

Il Manifesto di Leiden sulla bibliometria

Diana Hicks^a, Paul Wouters^b, Ludo Waltman^b, Sarah de Rijcke^b and Ismael Rafols^{c,d,e}

^a School of Public Policy, Georgia Institute of Technology, Atlanta, USA

^b Centre for Science and Technology Studies (CWTS), University of Leiden, The Netherlands

^c *Ingenio* (CSIC-UPV), Universitat Politècnica de València, València, Spain

^d Science Policy Research Unit (SPRU), University of Sussex, Brighton, UK

^e Observatoire des Science et des Techniques (OST-HCERES), Paris, France

(Traduzione in italiano di Hicks et al. (2015) The Leiden Manifesto for research metrics. *Nature*, 520, 429-431. <http://www.leidenmanifesto.org/>, ad opera di Edoardo Milotti, Università di Trieste)

I dati sull'attività scientifica vengono utilizzati sempre più per governare la scienza. Valutazioni della ricerca che una volta venivano realizzate a livello individuale ed eseguite da altri scienziati sono diventate una routine che si basa sulla bibliometria¹. Il problema è che adesso la valutazione è guidata dai dati invece che dal giudizio di esperti. Le metriche si sono moltiplicate: di solito con buone intenzioni ma non sempre con buone basi, e spesso male applicate. Poiché la valutazione viene sempre più affidata ad organizzazioni che non hanno una chiara idea delle buone pratiche e dell'interpretazione corretta dei dati, rischiamo di danneggiare il sistema proprio con gli strumenti che sono stati progettati per migliorarlo.

Prima del 2000 c'era il Science Citation Index su CD-ROM dell'Institute for Scientific Information (ISI), usato da esperti per analisi specialistiche. Nel 2002, Thomson-Reuters ha lanciato una piattaforma web integrata, rendendo così universalmente accessibile il database Web of Science. La competizione ha creato nuovi indici citazionali: Scopus della Elsevier (a partire dal 2004) e Google Scholar (la prima versione beta è del 2004). Sono stati introdotti strumenti di rete che confrontano agevolmente la produttività e l'impatto, come InCites (che utilizza Web of Science) e SciVal (che utilizza Scopus), così come software capace di analizzare i profili citazionali individuali per mezzo di Google Scholar (Publish or Perish, la cui prima versione è del 2007).

Nel 2005, Jorge Hirsch, un fisico dell'Università della California, San Diego, ha proposto l'*h*-index, diffondendo così il conteggio delle citazioni dei ricercatori individuali. L'interesse nell'impact factor dei giornali scientifici è cresciuto gradualmente dopo il 1995 (si veda l'insero "L'ossessione per l'impact factor").

Ultimamente, sono diventate popolari le metriche associate all'utilizzo social e ai commenti online – F1000Prime è stata fondata nel 2002, Mendeley nel 2008, e Altmetric.com (un prodotto di Macmillan Science and Education, che possiede il gruppo Nature) nel 2011.

¹ Wouters, P. in *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (eds Cronin, B. & Sugimoto, C.) 47–66 (MIT Press, 2014).

Come scienziati bibliometrici, sociologi e amministratori della ricerca scientifica, abbiamo osservato con crescente allarme la pervasiva e cattiva applicazione di indicatori per la valutazione della ricerca scientifica. Quelli che seguono sono solo alcuni esempi tra tanti. In tutto il mondo le università hanno sviluppato un'ossessione per il loro posizionamento nei rankings globali (come lo Shanghai Ranking e la lista di Times Higher Education), anche se queste liste sono basate, a nostra opinione, su dati inaccurati e indicatori arbitrari.

Alcuni reclutatori richiedono gli *h*-index dei candidati. Parecchie università basano le promozioni su valori soglia di *h*-index e sul numero di articoli in giornali "ad alto impatto". I CV dei ricercatori sono diventati delle opportunità per vantarsi di questi valori, soprattutto in biomedicina. Ovunque, i supervisori chiedono agli studenti di dottorato di pubblicare in giornali ad alto impatto e ad acquisire fondi esterni prima di essere pronti a farlo.

In Scandinavia ed in Cina, alcune università decidono la distribuzione dei fondi per la ricerca o degli incentivi sulla base di un numero: ad esempio, calcolando i punteggi di impatto individuale per distribuire risorse o dando un bonus ai ricercatori che pubblicano su un giornale scientifico con impact factor maggiore di 15.²

In molti casi, ricercatori e valutatori giudicano ancora in modo bilanciato. E tuttavia, l'abuso della bibliometria si è diffuso troppo per poterlo ignorare.

Per questo noi presentiamo il Manifesto di Leiden, che prende il nome dalla conferenza in cui è stato proposto (si veda <http://sti2014.cwts.nl>). I suoi dieci principi non sono nuovi per gli scienziati bibliometrici, benché nessuno di noi sarebbe capace di declamarli nella loro interezza perché finora non erano stati codificati. Si sa che luminari del campo come Eugene Garfield (fondatore dell'ISI) ne hanno pronunciati alcuni.³ Ma questi luminari non accompagnano i valutatori che fanno rapporto ad amministratori di università senza esperienza metodologica. E gli scienziati che cercano riferimenti bibliografici con i quali contestare una valutazione scoprono che il materiale è sparso in oscuri giornali a cui loro non hanno accesso.

Noi offriamo questo distillato di buone pratiche per la valutazione basata sulla bibliometria in modo che i ricercatori possano chiedere conto del loro operato ai valutatori, e i valutatori possano responsabilmente utilizzare gli indicatori bibliometrici.

DIECI PRINCIPI

1. La valutazione quantitativa dovrebbe aiutare la valutazione qualitativa degli esperti. Le metriche quantitative possono mettere in luce pregiudizi nella peer review e facilitare le decisioni. Questo dovrebbe rafforzare la peer review, perché giudicare i colleghi è difficile senza avere informazioni utili. Ma i valutatori non devono cedere alla tentazione di affidare la decisione ai numeri. Gli indicatori non devono sostituire il giudizio informato. Ciascuno resta responsabile della propria valutazione.

² Shao, J. & Shen, H. *Learned Publ.* **24**, 95–97 (2011).

³ Seglen, P. O. *Br. Med. J.* **314**, 498–502 (1997); Garfield, E. *J. Am. Med. Assoc.* **295**, 90–93 (2006).

2. Una misura della performance deve rapportarsi alla missione dell'istituzione, del gruppo o del singolo ricercatore. Si dovrebbero chiarire fin dall'inizio gli scopi programmatici, e gli indicatori utilizzati dovrebbero essere ben correlati a questi scopi. La scelta degli indicatori, e i modi in cui sono usati, dovrebbe tenere conto del più vasto contesto socio-economico e culturale. La ricerca che fa avanzare le frontiere della conoscenza accademica differisce dalla ricerca focalizzata sulla soluzione dei problemi sociali. Il giudizio potrebbe essere basato su meriti rilevanti per gli indirizzi politici, industriali o pubblici piuttosto che sull'idea di eccellenza accademica. Non c'è un solo modello valutativo che si applica a tutti i contesti.

3. Proteggete l'eccellenza nella ricerca che ha rilevanza locale. In molte parti del mondo l'eccellenza nella ricerca viene identificata con la pubblicazione in lingua inglese. La legge spagnola, per esempio, auspica che gli studiosi spagnoli pubblichino su giornali ad alto impatto. L'impact factor viene calcolato per i giornali indicizzati da Web of Science, che è basata negli Stati Uniti ed è per lo più anglofona. Questi aspetti pregiudiziali sono particolarmente problematici nelle scienze sociali e nelle scienze umane, in cui la ricerca ha maggiori aspetti regionali e nazionali. Molti altri campi hanno una dimensione nazionale o regionale – per esempio, l'epidemiologia dell'HIV nell'Africa sub-Sahariana.

Questo pluralismo e questa rilevanza sociale tendono ad essere soppressi per creare articoli che siano di interesse per i guardiani dell'impact factor: i giornali in lingua inglese. I sociologi spagnoli che hanno un alto numero di citazioni in Web of Science sono quelli che lavorano su modelli astratti oppure studiano dati provenienti dagli Stati Uniti. Si perde così la specificità dei sociologi che producono articoli ad alto impatto in spagnolo: argomenti come la legge locale sul lavoro, la rete sanitaria familiare per gli anziani, o il livello di occupazione degli immigrati.⁴ Delle metriche costruite per la letteratura scientifica non in inglese servirebbero a identificare e dare riconoscimento all'eccellenza nella ricerca che ha un rilievo locale.

4. Mantenete la raccolta dati e il processo analitico aperto, trasparente e semplice. La costruzione delle banche dati che sono richieste per la valutazione dovrebbe seguire regole stabilite con chiarezza, definite prima del completamento dell'analisi. Questa è stata per decenni la pratica usuale tra gli accademici e i gruppi commerciali che hanno costruito la metodologia di valutazione bibliometrica. Questi gruppi hanno fatto riferimento a protocolli pubblicati nella letteratura scientifica. Questa trasparenza permetteva il controllo incrociato. Per esempio, nel 2010 il dibattito pubblico sulle proprietà tecniche di un importante indicatore utilizzato da uno dei nostri gruppi (il Centre for Science and Technology Studies presso l'Università di Leiden in Olanda) ha portato ad una revisione nel calcolo di questo indicatore.⁵ Ai nuovi arrivi tra i gruppi commerciali dovrebbe venire richiesto di attenersi allo stesso standard; nessuno dovrebbe accettare una scatola nera come macchina per la valutazione.

La semplicità è una virtù in un indicatore perché aumenta la trasparenza. Ma metriche semplicistiche possono distorcere il dato (si veda il principio 7). I valutatori devono darsi da

⁴ López Piñeiro, C. & Hicks, D. *Res. Eval.* **24**, 78–89 (2015).

⁵ van Raan, A. F. J., van Leeuwen, T. N., Visser, M. S., van Eck, N. J. & Waltman, L. *J. Informetrics* **4**, 431–435 (2010).

fare per arrivare ad un giusto bilanciamento – indicatori semplici che non sminuiscano la complessità del processo di ricerca.

5. Permettete ai soggetti valutati di verificare i dati e l'analisi. Per assicurare la qualità dei dati, tutti i ricercatori inclusi negli studi bibliometrici dovrebbero poter controllare che la loro produzione scientifica sia stata correttamente identificata. Tutti quelli che dirigono e si occupano del processo di valutazione dovrebbero assicurare l'accuratezza dei dati, tramite una auto-verifica o con un audit di un soggetto esterno. Le università potrebbero implementare questo aspetto nei loro sistemi informativi e dovrebbe diventare un principio guida nella selezione dei providers di questi sistemi. Dati accurati, di alta qualità richiedono tempo e denaro per la cura e l'analisi. Nei vostri bilanci, assegnate dei fondi dedicati a questo scopo.

6. Tenete conto delle specificità tipiche dei campi di ricerca nelle pratiche di pubblicazione e di citazione. La migliore pratica consiste nel selezionare un insieme di possibili indicatori e di consentire a ciascun campo di scegliere quelli che preferiscono. Alcuni anni fa un gruppo di storici europei ha ricevuto un punteggio relativamente basso in una valutazione nazionale perché scrivevano libri invece di articoli indicizzati da Web of Science. Gli storici avevano la sfortuna di fare parte di un dipartimento di psicologia. Gli storici e i sociologi hanno bisogno di libri e di opere nella lingua nazionale per essere inclusi nei conteggi delle loro pubblicazioni; gli informatici vogliono poter contare i contributi a conferenze.

Il numero di citazioni varia da campo a campo: i giornali più prestigiosi in matematica hanno un impact factor vicino a 3; quelli di maggior prestigio in biologia cellulare hanno impact factors di circa 30. C'è bisogno di indicatori normalizzati, e il metodo più robusto di normalizzazione utilizza i percentili; ogni articolo è pesato sulla base del percentile al quale appartiene nella distribuzione di citazioni del campo a cui si applica (il top 1%, 10% o 20%, ad esempio). Una singola pubblicazione molto citata migliora leggermente la posizione di un'università in una graduatoria basata sui percentili, ma potrebbe spingerla fino in cima da una posizione a metà graduatoria in un ranking basato sulle medie citazionali.⁶

7. Basate la valutazione dei singoli ricercatori su un giudizio qualitativo del loro portfolio. Più si è anziani e più è alto l'*h*-index, anche in assenza di nuovi articoli. L'*h*-index varia secondo il campo: nelle scienze della vita arriva a 200; i fisici arrivano a 100 e i sociologi a 20-30.⁷ L'*h*-index dipende dal database: ci sono ricercatori informatici che hanno un indice di circa 10 in Web of Science ma di 20-30 in Google Scholar.⁸ Leggere e giudicare il lavoro di un ricercatore è molto più appropriato che affidarsi ad un numero. Anche quando si confrontano grandi numeri di ricercatori un approccio è migliore se considera più informazione sulle conoscenze, le esperienze, le attività e l'autorevolezza di un individuo.

8. Evitate la concretezza mal riposta e la falsa precisione. Gli indicatori scientifici e tecnologici possono essere concettualmente ambigui e incerti e richiedere ipotesi forti che non sono universalmente accettate. Il significato dei conteggi citazionali, per esempio, è stato dibattuto

⁶ Waltman, L. et al. *J. Am. Soc. Inf. Sci. Technol.* **63**, 2419–2432 (2012).

⁷ Hirsch, J. E. *Proc. Natl. Acad. Sci. USA* **102**, 16569–16572 (2005).

⁸ Bar-Ilan, J. *Scientometrics* **74**, 257–271 (2008).

a lungo. Perciò è bene utilizzare molteplici indicatori per produrre un'immagine più robusta e completa. Se è possibile quantificare incertezza ed errore, per esempio usando grafici con barre d'errore, questa informazione dovrebbe accompagnare i valori pubblicati degli indicatori. Se questo non è possibile, chi produce gli indicatori dovrebbe almeno evitare una falsa precisione. Per esempio, l'impact factor di un giornale scientifico viene pubblicato con tre decimali per evitare situazioni di pareggio. Però, data l'ambiguità concettuale e la variabilità casuale dei conteggi citazionali, non ha senso distinguere tra giornali sulla base di differenze molto piccole di impact factor. Evitate la falsa precisione: solo l'uso di un singolo decimale è giustificato.

9. Riconoscete gli effetti di sistema prodotti dalla valutazione e dagli indicatori. Gli indicatori modificano il sistema stabilendo un sistema di incentivi. Questi effetti dovrebbero venire previsti. Questo significa che è sempre meglio definire un insieme di indicatori – un solo indicatore è un invito a elaborare strategie e a modificare gli obiettivi (l'indicatore diventa lo scopo). Per esempio, negli anni '90, l'Australia finanziava la ricerca usando una formula basata principalmente sul numero di articoli pubblicati da un istituto. Le università potevano calcolare il "valore" di un articolo pubblicato in un giornale scientifico; nel 2000 questo corrispondeva a 800 dollari australiani in fondi di ricerca (circa 480 US\$ nel 2000). Come ci si poteva immaginare, il numero di articoli pubblicati da ricercatori australiani è aumentato, ma per lo più erano in giornali poco citati, e questo suggerisce una diminuzione qualitativa.⁹

10. Revisionate regolarmente gli indicatori e aggiornateli. La missione della ricerca e gli scopi della valutazione cambiano nel tempo e il sistema della ricerca co-evolve. Gli indicatori che una volta erano stati utili diventano inadeguati; compaiono nuovi indicatori. Bisogna passare in rassegna i sistemi di indicatori e forse modificarli. Dopo avere compreso gli effetti della sua formula semplicistica, nel 2010 l'Australia ha introdotto una nuova e più articolata iniziativa, Excellence in Research for Australia, che mette l'accento sulla qualità.

I PROSSIMI PASSI

Se si attiene a questi dieci principi, la valutazione della ricerca può giocare un ruolo importante nello sviluppo della scienza e delle sue interazioni con la società. La bibliometria può dare informazioni cruciali che sarebbe difficile raccogliere o capire per mezzo dell'esperienza individuale. Ma non si deve permettere che questa informazione quantitativa si trasformi da strumento ad obiettivo della valutazione.

Le migliori decisioni si prendono combinando statistiche robuste con la sensibilità agli scopi e alla natura della ricerca che viene valutata. C'è bisogno di dati sia quantitativi che qualitativi; ciascuno di essi ha una sua forma di obiettività. Le decisioni sulla scienza si devono basare su processi di alta qualità informati dai dati della più alta qualità.

⁹ Butler, L. *Res. Policy* **32**, 143–155 (2003).

Diana Hicks è professore di politiche pubbliche presso il Georgia Institute of Technology, Atlanta, Georgia, USA. **Paul Wouters** è professore di bibliometria e direttore, Ludo Waltman è un ricercatore, e **Sarah de Rijcke** è professore aggregato presso il Centre for Science and Technology Studies, Leiden University, Olanda. **Ismael Rafols** è un ricercatore di politiche scientifiche nel Consiglio Nazionale delle Ricerche spagnolo e presso l'Università Politecnica di Valencia, Spagna.

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

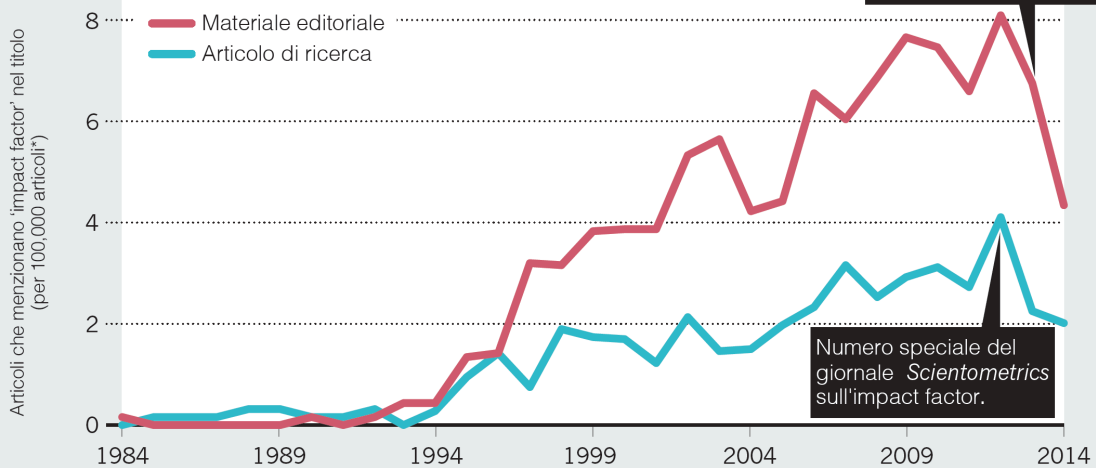
DATA SOURCE: THOMSON REUTERS WEB OF SCIENCE; ANALYSIS: D.H., L.W.

L'OSSESSIONE PER L'IMPACT FACTOR

Il crescente interesse per una misura grossolana — il conteggio medio delle citazioni degli articoli pubblicati in un giornale negli ultimi due anni — illustra la crisi nella valutazione della ricerca.

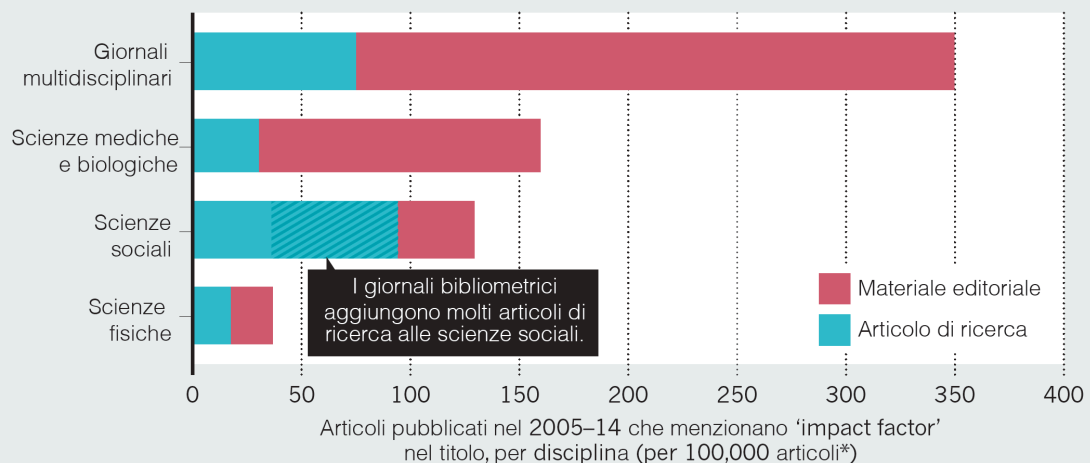
La dichiarazione DORA† chiede che si smetta di utilizzare l'impact factor di un giornale come indice di qualità della ricerca.

1 ARTICOLI CHE MENZIONANO 'IMPACT FACTOR' NEL TITOLO



Numero speciale del giornale *Scientometrics* sull'impact factor.

2 CHI E' PIU' OSSESSIONATO?



I giornali bibliometrici aggiungono molti articoli di ricerca alle scienze sociali.

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