtikrinti naudodamas vertinamųjų savikontrolę arba trečiųjų šalių auditą. Universitetai tokias priemones turėtų įdiegti savo informacinėse sistemose, pritaikytose mokslinių tyrimų apskaitai, ir šie principai turėtų būti įtraukti į technines sąlygas renkantis šių sistemų tiekėjus/kūrėjus. Surinkti tikslius ir kokybiškus duomenis ir juos apdoroti kainuoja laiko ir pinigų. Tai būtina įvertinti planuojant biudžeto lėšas.

6. Atsižvelgti į skirtingoms sritims būdingus publikavimo ir citavimo praktikos skirtumus. Gerąją praktiką laikytinas siekis parengti galimų rodiklių rinkinį ir leisti skirtingų sričių atstovams iš jų pasirinkti tinkamiausius. Prieš keletą metų, atliekant nacionalinį ekspertinį vertinimą, Europos istorikų grupė (humanitariniai mokslai) buvo įvertinta palyginti žemu reitingu, kadangi skelbė knygas, o ne straipsnius „Web of Science“ indeksuojamuose žurnaluose. Mokslininkams-humanitarams nepasisėkė, nes jie dirbo psichologijos, t. y. socialiniams mokslams priskirtame padalinyje. Istorikams ir socialinių mokslų tyrėjams svarbu, kad į vertinamų publikacijų sąrašus būtų įtrauktos ne tik knygos, bet ir nacionaline kalba paskelbta literatūra; informatikos atstovai siekia, kad būtų vertinami pranešimai konferencijose.

Cituojamumas mokslo srityse nėra vienodas: geriausiai vertinamų ląstelės biologijos žurnalų cituojamumo rodikliai (angl. *Impact Factor*) svyruoja apie 30, o geriausiai vertinamų matematikos žurnalų cituojamumo rodikliai tesiekia 3. Dėl šios priežasties būtina naudoti normalizuotus rodiklius, o pats patikimiausias normalizavimo metodas yra paremtas procentinėmis dalimis: kiekvienai publikacijai priskiriamas koeficientas, priklausomai nuo procentinės dalies, kuriai ji priskiriama pagal šios srities citavimo skirstinį (pavyzdžiui, TOP1, TOP10 arba TOP20 priklauso 1, 10 arba 20 proc. geriausiųjų). Viena itin gausiai cituojama publikacija universiteto poziciją reitinge, kuris grindžiamas procentinių dalių rodikliais, gali pagerinti tik šiek tiek, tačiau reitinguose, kuriuose vertinami citavimo vidurkiai, universitetas iš vidutinių gali pakilti į lyderius [7].

7. Kiekvieno tyrėjo vertinimą grįsti kokybiniu jo darbų vertinimu. Kuo mokslininkas vyresnis, tuo jo *h* indeksas didesnis, net jei naujų darbų nepateikta. *h* indeksas priklauso nuo cituojamumo srityje: gyvybės mokslų tyrėjų didžiausias *h* rodiklis sudaro 200, fizikos – 100, o socialinių mokslų – 20 arba 30 [8]. To paties mokslininko *h* indekso dydis skirtingose duomenų bazėse bus nevienodas: yra informatikos tyrėjų, kurių *h* indeksas „Web of Science“ duomenų bazėje yra apie 10, o „Google Scholar“ – 20 arba 30 [9]. Perskaityti ir įvertinti mokslininko darbą kur kas patikimiau nei pasitikėti vienu skaičiumi. Net lyginant dideles mokslininkų grupes tinkamiausiu laikytinas vertinimo būdas, kurį taikant atsižvelgiama į tyrėjų kompetencijas, patirtį, veiklas ir jų poveikį.

8. Vengti netinkamo konkretumo ir apgaulingo tikslumo. Mokslo ir technologijų vertinimui skirtų rodiklių trūkumas – konceptualus dviprasmiškumas ir neapibrėžtumas, be to, reikalingos aiškios prielaidos, bet jos nėra visuotinai priimtose. Pavyzdžiui, vis dar diskutuojama dėl citavimų skaičiaus reikšmės. Todėl gerąją praktiką laikoma taikyti kelis rodiklius, nes tai padeda susidaryti aiškesnį vaizdą, pagrįstą

nuomonių įvairovė. Jei neapibrėžtumą ir paklaidas galima išreikšti kiekybiškai, pavyzdžiui, naudojant paklaidų stulpelius, ši informacija turėtų būti pateikiama kartu su skelbiamomis rodiklių vertėmis. Jei to padaryti neįmanoma, rodiklių kūrėjai ir taikytojai turėtų bent jau vengti apgaulingo tikslumo. Pavyzdžiui, žurnalo cituojamumo rodiklis skelbiamas dešimtaines trupmenas nurodant iki trijų skaičių po kablelio, kad būtų kuo mažesnė tikimybė, jog bus gauti vienodi rezultatai. Tačiau, atsižvelgiant į konceptualų dviprasmiškumą ir atsitiktinį citavimų skaičiaus kintamumą, nėra jokios prasmės skirstyti žurnalus remiantis labai mažais cituojamumo rodiklio skirtumais. Venkite apgaulingo tikslumo: tik viena dešimtoji trupmenos dalis laikytina reikšminga.

9. Atpažinti sisteminį vertinimo ir rodiklių poveikį. Rodikliai keičia sistemą per paskatinimus, kuriems jie yra skirti. Tokį poveikį sistemai būtina numatyti iš anksto. Tai reiškia, kad visais atvejais pirmenybė turėtų būti teikiama rodiklių deriniams, nes turint tik vieną skaičių, galimos manipuliacijos ir tikslų perkėlimas (kai tikslu tampa pats vertinimo procesas). Pavyzdžiui, 1990 m. Australijoje universitetuose atliekamų tyrimų finansavimas buvo pagrįstas institucijos paskelbtų publikacijų skaičiumi. Universitetai galėjo suskaičiuoti recenzuojamuose žurnaluose paskelbtos publikacijos „vertę“, ji 2000 m. sudarė apie 800 Australijos dolerių (arba apie 480 JAV dolerių). Kaip ir galima buvo numanyti, Australijos mokslininkų publikacijų skaičius išaugo, tačiau daugiausia jų buvo paskelbta mažiau cituojamuose žurnaluose, o tai leidžia daryti prielaidą, kad nukentėjo straipsnių kokybė [10].

10. Reguliariai ir kruopščiai tikrinti bei atnaujinti rodiklius. Mokslinių tyrimų paskirtis ir mokslo vertinimo tikslai keičiasi, tobulėja ir pati mokslo sistema. Aktualumą praradusius rodiklius keičia naujai sukurti. Taikomus rodiklius būtina peržiūrėti ir, galbūt, tikslinti. Australijos valdžia, suvokusi pernelyg supaprastintos formulės poveikį, 2010 m. pritaikė sudėtingesnę mokslo vertinimo metodiką, kurios esminis dėmesys teko kokybei.

Tolesni žingsniai

Šis dešimties principų mokslo vertinimas gali svariai prisidėti prie mokslo plėtros ir mokslo sąveikos su visuomene. Mokslo vertinimo rodikliai gali suteikti svarbios informacijos, kurios ekspertas, vadovaudamasis asmenine kompetencija, negalėtų surinkti arba suprasti. Tačiau toks kiekybinių rodiklių taikymas neturėtų iš priemonės virsti tikslu.

Geriausi sprendimai priimami derinant patikimus rodiklius su vertinamo tyrimo įgyvendinamais tikslais bei darbo pobūdžiu. Reikalingi ir kiekybiniai, ir kokybiniai duomenys, nes visi – savaip objektyvūs. Mokslo politikų sprendimai turi remtis kiekybiškai atliktu ir patikimais duomenimis pagrįstu mokslo vertinimu.

Literatūra

1. Wouters, P. in *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (eds Cronin, B. & Sugimoto, C, MIT Press.), 47–66. (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 Piñero, 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).