

---

## Del Manifiesto de Leiden a las reformas de la evaluación: retos hacia un uso responsable de la bibliometría

From the *Leiden Manifesto* to the evaluation reforms:  
challenges towards a responsible use of bibliometrics

**Ismael Ràfols**

[\(i.rafols@cwts.leidenuniv.nl\)](mailto:i.rafols@cwts.leidenuniv.nl)

Centre for Science and Technology Studies (CWTS), Leiden University  
Science Policy Research Unit (SPRU), University of Sussex

---

Recibido: 30-05-2023; Revisado: 07-06-2023, 12-06-2023; Publicado: 29-06-2023

**Resumen:** Este comentario revisa los principios del *Manifiesto de Leiden* y su énfasis en la necesidad de contextualizar de los indicadores a los procesos específicos de evaluación por parte de expertos. Esta visión de la importancia de la contextualización y pluralización y el uso responsable de los indicadores es central en las reformas de la evaluación impulsadas en estos momentos por la Comisión Europea y una coalición internacional de instituciones de investigación (CoARA). Argumento que la implementación de estas reformas en España no será posible sin una reforma de la gobernanza que descentralice muchos procesos de evaluación y otorgue más autonomía a los centros de investigación para gestionar la contratación y promoción del personal docente e investigador.

**Palabras clave:** Evaluación, indicadores, bibliometría, reforma.

**Abstract:** This commentary reviews the principles of the *Leiden Manifesto* and its focus on the need of experts for contextualizing indicators to specific evaluation processes. This vision about the importance of contextualization and pluralization and the responsible use of indicators is central in the reforms of evaluation promoted by the European Commission and an international coalition of research institutions (CoARA). I argue that the implementation of these reforms in Spain will not be possible without a governance reform that devolves many evaluation processes and give more autonomy to research centres to manage recruitment and promotion of lecturers and researchers.

**Keywords:** Evaluation, indicators, bibliometrics, reform.

**Como citar este artículo/Citation:** Ràfols, I. (2023). Del *Manifiesto de Leiden* a la reforma de la evaluación: retos hacia un uso responsable de la bibliometría. *Enredadera: revista de la Red de Bibliotecas y Archivos del CSIC*, (39), 21-28. <https://doi.org/10.20350/digitalCSIC/15376>

---

## 1. INTRODUCCIÓN

Nos encontramos en un momento decisivo en la evolución de la evaluación científica. Después de dos o tres decenios en los que ha dominado la visión de que la evaluación cuantitativa era un buen (o quizás el mejor) método para evaluar de modo justo y objetivo, ahora se cuestionan los indicadores



bibliométricos y se están llevando a cabo reformas que limitan su uso, auspiciadas por la Comisión Europea, y apoyadas en el caso español, por organizaciones como la ANECA, la CRUE o el CSIC que eran reticentes al cambio hasta hace poco tiempo. ¿Hacia dónde va la evaluación? Partiendo del *Manifiesto de Leiden* (Hicks et al., 2015) del que fui promotor y autor, en este comentario me propongo analizar los cambios que se están produciendo en el uso evaluativo de indicadores bibliométricos, con un foco en la situación española.

## 2. DE LA CIENCIOMETRÍA ANALÍTICA A LA BIBLIOMETRÍA EVALUATIVA

En sus orígenes en los años 60, la bibliometría tenía un interés analítico para historiadores de la ciencia como Derek de Solla Price y sociólogos como Robert Merton. En los años 1980, esta especialidad científica se consolidó y empezó a extenderse su uso para aplicaciones evaluativas.

Hubo un debate duro y extenso de las limitaciones del uso de la bibliometría en evaluación, en los que los pioneros de los indicadores bibliométricos (por ejemplo, Martin & Irvine, 1983) propusieron que se trataba de ‘indicadores parciales’ que:

- i) sólo deben aplicarse a grupos o institutos (no a individuos)
- ii) reflejan influencia científica pero no la calidad
- iii) se necesita una batería de indicadores (no sólo un indicador) para capturar el desempeño de un grupo
- iv) sirven solo para comparaciones entre grupos similares
- v) debido a su carácter parcial, sólo deben tomarse como válidos cuando la batería de indicadores ofrece resultados consistentes (convergentes)

A pesar de estas precauciones, en la práctica, los indicadores ofrecen grandes ventajas a los gestores científicos en comparación con la evaluación por pares y fueron paulatinamente adoptados. Como instrumentos del *New Public Management*, que ha dominado la gestión pública con una perspectiva empresarial desde los 80, los indicadores permiten poner el énfasis en la eficiencia, el control administrativo externo sin conocimiento experto, y el estímulo por incentivos (Gläser & Laudel, 2007). Esto llevó a que se extendiera su uso sin tener en cuenta sus limitaciones, un uso que se aceleró extraordinariamente desde la primera década del siglo XXI cuando las grandes plataformas bibliométricas pasaron a ser accesible en internet. En los países líderes en ciencia su uso ha sido principalmente informal, pero en algunos países (generalmente periféricos) los adoptaron formalmente para aplicar la evaluación de modo burocrático, en contextos como el español o el italiano donde hay sospechas fundadas de nepotismo en la evaluación por pares (Rafols et al., 2016).

### 3. EL MANIFIESTO DE LEIDEN: HACIA UN USO RESPONSABLE DE LOS INDICADORES

La expansión del uso evaluativo de la bibliometría generó muchos problemas. Primero, a causa de los sesgos de los indicadores bibliométricos en contra de la investigación aplicada, de relevancia local, de campo, no publicadas en inglés, y en disciplinas menos académicas o fuera de las ciencias naturales (Molas-Gallart & Rafols, 2018). Segundo, a causa del comportamiento estratégico adoptado por muchos investigadores que se han focalizado en las actividades cubiertas por los indicadores y a expensas de otras como la docencia o la contribución social (Delgado-López-Cózar et al., 2021; Rijcke et al., 2016).

Las críticas al uso inapropiado de indicadores fue creciendo (Burrows, 2012; Weingart, 2005) hasta que se generó un movimiento de reforma dentro mismo de la comunidad científica que propuso la *Declaración de San Francisco sobre Evaluación de la Investigación (DORA)*<sup>1</sup> en 2015, en contra del uso de ciertos indicadores. Para los expertos en bibliometría (como yo), la primera declaración DORA era incómoda porque dejaba claro qué prácticas eran nocivas (usar el factor de impacto de revistas), pero no ofrecía alternativas en positivo sobre cómo sí que podían usarse los indicadores.

Así que, en paralelo a DORA, los bibliómetros organizamos debates sobre el uso evaluativo de los indicadores en los dos congresos de cienciometría en 2013 (Wouters, 2013; Wouters et al., 2013). Inicialmente se propuso la creación de “estándares de indicadores para evaluación”, de acuerdo con las prácticas en campos profesionales aplicados como la arquitectura o la ingeniería. Pero luego nos dimos cuenta de que los contextos de evaluación eran tan diversos que era más apropiado y pragmático proponer principios generales para la buena aplicación de indicadores. Así, en el congreso de indicadores de ciencia y tecnología del 2014, nació el *Manifiesto de Leiden*<sup>2</sup> (Hicks et al., 2015) con una voluntad propositiva: cómo facilitar el buen uso de indicadores. No se quería desecharlos, sino mejorar su uso. En este sentido, creo que no es casual que la autora líder del manifiesto, Diana Hicks, fuera discípula y colaboradora de Ben Martin, uno los pioneros de la bibliometría evaluativa.

El *Manifiesto* tiene dos ideas fundamentales. En primer lugar, los indicadores pueden apoyar la evaluación por expertos, pero no pueden sustituirlos: se necesita el experto para hacer una interpretación del indicador (Principio 1, P1). En segundo lugar, los indicadores deben adecuarse al contexto. Misiones de investigación distintas (más académica, más aplicada, de un campo científico u otro) requieren indicadores distintos (P2 y P6). Y la atención al contexto local es especialmente importante, en especial con relación a la lengua de uso y a las temáticas sociales, culturales o aplicaciones (P3). Estas dos ideas (primacía del experto y contextualización) tienen especial relevancia en la evaluación

---

<sup>1</sup> <https://sfdora.org/>

<sup>2</sup> La página web <http://www.leidenmanifesto.org/> incluye traducciones al catalán, castellano, euskera y gallego (y otras 21 lenguas), y una versión en video.

individual, donde la bibliometría mostró que los indicadores no son estadísticamente robustos y los contextos cambiantes (P6).

Además, el *Manifiesto* aporta principios sobre buenas prácticas en relación con la transparencia (P4), la verificación de datos (P5), la incertidumbre y los errores en las medidas (P8), la anticipación de efectos sistémicos de la evaluación (P9) y la necesidad de adaptación a condiciones cambiantes (P10).

En 2015 también se publicó *The metric tide (La marea métrica)* que proponía un marco conceptual más amplio de debate, inspirado en conceptos de ciencia responsable (RRI: *Responsible Research and Innovation*), apoyándose en aprendizajes del buen uso de tecnologías en sociedad (Wilsdon et al., 2015). Después ha habido otras iniciativas importantes como el Foro Latinoamericano de Evaluación Científica (FOLEC-CLACSO<sup>3</sup>), la Iniciativa de Helsinki sobre multilingüismo<sup>4</sup> (2019), los principios de Hong Kong (Moher et al., 2020), o el marco SCOPE de evaluación para gestores científicos<sup>5</sup> – y documentos institucionales varios apoyando buenas prácticas (Saenen et al., 2019).

#### 4. LAS REFORMAS EN LA EVALUACIÓN Y LA CONTEXTUALIZACIÓN DE INDICADORES

A pesar de todas estas las declaraciones y documentos de buenas prácticas, las reformas de los sistemas de evaluación han sido menores. Hay una gran reticencia a cambiar las prácticas porque las alternativas requieren más esfuerzo y recursos. Además, las alternativas incomodan a la mayoría del *establishment* porque cuestionan la jerarquía de méritos académicos utilizada hasta ahora y, por lo tanto, implícitamente también los méritos de muchos de los científicos en cargos políticos o ejecutivos.

Ahora bien, esta falta de cambios genera un problema político, porque la evaluación bibliométrica supone también una barrera para la adopción de reformas hacia una ciencia abierta, socialmente responsable y participativa, justamente el tipo de ciencia que es percibida como necesaria para que la investigación contribuya a la innovación sostenible. Ante esta inmovilidad, la *Comisión Europea* impulsó en 2021 un acuerdo de principios y una coalición de instituciones para el avance de la reforma en la evaluación (*CoARA*<sup>6</sup>) que supone un empuje importantísimo hacia las reformas.

El acuerdo de *CoARA* recoge las líneas principales de pensamiento de DORA, el *Manifiesto de Leiden* o los *Principios de Hong Kong* en su rechazo a indicadores como el Factor de Impacto de revistas y el índice h, y en la adopción de criterios de evaluación más diversos y pluralistas. Pero ante tal flexibilidad, ¿cuál es el papel de la bibliometría en este nuevo contexto? *CoARA* propone

---

<sup>3</sup> <https://www.clacso.org/folec/>

<sup>4</sup> <https://www.helsinki-initiative.org/es>

<sup>5</sup> <https://inorms.net/scope-framework-for-research-evaluation/>

<sup>6</sup> <https://coara.eu/>

que se tiene que hacer un uso responsable de los indicadores para apoyar la revisión por pares. ¿Pero qué es un uso responsable?

Este es un punto ambiguo, en el que las declaraciones y manifiesto anteriores pueden servir de referencia. Por ejemplo, a mi parecer un uso responsable seguirá los principios del *Manifiesto de Leiden* en temas como la transparencia y verificación de datos, los límites de precisión de muestras pequeñas, y la contextualización. La contextualización cuantitativa de indicadores se consigue con baterías de indicadores que permitan medir las dimensiones de interés, no solamente de impacto académico<sup>7</sup>, sino también en mejoras en salud, agricultura u otros ámbitos de aplicación. La contextualización cualitativa se consigue cuando los evaluadores conocen bien el campo científico y/o su aplicación, entienden los sesgos de los indicadores, y son capaces de interpretar los indicadores teniendo en cuenta las condiciones particulares: el grupo A es muy citado porque trabaja en un tema de moda, pero el grupo B también tiene un buen nivel de citas teniendo en cuenta que su tema es local (con lo que recibirá menos citas) pero socialmente importante en el territorio del financiador.

Lo ideal sería que la contextualización de los indicadores combinara conocimiento de indicadores y contara, además, con la participación de expertos del campo o beneficiarios de la investigación (Ràfols, 2019), pero este tipo de implementación requiere una evaluación descentralizada y formativa que choca con la gobernanza científica española.

## 5. LOS RETOS EN ESPAÑA: LA NECESIDAD DE DESCENTRALIZAR LOS PROCESOS DE EVALUACIÓN

Las principales instituciones científicas españolas han firmado el acuerdo y se han comprometido a aplicar sus principios. Sin embargo, su aplicación representa un gran reto en el contexto español por dos razones principales:

- i) la confusión entre medición y evaluación
- ii) la centralización de la evaluación individual en agencias.

En primer lugar, como en la tradición gubernativa del estado español se entiende a menudo la evaluación como auditoría o rendición de cuentas (Molas-Gallart, 2012), los indicadores se utilizan como baremos administrativos. Se entiende evaluar con controlar o medir si los investigadores o institutos están “produciendo” lo que deberían. Y se acaba confundiendo evaluar con ‘medir’, que es como confundir el valor o la valoración con el ‘precio’, algo que, como dijeron Quevedo y Machado, sólo hace el ‘necio’.

En otros países, la evaluación tiene, en términos relativos, unas funciones más de distribución de recursos y de aprendizaje. Para funciones que miran al futuro,

---

<sup>7</sup> En el caso de indicadores de citas, la contextualización se consigue aplicando bibliometría sofisticada con una caracterización de campos científicos precisa (en mi instituto, CWTS, usamos 4,000 campos en vez de las 200-300 categorías disciplinarias), y descripciones estadísticas rigurosas (por ejemplo, con percentiles); pero este nivel de rigor no es posible en la práctica en muchas evaluaciones.

está claro que la medición no basta para la valoración de la investigación. No basta con tener una medición del pasado para pensar qué se quiere financiar o dónde dirigirse: hay que razonar por qué se acepta o no un proyecto o por qué se toma una línea de investigación u otra. Además, en países como el Reino Unido o Dinamarca, la distribución de fondos por financiación competitiva de agencias es mucho más elevada y, por lo tanto, tiene un peso más importante que en España. En estados como los Países Bajos la evaluación se utiliza para ayudar a reflexionar y desarrollar estrategias de futuro (por ejemplo, en la evaluación de departamentos holandesa). Es más fácil entender la importancia de la contextualización y la pluralización de criterios fomentada por CoARA en evaluaciones que tienen por objetivo distribuir recursos o desarrollar estrategias, que en la tradición española de auditoría (Rafols et al., 2016).

En segundo lugar, la centralización de la evaluación individual en agencias externas (la ANECA y las autonómicas) resulta en procesos con miles de expedientes que no pueden ser abordados con el tiempo y los recursos que requeriría una evaluación contextualizada. Estos cuellos de botella administrativos se resolverían devolviendo a las organizaciones de investigación la autonomía para contratar y promocionar su propio personal (como ocurre en la mayoría de países europeos), y focalizando los recursos de las agencias externas para la evaluación de unidades institucionales como departamentos y facultades (Ràfols & Molas-Gallart, 2022). Además, trasladar la evaluación del individuo al grupo o centro tiene la ventaja de permitir una diversificación de perfiles dentro del grupo. Si me permiten la metáfora, la ciencia como el fútbol se juega en equipo y lo que cuenta no son los virtuosismos del jugador sino la consecución de goles (como hay *goals* científicos o sociales). La nueva ley de universidades (LOSU) ya posibilita desarrollar algunos pasos en esa dirección, como la acreditación institucional, que algunas agencias ya han estado ensayando.

## **6. CONCLUSIÓN: DIVERSIDAD, HUMILDAD CUANTITATIVA Y REFORMAS EN LA GOBERNANZA**

A las propuestas hacia un uso responsable de los indicadores en evaluación de hace unos 10 años, han seguido recientemente propuestas de reforma a la evaluación en Europa, CoARA. CoARA propone una implementación autónoma de la evaluación en cada institución. Desde el punto de vista de indicadores, esto significa que hay que hacer acto de ‘humildad cuantitativa’ porque los indicadores no son universales ni hablan por sí mismos: hay que adecuar el uso de los indicadores e interpretarlos según los contextos específicos de las organizaciones.

Esta flexibilidad choca en España con la confusión entre medición bibliométrica y evaluación, y con una organización individual y centralizada de muchos de los procesos de evaluación. En consecuencia, parece que la reforma de la evaluación en España necesitará de importantes cambios estructurales, de gobernanza. Por un lado, de un cambio cultural sobre qué significa evaluar, con menos énfasis en la medida de auditoría y más en la valoración formativa. Por

otro lado, de una devolución de los procesos evaluativos individuales a las universidades y centros de investigación.

Si queremos que la ciencia sirva a la sociedad para afrontar los retos sociales y ambientales del siglo XXI, la investigación debe ser valorada y, por lo tanto, evaluada, de acuerdo con la diversidad de contextos de usos del conocimiento – y eso requiere pluralidad en el uso e interpretación de indicadores, que sólo puede conseguirse con reformas de la gobernanza de la evaluación que permitan su contextualización.

## 5. REFERENCIAS

- Burrows, R. (2012). Living with the H-Index? Metric Assemblages in the Contemporary Academy. *The Sociological Review*, 60(2), 355–372. <https://doi.org/10.1111/j.1467-954X.2012.02077.x>
- Delgado-López-Cózar, E., Ràfols, I., & Abadal, E. (2021). Carta: Por un cambio radical en la evaluación de la investigación en España. *El profesional de la información*, 30(3), e300309. <https://doi.org/10.3145/epi.2021.may.09>
- Gläser, J., & Laudel, G. (2007). The Social Construction Of Bibliometric Evaluations. In R. Whitley & J. Gläser (Eds.), *The Changing Governance of the Sciences* (Vol. 26, pp. 101–123). Springer Netherlands. [http://dx.doi.org/10.1007/978-1-4020-6746-4\\_5](http://dx.doi.org/10.1007/978-1-4020-6746-4_5)
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2), 61–90. [https://doi.org/10.1016/0048-7333\(83\)90005-7](https://doi.org/10.1016/0048-7333(83)90005-7)
- Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., Coriat, A.-M., Foeger, N., & Dirnagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, 18(7), e3000737. <https://doi.org/10.1371/journal.pbio.3000737>
- Molas-Gallart, J. (2012). Research Governance and the Role of Evaluation: A Comparative Study. *American Journal of Evaluation*, 33(4), 583–598. <https://doi.org/10.1177/1098214012450938>
- Molas-Gallart, J., & Rafols, I. (2018). Why bibliometric indicators break down: Unstable parameters, incorrect models and irrelevant prope. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, (40). <https://doi.org/10.1344/BiD2018.40.23>
- Ràfols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation*, 28(1), 7–22. <https://doi.org/10.1093/reseval/rvy030>
- Ràfols, I., & Molas-Gallart, J. (2022). How to reform research evaluation in Spain. Institutional accreditation as a response to the European Agreement on research assessment. Letter. *El Profesional de la información*, 31(6), e310601. <https://doi.org/10.3145/epi.2022.nov.01>
- Rafols, I., Molas-Gallart, J., Chavarro, D. A., & Robinson-Garcia, N. (2016). *On the Dominance of Quantitative Evaluation in 'Peripheral' Countries: Auditing Research with Technologies of Distance* (SSRN Scholarly Paper ID 2818335). Social Science Research Network. <https://papers.ssrn.com/abstract=2818335>
- Rijcke, S. D., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>

- Saenen, B., Morais, R., Gaillard, V., & Borrell-Damián, L. (2019). *Research Assessment in the Transition to Open Science*. European University Association (EUA). <https://eua.eu/resources/publications/888:research-assessment-in-the-transition-to-open-science.html>
- Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1), 117–131. <https://doi.org/10.1007/s11192-005-0007-7>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Wouters, P. (2013, October 3). *Bibliometrics of individual researchers – the debate in Berlin*. <https://citationculture.wordpress.com/2013/10/03/bibliometrics-of-individual-researchers-the-debate-in-berlin/>
- Wouters, P., Glänzel, W., Gläser, J., & Rafols, I. (2013). The dilemmas of performance indicators of individual researchers—An urgent debate in bibliometrics. *ISSI Newsletter*, 9(3), 48–53. <https://www.issi-society.org/media/1139/newsletter35.pdf>