

## **Bibliometrik: Manifesto Leiden untuk metrik penelitian**

Diana Hicks<sup>a</sup>, Paul Wouters<sup>b</sup>, Ludo Waltman<sup>b</sup>, Sarah de Rijcke<sup>b</sup> and Ismael Rafols<sup>cde</sup>

<sup>a</sup> School of Public Policy, Georgia Institute of Technology, Atlanta, AS

<sup>b</sup> Centre for Science and Technology Studies (CWTS), University of Leiden, Belanda

<sup>c</sup> Ingenio (CSIC-UPV), Universitat Politècnica de València, València, Spanyol

<sup>d</sup> Science Policy Research Unit (SPRU), University of Sussex, Brighton, Inggris Raya

<sup>e</sup> Observatoire des Science et des Techniques (OST-HCERES), Paris, Perancis

(diterjemahkan oleh Surya Dalimunthe dan Afrilya dari Tim Sains Terbuka Indonesia)

Data semakin banyak digunakan untuk mengendalikan ilmu pengetahuan. Penilaian penelitian yang dulu dilangsungkan sesekali oleh rekan sejawat telah menjadi rutin dan bergantung pada bibliometrik (1). Masalahnya adalah penilaian sekarang lebih ditentukan oleh data bibliometrik daripada pertimbangan sejawat. Bibliometrik mewabah: seringkali untuk tujuan baik, namun tidak selalu berdasar kuat, sehingga sering diterapkan salah. Kita berpotensi merusak sistem dengan alat yang diciptakan untuk memperbaikinya, seiring kebijakan penilaian yang diterapkan oleh berbagai organisasi tanpa pengetahuan atau petunjuk tentang praktik dan pemahaman yang baik.

Sebelum tahun 2000, ada Indeks Pengutipan Sains (Science Citation Index) dalam bentuk CD-ROM dari Institute for Scientific Information (ISI), yang digunakan oleh para pakar untuk menganalisis sitasi. Pada tahun 2002, Thomson Reuters meluncurkan sebuah platform web terintegrasi, yang menjadikan pangkalan data Web of Science dapat diakses secara luas. Kemudian, indeks pengutipan tandingan bermunculan: Scopus dari Elsevier (dirilis pada tahun 2004) dan Google Scholar (versi beta dirilis pada tahun 2004). Alat berbasis web untuk membandingkan dengan mudah produktivitas dan dampak penelitian lalu muncul juga, seperti InCites (berbasis Web of Science) dan SciVal (berbasis Scopus), beserta perangkat lunak untuk menganalisis profil pengutipan individual melalui Google Scholar (Publish or Perish, dirilis pada tahun 2007).

Pada tahun 2005, Jorge Hirsch, seorang fisikawan dari Universitas California, San Diego, menciptakan indeks-h, yang mempopulerkan penghitungan kutipan di tingkat individu peneliti. Minat terhadap faktor dampak jurnal juga terus meningkat sejak tahun 1995 (lihat 'Impact-factor obsession').

Akhir-akhir ini, bibliometrik yang terkait dengan penyebaran sosial dan komentar daring juga mendapatkan momentumnya - F1000Prime didirikan pada tahun 2002, Mendeley pada tahun 2008, dan Altmetric.com (didukung oleh Macmillan Science and Education, induk perusahaan Nature Publishing Group) pada tahun 2011.

Sebagai ilmuwan bibliometrik, ilmuwan sosial, dan administrator penelitian, kami telah menyaksikan dengan gelisah kesalahan penerapan indikator yang meluas untuk menilai kinerja ilmiah. Berikut ini hanya beberapa contoh saja. Di seluruh dunia, universitas telah menjadi terobsesi dengan posisi mereka dalam peringkat dunia (seperti Shanghai Ranking dan daftar Times Higher Education), bahkan ketika

peringkat ini didasarkan pada, dalam pandangan kami, data yang tidak akurat dan indikator yang sewenang-wenang.

Beberapa perekrut meminta nilai indeks-h untuk kandidat yang ingin mereka pekerjakan. Beberapa universitas mendasarkan penilaian mereka pada nilai indeks-h tertentu dan jumlah artikel yang terbit dalam jurnal 'berdampak tinggi'. Dokumen riwayat hidup para ilmuwan menjadi tempat untuk memamerkan angka-angka ini, terutama dalam bidang biomedis. Di mana-mana, para pembimbing meminta mahasiswa doktoral untuk menerbitkan di jurnal berdampak tinggi dan mendapatkan pendanaan eksternal sebelum mereka siap.

Di Skandinavia dan China, beberapa universitas mengalokasikan dana penelitian atau bonus berdasarkan penilaian kuantitatif: misalnya, dengan menghitung skor dampak individu untuk mengalokasikan 'sumber daya berbasis kinerja' atau dengan memberi peneliti bonus untuk penerbitan dalam jurnal dengan faktor dampak yang lebih tinggi dari 15 (ref. 2).

Dalam banyak kasus, peneliti dan penilai masih menggunakan metode penilaian yang seimbang. Namun penyalahgunaan bibliometrik telah terlalu meluas untuk bisa diabaikan begitu saja.

Oleh itu, kami mengajukan Manifesto Leiden, yang dinamai berdasarkan tempat konferensi pendiskusiannya (lihat <http://sti2014.cwts.nl>). Sepuluh prinsipnya bukanlah hal yang baru untuk para ilmuwan bibliometrik, walaupun tidak satu pun dari kami yang bisa menghafal semua prinsipnya karena barunya penyusunan prinsip-prinsip ini. Para tokoh dalam bidang ini, seperti Eugene Garfield (pendiri ISI), tercatat turut menyuarakan sebagian prinsip-prinsip ini (3, 4). Tetapi mereka tidak bisa hadir ketika para penilai memberikan kepada para administrator kampus yang bukan pakar dalam bidang ini. Para ilmuwan yang mencari literatur terdahulu untuk mempertanyakan sebuah penilaian menemukan bahwa literatur-literatur relevan tersebar dalam apa yang, bagi mereka, adalah jurnal-jurnal tersembunyi yang tidak bisa diakses.

Kami mengajukan kumpulan praktik-praktik terbaik untuk penilaian berbasis bibliometrik ini agar para ilmuwan dapat meminta pertanggungjawaban para penilai, dan para penilai dapat mempertanggungjawabkan indikator-indikator mereka.

## **Sepuluh prinsip**

### **1) Penilaian kuantitatif harus mendukung penilaian kualitatif para pakar.**

Bibliometrik kuantitatif dapat mencegah kecenderungan bias dan membantu pertimbangan dalam peninjauan sejawat. Seharusnya ia memperkuat penilaian sejawat, karena sulit membuat penilaian sejawat tanpa berbagai informasi relevan. Namun, penilai tidak boleh tergoda untuk menggantikan penilaian dengan angka-angka saja. Indikator tidak boleh menggantikan pertimbangan menyeluruh. Setiap orang harus bertanggung jawab atas penilaian mereka.

### **2) Ukur kinerja berdasarkan misi penelitian lembaga, kelompok atau peneliti.**

Tujuan program harus dinyatakan sejak awal, dan indikator yang digunakan untuk menilai kinerja harus berhubungan jelas dengan tujuan tersebut. Pilihan indikator, dan cara penggunaannya, harus mempertimbangkan konteks sosio-ekonomi dan budaya yang lebih luas. Para ilmuwan memiliki beragam misi penelitian. Penelitian yang memajukan batas ilmu pengetahuan berbeda dengan penelitian yang berfokus pada solusi untuk masalah-masalah masyarakat. Peninjauan dapat didasarkan pada manfaat yang relevan dengan kebijakan, industri atau publik daripada dibatasi hanya terhadap keunggulan akademis. Tidak ada model penilaian tunggal yang berlaku pada semua konteks.

### **3) Melindungi keunggulan penelitian lokal.**

Di banyak bagian dunia, keunggulan penelitian disamakan dengan terbitan berbahasa Inggris. Undang-undang Spanyol, misalnya, menyatakan bahwa para ilmuwan Spanyol sebaiknya menerbitkan makalah di jurnal berdampak tinggi. Faktor dampak jurnal dihitung berdasarkan Web of Science yang cenderung berbahasa Inggris dan berbasis di Amerika. Bias-bias ini sangat bermasalah terutama dalam bidang ilmu sosial dan humaniora, di mana penelitian lebih bersifat regional dan nasional. Banyak bidang lain juga memiliki dimensi nasional atau regional - contohnya, epidemiologi HIV di Afrika sub-Sahara.

Kemajemukan dan relevansi sosial ini cenderung dikesampingkan untuk menciptakan makalah yang menarik bagi punggawa faktor dampak: jurnal berbahasa Inggris. Sosiolog Spanyol yang banyak dikutip di Web of Science jadi menulis tentang model abstrak atau meneliti tentang data Amerika. Hilang jadinya kekhususan sosiolog yang menulis makalah di jurnal berbahasa Spanyol yang berdampak tinggi: topik seperti undang-undang tenaga kerja lokal dan perawatan kesehatan keluarga untuk orang tua atau imigran (5). Bibliometrik yang dibangun dengan literatur non-bahasa Inggris yang berkualitas tinggi bisa berfungsi untuk menandai dan menghargai keunggulan penelitian lokal.

### **4) Pastikan pengumpulan data dan proses analitik terbuka, transparan dan sederhana.**

Penciptaan pangkalan data untuk penilaian harus mengikuti aturan yang dinyatakan secara jelas dan ditetapkan sebelum penelitian selesai. Ini adalah praktik umum di kalangan kelompok akademis dan komersial yang membangun metodologi penilaian bibliometrik selama beberapa dekade terakhir. Kelompok-kelompok ini merujuk pada protokol yang diterbitkan dalam literatur ilmiah. Transparansi ini memungkinkan pengawasan dan perbandingan. Misalnya, pada tahun 2010, debat publik tentang sifat teknis dari indikator penting yang digunakan salah satu kelompok kami (Centre for Science and Technology Studies di Universitas Leiden, Belanda) menyebabkan adanya revisi dalam penghitungan indikator ini (6). Pendetak komersial baru harus merujuk pada standar yang sama; tidak seorangpun harus menerima metode penilaian yang tertutup.

Kesederhanaan penting dalam sebuah indikator karena ia meningkatkan transparansi. Tetapi bibliometrik simplistik dapat mendistorsi penilaian (lihat prinsip 7). Penilai harus berusaha mencapai keseimbangan - indikator sederhana sesuai dengan kompleksitas proses penelitian.

## **5) Mengizinkan orang-orang yang dinilai untuk memverifikasi data dan analisis.**

Untuk memastikan kualitas data, semua peneliti yang dinilai secara bibliometrik harus dapat memeriksa apakah karya mereka telah diidentifikasi dengan benar. Semua orang yang menjalankan proses penilaian harus memastikan keakuratan data, melalui verifikasi sendiri atau audit pihak ketiga. Universitas dapat menerapkan prinsip ini dalam sistem informasi penelitian mereka dan prinsip ini harus menjadi panduan dalam pemilihan penyedia sistem informasi tersebut. Data yang akurat dan data berkualitas tinggi memerlukan waktu dan uang. Anggaran harus disediakan.

## **6) Pertimbangkan variasi kutipan dan publikasi berdasarkan bidang studi.**

Praktik terbaik adalah memilih sekumpulan indikator dan mengizinkan masing-masing bidang studi untuk memilih indikator yang sesuai. Beberapa tahun yang lalu, sekelompok sejarawan Eropa menerima penilaian yang relatif rendah di tingkat nasional karena mereka menulis buku, bukan artikel dalam jurnal yang diindeks oleh Web of Science. Para sejarawan ini kurang beruntung karena menjadi bagian dari departemen psikologi. Sejarawan dan ilmuwan sosial membutuhkan buku dan literatur dalam bahasa nasional untuk dimasukkan dalam hitungan jumlah publikasi mereka; ilmuwan komputer membutuhkan makalah konferensi untuk dimasukkan dalam hitungan ini.

Tingkat kutipan bervariasi berdasarkan bidang studi: jurnal peringkat teratas dalam bidang studi matematika berfaktor dampak sekitar 3; dalam biologi selular berfaktor dampak sekitar 30. Indikator dan metode yang sudah dinormalisasi diperlukan, dan normalisasi paling tepat didasarkan pada persentil: setiap makalah dihitung berdasarkan persentil dalam distribusi kutipan bidangnya (1%, 10% atau 20% teratas, misalnya). Satu publikasi yang banyak dikutip dapat meningkatkan posisi universitas dalam peringkat yang didasarkan pada indikator persentil, dan mendorong universitas tersebut dari peringkat menengah ke atas ranking yang dibangun dengan rerata kutipan (7).

## **7) Nilai peneliti individu berdasarkan pertimbangan kualitatif terhadap portofolio mereka.**

Semakin tua usia seorang peneliti, semakin tinggi indeks-h nya, walaupun tanpa publikasi baru. Indeks-h bervariasi berdasarkan bidang studi: ilmuwan biologi maksimal 200; fisikawan maksimal 100 dan ilmuwan sosial maksimal 20-30 (8). Ia juga tergantung pada pangkalan data: ada peneliti dalam bidang studi ilmu komputer yang memiliki indeks-h sekitar 10 pada Web of Science namun sekitar 20-30 pada Google Scholar (9). Membaca dan menilai karya seorang peneliti jauh lebih tepat daripada bergantung pada angka. Bahkan ketika membandingkan peneliti dalam jumlah banyak, pendekatan penilaian yang mempertimbangkan secara menyeluruh keahlian, pengalaman, aktivitas, dan pengaruh individu adalah yang terbaik.

## **8) Hindari kepastian yang tidak pada tempatnya dan presisi yang salah.**

Indikator sains dan teknologi rentan terhadap ambiguitas dan ketidakpastian konseptual dan membutuhkan asumsi kuat yang belum tentu diterima secara universal. Makna hitungan kutipan, misalnya, telah lama diperdebatkan. Dengan demikian, praktik terbaik menggunakan berbagai indikator untuk memberikan gambaran yang lebih kuat dan pluralistik. Jika ketidakpastian dan kesalahan dapat dikuantifikasi, misalnya dengan menggunakan diagram kesalahan, diagram ini harus menyertai nilai indikator yang dipublikasikan. Jika tidak memungkinkan, pembuat indikator setidaknya harus menghindari presisi palsu. Misalnya, faktor dampak jurnal ditampilkan hingga desimal ketiga untuk menghindari kesamaan nilai. Namun, mengingat ambiguitas konseptual dan variabilitas acak dari penghitungan pengutipan, tidak masuk akal membedakan jurnal berdasarkan perbedaan faktor dampak yang sangat kecil. Hindari presisi yang salah: hanya satu desimal yang perlu.

### **9) Kenali dampak sistemik terhadap penilaian dan indikator.**

Indikator mengubah sistem melalui dampak insentif yang terhasil. Dampak-dampak ini harus diantisipasi. Ini berarti bahwa seperangkat indikator selalu lebih baik - indikator tunggal akan mengundang manipulasi dan pergeseran tujuan (di mana penilaian yang menjadi tujuan). Sebagai contoh, pada tahun 1990an, universitas Australia mendanai penelitian universitas dengan menggunakan formula yang sebagian besar didasarkan pada jumlah publikasi makalah. Universitas dapat menghitung 'nilai' sebuah makalah dalam jurnal tertinjau; pada tahun 2000, nilainya adalah 800 \$Aus (sekitar 480 \$AS pada tahun 2000). Bisa ditebak, jumlah makalah yang diterbitkan oleh peneliti Australia naik, tetapi pada jurnal yang kurang dikutip, yang menunjukkan bahwa kualitas makalah berkurang (10).

### **10) Periksa dan perbarui indikator secara teratur.**

Misi penelitian dan tujuan penilaian bergeser dan sistem penelitian sendiri berkembang. Bibliometrik yang dulu mencukupi menjadi tidak memadai; bibliometrik yang baru muncul. Sistem indikator harus ditinjau dan jika perlu diubah. Menyadari dampak dari formula simplistik, Australia pada tahun 2010 memperkenalkan inisiatif Excellence in Research yang lebih kompleks, dan menekankan kualitas.

### **Langkah-langkah selanjutnya**

Berdasarkan sepuluh prinsip ini, penilaian penelitian dapat memainkan bagian penting dalam pengembangan sains dan interaksinya dengan masyarakat. Bibliometrik penelitian dapat memberikan informasi penting yang sulit dikumpulkan atau dipahami melalui keahlian individu. Tetapi informasi kuantitatif ini tidak boleh dibiarkan berubah dari alat menjadi tujuan.

Keputusan terbaik dapat diambil dengan menggabungkan statistik yang kokoh dengan kepekaan terhadap tujuan dan sifat penelitian yang dinilai. Kedua-dua bukti kuantitatif dan kualitatif diperlukan; masing-masing objektif dengan caranya sendiri. Pengambilan keputusan tentang sains harus didasarkan pada proses berkualitas tinggi yang didasarkan pada data berkualitas tertinggi.

Nature 520, 429–431 (23 April 2015), doi: 10.1038/520429a

## Referensi

1. Wouters, P. in *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (ed. 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).