

## MANIFESTO DE LEIDEN SOBRE AS MÉTRICAS PARA AVALIAR A INVESTIGACIÓN

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>c,d,e</sup>

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

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

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

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

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

Versión galega de Hicks et al. *The Leiden Manifesto for research metrics*. *Nature*, v. 520, p. 429-431, 2015, <http://www.leidenmanifesto.org/>

Cada vez máis empréganse datos para a xestión da ciencia. As avaliacións da investigación, que antes facían os pares consonte se lles solicitaba, son agora sistemáticas e baseadas en métricas formais<sup>1</sup>. O problema é que a avaliación depende máis desas métricas ca dos xuízos de valor. Os indicadores multiplícanse: xeralmente ben intencionados, non sempre ben fundamentados, con frecuencia, mal aplicados. Corremos o risco de prexudicar o sistema coas propias ferramentas destinadas a melloralo, xa que as institucións realizan cada vez máis as avaliacións sen un coñecemento suficiente nin un bo asesoramento sobre as boas prácticas e a axeitada interpretación dos indicadores.

Antes de 2000 as persoas expertas usaban para as súas análises especializadas o Science Citation Index (SCI) do Instituto de Información Científica (ISI) na súa versión en CD-ROM. En 2002, Thomson Reuters puxo en marcha unha plataforma web integrada, facendo accesible a un público máis amplo a base de datos da Web of Science. Axiña xurdiron índices de citas alternativos: a base de datos Scopus de Elsevier (lanzada en 2004) e Google Scholar (versión beta lanzada en 2004). Creáronse outras ferramentas en liña para facilitar a comparación da produtividade da investigación institucional e o seu impacto, como InCites (que usa datos da Web of Science) e SciVal (con datos de Scopus), e tamén programas para analizar perfís individuais de citación utilizando datos de Google Scholar (Publish or Perish, lanzado en 2007).

En 2005 Jorge Hirsch, físico da Universidade de California en San Diego, propuxo o índice *h*, popularizando a cuantificación das citas de cada investigador individualmente. O interese polo factor de impacto das publicacións periódicas aumentou constantemente desde 1995.

Máis recentemente gañan forza as métricas relacionadas coas redes sociais e os comentarios en liña, como F1000 Prime, creado en 2002, Mendeley, en 2008, e Altmetric.com en 2011 (financiado por Macmillan Science and Education, propietarios de Nature Publishing Group).

Como especialistas en cienciometría, científicos sociais e xestores de investigación, observamos con crecente alarma a aplicación errónea dos indicadores na avaliación do rendemento científico. Os seguintes son só algúns dos moitos exemplos posibles. En todo o mundo, as universidades obsesionáronse coa súa posición nas clasificacións mundiais (como a de Shanghai e a do *Times Higher Education*), aínda que estas listaxes están baseadas, na nosa opinión, en datos incorrectos e indicadores arbitrarios.

Algunhas institucións solicítanlles o seu índice *h* ás persoas que se presentan ás súas ofertas de emprego. Varias universidades basean as súas promocións de persoal en valores mínimos do índice *h* e no número de artigos publicados en revistas de "alto impacto". Os currículos de investigadores e investigadoras convertéronse en espazos onde alardear destas "puntuacións",

especialmente no ámbito da biomedicina. En todas partes, os titores pídenlle ao seu alumnado de doutoramento que publiquen en xornais de "alto impacto" e obteñan financiamento externo antes de estaren preparados.

En Escandinavia e en China algunhas universidades distribúen fondos de investigación ou incentivos baseándose nun número: por exemplo, calculando as puntuacións individuais de impacto para asignar "recursos por rendemento", ou ofrecéndolles unha gratificación aos investigadores por aparecer en revistas cun factor de impacto superior a 15<sup>2</sup>.

En moitos casos, quen investiga e avalía aínda se esforza por manter un xuízo equilibrado. Porén, o abuso dos indicadores na investigación tornouse algo demasiado estendido como para ignoralo.

É por iso que presentamos o Manifesto de Leiden, que recibe o nome da conferencia da que xurdiu (véxase <http://sti2014.cwts.nl>). Os seus dez principios non son novos para as persoas especialistas en cienciometría, aínda que ninguén de nós sería quen de recitalos na súa totalidade debido a que estaban sen sistematizar ata o de agora. Celebidades no campo da cienciometría como Eugene Garfield (fundador do ISI), xa mencionaran algúns destes principios<sup>3,4</sup>. Pero estes expertos non están presentes cando os avaliadores informan os xestores universitarios, que tampouco son expertos nesta metodoloxía. Os científicos que buscan bibliografía para rebater unha avaliación descubren que a información que precisan está espallada nas que, para eles, son unhas publicacións inintelixibles e ás que non teñen acceso.

Ofrecemos esta síntese de boas prácticas de avaliación baseada en indicadores métricos para que os investigadores poidan pedirlle contas a quen os avalía, e para que estes avaliadores poidan pedirilles contas aos seus indicadores.

## Os dez principios

**1. A avaliación cuantitativa debe servirle de apoio ao asesoramento cualitativo e experto.** Os indicadores cuantitativos poden corrixir as tendencias pouco obxectivas na revisión feita por pares e facilitar a deliberación. Isto debería reforzar a revisión por pares, xa que é difícil emitir xuízos de valor sobre colegas sen dispor de abundante información relevante. Non obstante, os avaliadores non deben ceder á tentación de basear as súas decisións só en números. Os indicadores non deberían substituír o xuízo informado. Cada persoa debe responsabilizarse das súas valoracións.

**2. Mida os resultados segundo os obxectivos de investigación da institución, grupo ou investigador.** Os obxectivos dun programa de investigación deberíanse indicar desde o principio, e os indicadores utilizados para avaliar os resultados deberían estar claramente ligados a eses obxectivos. A elección dos indicadores e o xeito de utilizalos debería ter en conta contextos socioeconómicos e culturais máis amplos. Os científicos teñen distintas misións. A investigación que avanza nas fronteiras do coñecemento académico difire da que se centra en proporcionar solucións aos problemas da sociedade. A avaliación pode basearse nos méritos relevantes para as políticas públicas, a industria ou os cidadáns en xeral, no canto de basearse nas nocións académicas de excelencia. Non hai un único modelo de avaliación que se aplique a todos os contextos.

**3. Protexa a excelencia na investigación de interese local.** En moitas partes do mundo a excelencia na investigación está asociada á publicación en lingua inglesa. A lei española, por exemplo, menciona explicitamente a conveniencia de que os investigadores españois publiquen en revistas de alto impacto. O factor de impacto calcúlase para revistas indexadas na Web of

Science, base de datos centrada en publicacións dos Estados Unidos e aínda maioritariamente en inglés. Esta tendencia é especialmente problemática para as ciencias sociais e as humanidades, áreas nas que a investigación está máis orientada cara a temas rexionais e nacionais. Moitas outras áreas teñen unha dimensión nacional ou rexional, como a epidemioloxía do VIH na África subsahariana.

Este pluralismo e relevancia social adoitan eliminarse para publicar artigos de interese para os gardiáns do alto impacto: as revistas en lingua inglesa. Os sociólogos españois que son altamente citados na Web of Science traballaron con modelos abstractos ou con datos dos Estados Unidos. Neste proceso pérdese a especificidade dos sociólogos en traballos en español de alto impacto: cuestións como a lexislación laboral local, os servizos de saúde para persoas maiores ou o emprego das persoas inmigrantes<sup>5</sup>. Unhas métricas baseadas en revistas de alta calidade publicadas en idiomas distintos do inglés servirían para identificar e recompensar a excelencia na investigación de interese local.

#### **4. Os procesos de recompilación e análise de datos deben ser abertos, transparentes e simples.**

A construción das bases de datos necesarias para a avaliación debe seguir regras claramente definidas e fixadas antes de rematar a investigación. Esta foi a práctica habitual entre os grupos académicos e empresariais que desenvolveron a metodoloxía de avaliación bibliométrica durante décadas. Estes grupos baseáronse en protocolos publicados na literatura revisada por pares. A transparencia fixo posible o seu escrutinio. Por exemplo, en 2010, o debate público sobre as propiedades técnicas dun importante indicador utilizado por un dos nosos grupos (o Centro para os Estudos da Ciencia e a Tecnoloxía, CWTS, da Universidade de Leiden, Países Baixos) levou a unha revisión do cálculo deste indicador<sup>6</sup>. Os novos participantes do sector privado deberían axustarse aos mesmos estándares; ninguén debería aceptar avaliacións automáticas de orixe opaca.

A simplicidade é unha virtude nun indicador porque favorece a transparencia. Pero as métricas simplistas poden provocar distorsións nos resultados (véxase o principio 7). Os avaliadores deben esforzarse por atopar un equilibrio: indicadores sinxelos e fieis á complexidade do proceso de investigación.

**5. Permítalles ás persoas avaliadas verificar os datos e as análises.** Para garantir a calidade dos datos, todos os investigadores incluídos nos estudos bibliométricos deben poder comprobar se as súas publicacións foron identificadas correctamente. Quen dirixa e xestione os procesos de avaliación debe asegurar a exactitude dos datos, sexa por verificación propia ou por auditoría de terceiros. As universidades poderían aplicar este principio nos seus sistemas de información para a investigación, e debería ser un principio básico na selección dos provedores destes sistemas. A recollida e o procesamento de datos precisos e de alta calidade requiren tempo e diñeiro. Resérvelle unha parte dos orzamentos.

#### **6. Teña en conta as diferenzas nas prácticas de publicación e citación entre campos científicos.**

A práctica máis axeitada é seleccionar un conxunto de indicadores posibles e permitir que as distintas áreas elixan entre eles. Hai uns anos, un grupo europeo de historiadores recibiu unha cualificación relativamente baixa nunha revisión nacional feita por pares porque escribiron libros en lugar de artigos en revistas indexadas na Web of Science. Estes historiadores tiveron a mala sorte de formar parte dun departamento de psicoloxía. Os historiadores e os científicos sociais precisan que os libros e os traballos publicados na súa lingua se inclúan nos seus recontos de publicacións, namentres que os especialistas en tecnoloxías da computación precisan que sexan tidos en conta os traballos presentados en conferencias.

As taxas de citación varían segundo os campos: as revistas mellor valoradas en matemáticas teñen un factor de impacto ao redor dun 3, mentres que no eido da bioloxía celular teñen factores de impacto ao redor dun 30. Necesítanse indicadores normalizados, e o método de normalización máis fiable é o baseado en percentís: cada artigo pondérase segundo o percentil en que se atope na distribución de citacións da súa área (no 1%, 10% ou 20% dos mellores, por exemplo). Unha única publicación altamente citada mellora só lixeiramente a posición dunha universidade nunha clasificación baseada en indicadores de percentís, pero pode impulsar a universidade desde posicións medias ata os primeiros postos nunha clasificación baseada na media de citas<sup>7</sup>.

**7. Basee a avaliación individual no xuízo cualitativo das publicacións.** Canto máis veterano é un investigador, maior será o seu índice  $h$ , aínda que non publique novos artigos. O índice  $h$  varía segundo a área: os investigadores das ciencias da vida poden chegar a 200, os de física a 100 e os de ciencias sociais a 20-30<sup>8</sup>. Depende tamén da base de datos: hai investigadores en informática que teñen un índice  $h$  de 10 en Web of Science, pero de 20 ou 30 en Google Scholar<sup>9</sup>. É moito máis apropiado ler e xulgar o traballo dun investigador que confiar nun número. Mesmo cando se compara un gran número de investigadores, é preferible un enfoque que inclúa información diversa sobre o coñecemento, a experiencia, as actividades e a influencia de cada individuo.

**8. Evite a concreción inadecuada e a falsa precisión.** Os indicadores científicos e tecnolóxicos son propensos á ambigüidade e á incerteza conceptuais; requiren de fortes hipóteses que non son universalmente aceptadas. O significado do reconto de citas, por exemplo, ten sido amplamente discutido. É máis adecuado para a avaliación empregar varios indicadores, que proporcionen unha imaxe máis sólida e plural da investigación. Se se poden cuantificar as incertezas e os erros, por exemplo utilizando barras de erro, esta información debería acompañar os valores dos indicadores publicados. Se isto non é posible, os produtores dos indicadores deberían cando menos evitar a falsa precisión. Por exemplo, o factor de impacto das revistas publícase con tres decimais para evitar empates. Non obstante, dada a ambigüidade conceptual e a variabilidade aleatoria dos recontos de citas, non ten sentido comparar revistas baseándose en diferenzas mínimas no factor de impacto. Evite a precisión falsa: só se pode xustificar un dígito.

**9. Identifique os efectos sistémicos da avaliación e dos indicadores.** Os indicadores cambian o sistema de investigación a través dos incentivos que establecen. Estes efectos deberíanse prever. Isto significa que sempre é preferible empregar un conxunto de indicadores, xa que un único indicador invita a manipulacións ou a desviacións dos obxectivos (onde a medida se converte nun fin en si mesma). Por exemplo, na década de 1990 Australia financiou a investigación universitaria a través dunha fórmula baseada principalmente no número de artigos publicados por cada institución. As universidades podían calcular o "valor" de cada artigo publicado en revistas avaliadas por pares; no ano 2000 equivalían a 800 dólares australianos por artigo (ao redor de 480 dólares dos EUA) dos fondos de investigación. Como era de esperar, aumentou a cantidade de artigos publicados por investigadores australianos, pero en revistas menos citadas, o que suxire unha diminución da calidade dos artigos<sup>10</sup>.

**10. Revise e actualice os indicadores periodicamente.** As misións da investigación científica e os obxectivos da avaliación van mudando e o propio sistema de investigación evoluciona con eles. Parámetros que foron útiles no pasado tórnase inadecuados e xorden outros novos. Os sistemas de indicadores deben ser revisados e quizais modificados. Ao se decataren das consecuencias da súa fórmula simplista, en 2010 Australia lanzou a iniciativa "Excelencia en investigación para Australia", máis complexa e con énfase na calidade.

## Os seguintes pasos

Respectando estes dez principios, a avaliación da investigación pode desempeñar un papel importante no desenvolvemento da ciencia e as súas interaccións coa sociedade. Os indicadores de investigación poden proporcionar información crucial que sería difícil de reunir ou de comprender por medio da experiencia individual. Pero non se debe permitir que esta información cuantitativa pase de ser un instrumento a transformarse nun fin en si mesma.

As mellores decisións tómanse combinando estatísticas sólidas con sensibilidade aos obxectivos e á natureza da investigación que se está a avaliar. Precísanse tanto os datos cuantitativos como os cualitativos; cada un deles é obxectivo ao seu xeito. A toma de decisións na ciencia debe estar baseada en procesos de alta calidade fundamentados en datos tamén de máxima calidade.

## Referencias

1. WOUTERS, P. The citation: From culture to infrastructure. In: CRONIN, B.; SUGIMOTO, C. (Eds.). **Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact**. Cambridge, MA: MIT Press, 2014. p. 47–66
2. SHAO, J.; SHEN, H. The outflow of academic papers from China: why is it happening and can it be stemmed? **Learned Publishing**, v. 24, p. 95–97, 2011
3. SEGLEN, P. O. Why the impact factor of journals should not be used for evaluating research. **British Medical Journal**, v. 314, n. 7079, p. 498–502, 1997
4. Garfield, E. J. The history and meaning of the journal impact factor. **Journal of the American Medical Association**, v. 95, n. 1, p. 90–93, 2006
5. LÓPEZ PIÑEIRO, C.; HICKS, D. Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. **Research Evaluation**, v. 24, n. 1, p. 78–89, 2014
6. VAN RAAN, A. F. J.; VAN LEEUWEN, T. N.; VISSER, M. S. et al. Rivals for the crown: Reply to Opthof and Leydesdorff. **Journal of Informetrics**, v. 4, n. 3, p. 431–435, 2010
7. WALTMAN, L.; CALERO-MEDINA, C.; KOSTEN, J. et al. The Leiden Ranking 2011/2012: Data Collection, Indicators, and Interpretation. **Journal of the American Society for Information Science and Technology**, v. 63, n. 12, p. 2419–2432, 2012
8. HIRSCH, J. E. An index to quantify an individual's scientific research output. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 46, p. 16569–16572, 2005
9. BAR-ILAN, J. Which h-index?—A comparison of WoS, Scopus and Google Scholar. **Scientometrics**, v. 74, n. 2, p. 257–271, 2008
10. BUTLER, L. Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. **Research Policy**, v. 32, p. 143–155, 2003