

Leidenský manifest k měření výzkumu

„Používejte při hodnocení výzkumu těchto deset principů,“ vyzývají **Diana Hicks, Paul Wouters** a kol.

Řízení vědy se stále více opírá o data. Hodnocení výzkumu, které bylo dříve příležitostné a prováděli jej mezi sebou kolegové, se dnes stalo rutinní a opírá se o různé metriky¹. Problém spočívá v tom, že hodnocení se dnes řídí spíše čísly než úsudkem. Měření je všudypřítomné. Většinou je dobře míněné, ne vždy vychází z relevantních informací, často je nesprávně použité. Hrozí, že poškodíme systém nástrojů, které byly navrženy za účelem jeho zdokonalení. Organizace totiž stále častěji zavádějí hodnocení, aniž by věděly nebo si nechaly poradit, jaká je jeho dobrá praxe a jak interpretovat jeho výsledky.

Před rokem 2000 používali experti pro odborné analýzy Science Citation Index na CD-ROMu, který vydával Institute for Scientific Information (ISI). V roce 2002 spustili Thomson Reuters integrovanou internetovou platformu, čímž zpřístupnili databázi Web of Science široké odborné veřejnosti. Následně byly vytvořeny konkurenční citační indexy: Scopus společnosti Elsevier (spuštěný v roce 2004) a Google Scholar (verze beta byla spuštěna v roce 2004). Vznikly internetové nástroje pro snadné srovnávání institucí dle produktivity a dopadu výzkumu, jako například InCites (využívající Web of Science) a SciVal (využívající Scopus), a rovněž software pro analýzu jednotlivých citačních profilů, využívající Google Scholar (Publish or Perish, spuštěný v roce 2007).

V roce 2005 navrhl fyzik z Kalifornské univerzity v San Diegu Jorge Hirsh tzv. H-index, který zpopularizoval sčítání citací jednotlivých výzkumníků. Od roku 1995 trvale vzrůstá zájem o impakt faktor časopisů (viz grafy „Posedlost impakt faktorem“).

Poslední dobou nabývá na významnosti měření referencí na sociálních sítích a online komentářů – v roce 2002 vzniká F1000Prime, v roce 2008 Mendeley a v roce 2011 Altmetric.com (za podpory nakladatelství Macmillan Science and Education, které je vlastníkem nakladatelství Nature Publishing Group).

Jako odborníci na scientometrii, sociální vědci a administrátoři výzkumu se vzrůstajícími obavami sledujeme všudypřítomné zavádějící užívání indikátorů při hodnocení vědecké výkonnosti. Níže uvádíme jen několik z mnoha příkladů. Univerzity na celém světě jsou posedlé svým umístěním v globálních žebříčcích (jako Šanghajský žebříček nebo žebříček portálu Times Higher Education), přestože tyto seznamy jsou dle našeho názoru založené na nepřesných datech a nahodilých indikátorech.

Některá pracoviště požadují po kandidátech jejich H-index. Řada univerzit rozhoduje o povýšení pracovníků na základě jejich H-indexu a počtu článků publikovaných v časopisech s vysokým impakt faktorem. Výzkumníci, zejména v oblasti biomedicíny, se těmito skóre chlubí ve svých životopisech. V podstatě všude chtějí školitelé po doktorandech, aby publikovali v časopisech s vysokým impakt faktorem a získali externí financování ještě před završením studia.

Ve Skandinávii a v Číně přidělují některé univerzity finance na výzkum nebo odměny na základě čísel: kupříkladu přidělují zvláštní zdroje na základě skóre impaktu nebo udělují výzkumníkům odměny za publikaci v časopisech s impakt faktorem vyšším než 15 (pozn. 2).

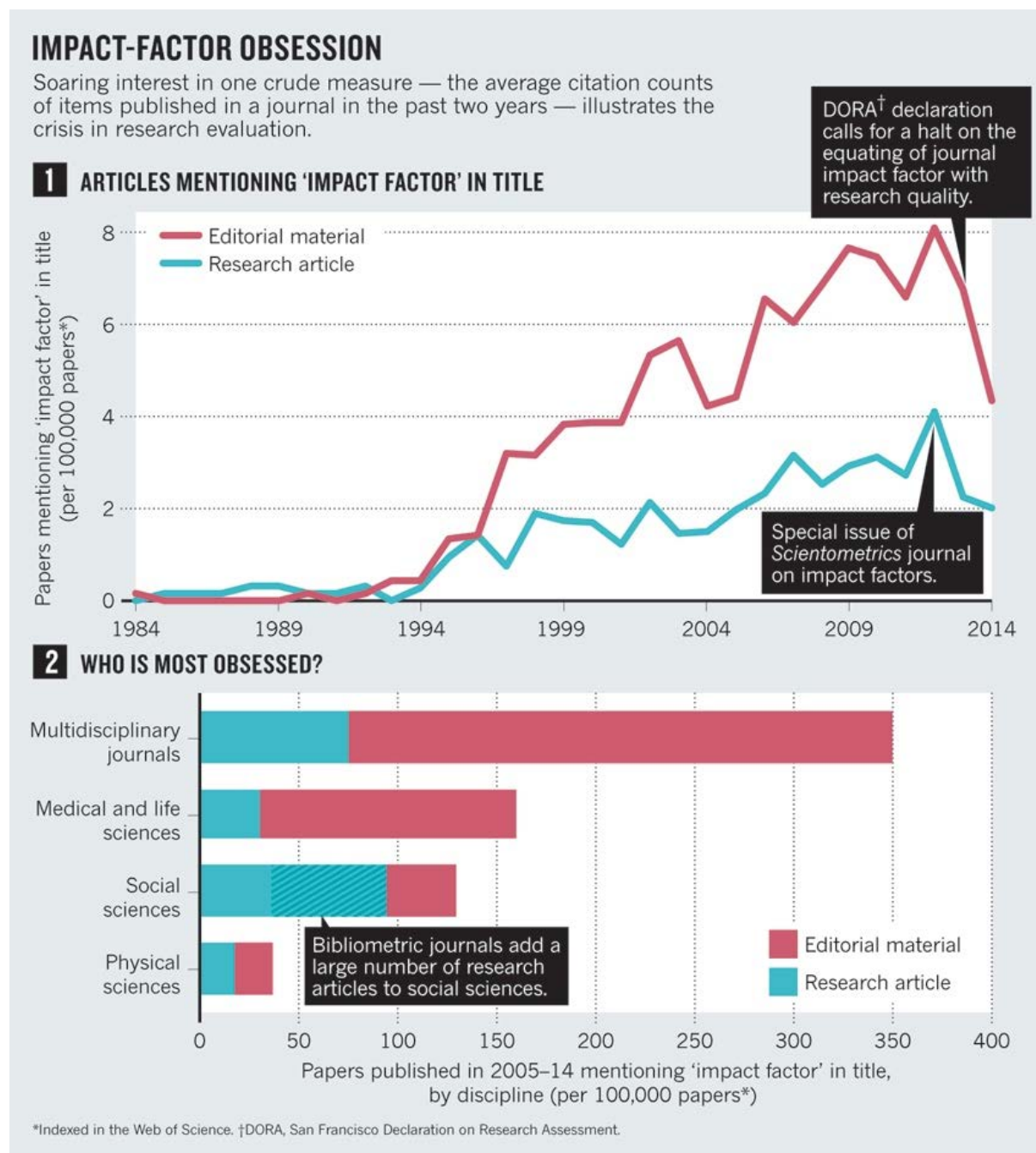
V mnoha případech výzkumníci a hodnotitelé stále posuzují výzkum vyváženě. Přesto se zneužívání měření výzkumu rozšířilo natolik, že ho nelze ignorovat.

Proto představujeme Leidenský manifest, nazvaný podle konference, na níž vznikl (viz <http://sti2014-cwts.nl>). Jeho deset principů není pro odborníky na scientometrii žádnou novinkou. Nikdo z nás by je však nebyl schopen vyjmenovat v celém jejich rozsahu, protože dodnes nebyly takto jednotně kodifikovány. Autority na tomto poli, jako například Eugene

Garfield (zakladatel ISI), některé z těchto principů v minulosti deklarovaly³⁴. Nemají už ale vliv na to, když hodnotitelé podávají zprávu představitelům univerzit, kteří nejsou odborníky na příslušnou metodologii. Vědci, kteří hledají publikace, pomocí níž by mohli hodnocení rozporovat, nacházejí materiál rozptýlený v časopisech z jejich pohledu obskurních, k nimž nemají přístup.

Přinášíme toto shrnutí dobré praxe v hodnocení výzkumu založeném na měření, aby mohli výzkumníci kriticky posoudit práci hodnotitelů a hodnotitelé zvážit vhodnost indikátorů.

Posedlost impakt faktorem:



DESET PRINCIPŮ

- 1. Kvantitativní hodnocení by mělo sloužit jako podpora kvalitativního, odborného posouzení.** Kvantitativní měření může zabránit zaujatosti při „peer review“ a pomoci při rozhodování. To by mělo „peer review“ posílit, protože posuzovat kolegy bez dostatku relevantních informací je obtížné. Hodnotitelé však nesmějí být v pokušení přenechat rozhodování číslům. Indikátory by neměly nahradit informované posuzování. Všichni nesou odpovědnost za svá hodnocení.
- 2. Měřte výkonnost ve vztahu k výzkumným cílům instituce, skupiny nebo výzkumníka.** Na začátku by měly být stanoveny programové cíle a indikátory použité k hodnocení výkonnosti by se měly k těmto cílům jasně vztahovat. Výběr indikátorů a způsoby jejich užití by měly zohledňovat širší socio-ekonomický a kulturní kontext. Vědci mají různé výzkumné cíle. Výzkum, který posouvá hranice akademického vědění, se liší od výzkumu zaměřeného na hledání řešení společenských problémů. Posuzování je tedy možné založit i na přínosu pro státní správu a samosprávu, průmysl nebo veřejnost spíše než na představách akademiků o vynikajícím výzkumu. Neexistuje model hodnocení použitelný ve všech kontextech.

- 3. Je třeba chránit vynikající výzkum regionálního významu.** V mnoha částech světa je vynikající výzkum ztotožňován s publikováním v anglickém jazyce. Například španělský zákon stanovuje jako žádoucí, aby španělští badatelé publikovali v časopisech s vysokým impakt faktorem. Impakt faktor se vypočítává pro časopisy zařazené na Web of Science, který má sídlo v USA a převažuje na něm angličtina. Takové zvýhodňování je zvláště problematické v sociálních a humanitních vědách, kde se výzkum více vztahuje k danému regionu či státu. Národní nebo regionální dimenzi má i mnoho dalších oborů – jako například epidemiologie HIV v subsaharské Africe.

Tento pluralismus a společenský význam výzkumu jsou potlačovány tvorbou článků, o které mají zájem časopisy s vysokým impakt faktorem – anglicky psané časopisy. Španělští sociologové, kteří pracují na abstraktních modelech nebo analyzují data z USA, dosahují na Web of Science vysoké citovanosti. Zapadnout specifika sociologů, kteří publikují ve vysoce vlivných španělsky psaných člancích o tématech jako místní pracovní právo, neformální zdravotní péče o seniory nebo zaměstnanost přistěhovalců.⁵ Měření zohledňující kvalitní neanglojazyčné publikace by pomohlo rozpoznat a ocenit vynikající výzkum regionálního významu.

- 4. Sběr a analýza dat by měly být otevřené, transparentní a jednoduché.** Tvorba databází potřebných pro hodnocení by se měla řídit jasnými a dopředu stanovenými pravidly. Po desítky let to byla běžná praxe akademiků a komerčních skupin, které se podílely na tvorbě metodologie bibliometrických hodnocení. Tyto skupiny se odkazovaly na metodiky publikované v recenzovaných časopisech. Taková transparentnost umožňovala pečlivý dohled. Například v roce 2010 vedla veřejná debata o technickém charakteru důležitého indikátoru používaného jednou z našich skupin (Centrum pro studium vědy a technologie Leidenské univerzity v Nizozemsku) ke změně výpočtu tohoto indikátoru⁶. Pro komerční subjekty, které v současnosti vstupují do sféry hodnocení vědy, by měla platit stejná kritéria. Nikdo by neměl akceptovat hodnocení, jehož mechanismus není transparentní.

Jednoduchost je předností indikátoru, protože umožňuje transparentnost. Zjednodušující měření však může zkreslit data (viz princip 7). Hodnotitelé musejí usilovat o rovnováhu – používat jednoduché indikátory, které vystihují složitost výzkumného procesu.

- 5. Ti, kteří jsou hodnoceni, by měli mít možnost ověřit data a analýzy.** V zájmu zajištění kvality dat by měli mít všichni výzkumníci zahrnutí v bibliometrických studiích možnost ověřit, že jejich výstupy byly správně zpracovány. Ti, kteří jsou zodpovědní za nastavování a řízení procesů hodnocení, by měli zajistit přesnost dat – kontrolu

mohou provádět sami hodnocení či třetí strana. Univerzity by mohly tento princip zahrnout do svých informačních systémů a zohlednit ho při výběru poskytovatelů těchto systémů. Třídění a zpracování přesných, kvalitních dat vyžaduje čas a peníze a mělo by se s nimi počítat v rozpočtu.

- 6. Je nutné zohledňovat rozdíly mezi obory v publikační a citační praxi.** Dobrou praxí je nabídnout jednotlivým oborům soubor možných indikátorů a umožnit jim, aby si z nich vybraly. Před několika lety získala skupina historiků v jedné evropské zemi při národním hodnocení poměrně nízké skóre. Běžně totiž publikovali monografie, nikoli články v časopisech zařazených na Web of Science. Měli tu smůlu, že byli součástí katedry psychologie. Pro historiky a sociální vědce je důležité, aby se mezi jejich výstupy počítaly i monografie a publikace v národním jazyce, pro informatiky jsou to zase příspěvky na konferencích.

Počty citací se liší dle oborů: přední matematické časopisy mají impakt faktor kolem 3, cytologické asi 30. Je třeba standardizovat indikátory. Nejspolehlivější metodou je standardizace na základě percentilů: každý článek by měl být hodnocen na základě percentilu, k němuž se řadí v citační distribuci ve svém oboru (například horní 1 %, 10 % nebo 20 %). V žebříčku založeném na percentilních indikátorech zlepšila jedna hojně citovaná publikace postavení univerzity jen nepatrně, ale v žebříčku založeném na citačním průměru ji může výrazně posunout ze středních příček k vrcholu⁷.

- 7. Hodnocení jednotlivých výzkumníků by mělo být založeno na kvalitativním posouzení jejich portfolií.** Čím jste starší, tím vyšší je váš H-index, i když nepublikujete nové články. H-index se liší podle oborů: špičkoví přírodovědci dosahují 200, fyzikové 100 a sociální vědci 20–30 (pozn. 8). Rovněž záleží na databázi: například někteří výzkumníci v informatice mají na Web of Science H-index kolem 10, kdežto na Google Scholar 20–30⁹. Je tedy vhodnější číst a posuzovat práci výzkumníka než spoléhat na jedno číslo. I při srovnávání velkého počtu výzkumníků je nejlepší zvolit takový přístup, který zohlední jejich individuální odborné znalosti, zkušenosti, aktivity a vliv v oboru.

- 8. Vyhněte se nemístné konkrétnosti a falešné přesnosti.** Indikátory popisující vědu a technologie jsou často pojmově nejednoznačné a nejasné a stojí na silných předpokladech, které nejsou všeobecně přijímané. Dlouhodobě se vede debata například o smyslu sčítání citací. K dobré praxi patří používat různé indikátory a poskytnout tak spolehlivější a komplexnější přehled. Je-li možné nejistoty a chyby vyjádřit kvantitativně, například pomocí intervalu spolehlivosti, měla by být tato informace uvedena spolu s hodnotou indikátoru. Pokud kvantifikace možná není, tvůrci indikátorů by se měli přinejmenším vyvarovat falešné přesnosti. Impakt faktor časopisu se například uvádí s přesností na tři desetinná místa, aby se předešlo stejným výsledkům. Avšak vzhledem k pojmové nejednoznačnosti a náhodné variabilitě při sčítání citací nemá smysl rozlišovat mezi časopisy na základě takto malých rozdílů v impakt faktoru. Vyhněte se falešné přesnosti: opodstatněné je jen jedno desetinné číslo.

- 9. Věnujte pozornost vlivu hodnocení a indikátorů na systém.** Indikátory mění systém prostřednictvím pobídek, které vytvářejí. Tyto dopady bychom měli být schopni předvídat. To znamená, že je vždy vhodnější používat soubor indikátorů – jediný indikátor povede ke snaze vyžrát na hodnocení a k tomu, že se cílem stane samotné měření. Například v devadesátých letech 20. století financovala Austrálie univerzitní výzkum pomocí vzorce, který byl založený převážně na počtu článků publikovaných jednotlivými instituty. Univerzity si tak mohly vypočítat „hodnotu“ jednoho článku v recenzovaném časopise – v roce 2000 to bylo 800 australských dolarů (480 amerických dolarů) na financování výzkumu. Dle očekávání se počet článků publikovaných australskými výzkumníky zvýšil. Články však byly publikovány v méně citovaných časopisech, což napovídá, že se snížila jejich kvalita¹⁰.

10. Indikátory by měly být pravidelně přezkoumávány a aktualizovány. Cíle výzkumu a cíle hodnocení se mění a vyvíjí se i samotný výzkumný systém. Kdysi užitečné nástroje měření přestávají dostačovat, objevují se nové. Systémy indikátorů je tedy třeba přezkoumávat a podle potřeby měnit. V Austrálii si dopady užívání zjednodušujícího vzorce uvědomili a v roce 2010 zavedli komplexnější systém Excellence in Research for Australia, který klade důraz na kvalitu.

DALŠÍ KROKY

Při dodržování těchto deseti principů může hrát hodnocení výzkumu důležitou roli v rozvoji vědy a jejího vztahu ke společnosti. Scientometrie může poskytnout zásadní informace, které by bylo obtížné shromažďovat nebo interpretovat na základě individuálních odborných znalostí. Kvantitativní informace se však nesmějí přeměnit z nástroje na cíl.

Nejlepší rozhodnutí vznikají spojením spolehlivých statistických údajů s citlivostí vůči cílům a povaze hodnoceného výzkumu. Zapotřebí jsou kvantitativní i kvalitativní data, oba druhy jsou svým způsobem objektivní. Rozhodování o vědě musí být založeno na kvalitních procesech, které pracují s kvalitními daty.

Diana Hicks je profesorka veřejné politiky na Georgijském technickém institutu v Atlantě ve státě Georgie v USA.

Paul Wouters je profesor scientometrie a ředitel, **Ludo Waltman** výzkumný pracovník a **Sarah de Rijcke** odborná asistentka Centra pro studia vědy a technologie Leidenské univerzity v Nizozemsku.

Ismael Rafols se věnuje výzkumu vědní politiky ve Španělské národní radě pro výzkum a na Polytechnické univerzitě ve Valencii ve Španělsku.

E-mail: diana.hicks@pubpolicy.gatech.edu

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).