



Palabra Clave (La Plata)
ISSN:
ISSN: 1853-9912
palabraclave@fahce.unlp.edu.ar
Universidad Nacional de La Plata
Argentina

La marea métrica y su resaca: la experiencia británica en evaluación y gestión de la investigación

 López Ruiz, Osvaldo Javier

La marea métrica y su resaca: la experiencia británica en evaluación y gestión de la investigación

Palabra Clave (La Plata), vol. 12, núm. 2, e180, 2023

Universidad Nacional de La Plata

Disponible en: <https://www.redalyc.org/articulo.oa?id=350574508002>

DOI: <https://doi.org/10.24215/18539912e180>



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-Compartirlgual 4.0 Internacional.



La marea métrica y su resaca: la experiencia británica en evaluación y gestión de la investigación

The metric tide and its hangover: the British experience in research evaluation and management

Osvaldo Javier López Ruiz

Instituto de Ciencias Humanas, Sociales y Ambientales

(INCIHUSA-CONICET), Argentina

olopezruiz@mendoza-conicet.gob.ar

 <https://orcid.org/0000-0002-9685-3512>

DOI: <https://doi.org/10.24215/18539912e180>

Recepción: 15 Noviembre 2022

Aprobación: 09 Febrero 2023

Publicación: 03 Abril 2023



Acceso abierto diamante

Resumen

Este artículo revisa lo que ha sido la experiencia británica con el uso de las métricas en la evaluación y gestión de la investigación, pasando desde un primer momento de implementación “entusiasta”, a otro de revisión cautelosa y preocupada por los efectos distorsivos que éstas han tenido en las prácticas científicas y en sus resultados. Actualmente se plantea la necesidad de un cambio cultural en el uso de las métricas que revierta las consecuencias dañinas que está teniendo la utilización poco responsable de las mismas para el sistema científico-académico, sus instituciones y sus individuos. Tras la implementación del *New Public Management*, el Reino Unido fue uno de los primeros países en establecer una cultura de auditoría basada en criterios *manageriales* traídos del sector privado. Lo que muestra la evidencia analizada en los informes de una revisión de expertos independientes convocados por el gobierno británico es que la racionalidad *managerial* acaba, en muchos casos, prevaleciendo sobre la racionalidad de la propia ciencia y los miembros de la comunidad científica terminan teniendo que adaptarse, desviándose de sus objetivos y, a veces, teniendo que tomar atajos a las buenas prácticas que la propia comunidad ha establecido. Frente a esta situación, una serie de iniciativas vienen siendo promovidas en el Reino Unido para contrarrestar esta tendencia. Como esta situación se registra también en otros países, son revisadas otras propuestas recientes para devolverle a la comunidad académica y a las culturas de investigación el liderazgo efectivo y el control sobre su propia gestión.

Palabras clave: Racionalidad, Métricas, Evaluación de la investigación, Gestión de la investigación, Reino Unido.

Abstract

This article reviews the British experience with the use of metrics in the evaluation and management of research, moving from an initial period of “enthusiastic” implementation to a period of cautious review and concern about the distorting effects they have had on scientific practices and their results. There is currently a concern and a need for a cultural change in the use of metrics that reverses the harmful consequences of their irresponsible use for the scientific-academic system, its institutions and individuals. Following the implementation of New Public Management, the UK was one of the first countries to establish an audit culture based on managerial criteria brought in from the private sector. What the evidence analysed in the reports of a review of independent experts convened by the British government shows is that managerial rationality ends up, in many cases, prevailing over the rationality of science itself. Members of the scientific community end up having to adapt, deviating from their objectives and, sometimes, having to take shortcuts to the good practices that the community itself has established. In response to this situation, a number of initiatives are being promoted in the UK to counteract this trend. As this situation is also found in other countries, other recent proposals to restore to the academic community and research cultures effective leadership and control over their own management are reviewed.

Keywords: Rationality, Metrics, Research evaluation, Research management, United Kingdom.

1. Introducción

Hace más de dos décadas, Marilyn Strathern (2000), profesora de antropología social de la *University of Cambridge*, editó un libro sobre un tema muy distinto de aquellos por los que es conocida. Su título es “*Audit cultures: anthropological studies in accountability, ethics and the academy*” (Culturas de auditoría: estudios antropológicos en *accountability*, ética y la academia, traducción propia). En ese momento declaraba que el objetivo del libro era captar un tipo de cultura en proceso de formación porque la auditoría y la *accountability* (la rendición de cuentas) se habían vuelto omnipresentes en muy distintos ámbitos de lo social. En la academia, no obstante, siempre se ha estado familiarizado con cierto tipo de auditoría a través de las evaluaciones a la investigación y la enseñanza. Es casi imposible de criticar ya que promueve valores que los académicos aprecian, como la responsabilidad, la apertura de la investigación y la ampliación del acceso. Sin embargo, ya entonces se extendía la precepción de que estas nuevas prácticas de auditoría y rendición de cuentas obstruían las buenas prácticas científicas y tenían consecuencias nefastas para la producción intelectual.

Dos décadas antes, en el Reino Unido habían empezado a implementarse cambios radicales en la gestión de lo público a partir de la llegada al poder de Margaret Thatcher en 1979 y sus políticas neoliberales. Una reforma que a principios de los años noventa fue bautizada como *New Public Management* (nueva gestión pública) y que fue exportada al resto del mundo, en buena medida, a través de los organismos internacionales. Es por ello que la experiencia británica en la evaluación y gestión de la investigación si bien puede pensarse como un estudio de caso de la forma en la que ciencia es gestionada en un país determinado, también puede leerse –y este esperamos que sea nuestro aporte con el presente artículo–, como la posibilidad de poner la actual discusión de la evaluación y la gestión de la ciencia a nivel mundial bajo una perspectiva más amplia. Dicha perspectiva tiene que ver con transformaciones culturales profundas que alteran valores y visiones sobre lo que es la ciencia, su sentido para la sociedad, así como sobre el papel de los Estados y los organismos de financiación de la investigación; y, finalmente, también sobre el sentido de investigar y pertenecer a una comunidad científico-académica para los propios investigadores.

2. 2007: *New Public Management* y la búsqueda fervorosa de indicadores y métricas

A mediados del 2007, Gordon Brown asume como Primer Ministro del Reino Unido. Durante los diez años precedentes, cuando ese puesto lo ocupaba su antecesor, Tony Blair, Brown había estado a cargo del tesoro británico y de la economía y las finanzas de su país. Desde ese lugar había impulsado grandes reformas en el Estado. Recordemos que el proyecto de modernización del Nuevo Laborismo de Blair y Brown representaba la extensión de los métodos del *New Public Management* que se habían empezado a aplicar en el Reino Unido desde la década de 1980 durante los gobiernos conservadores de Thatcher y Major.¹ Como ejemplo de esto, en junio del 2007, justo antes de la asunción de Brown como Primer Ministro, un nuevo ministerio fue creado con el objetivo de dar coherencia y efectividad a las inversiones del gobierno en ciencia e investigación. El *Department for Innovation, Universities and Skills* (Departamento para la Innovación, las Universidades y las Competencias, en adelante DIUS). De él pasaron a depender los *Research Councils*, RCs (Consejos de Investigación Británicos) encargados de financiar la investigación y los posgrados en el Reino Unido, con lo que asumió para sí la tarea de mantener y desarrollar, con estándares internacionales, toda la base de investigación, mientras que buscaba perfeccionar la articulación de esa base como forma de apoyo a la innovación en todos los sectores de la economía (López Ruiz, 2010).

En ese contexto y en los días previos a la asunción de Brown, tuvimos la oportunidad de entrevistar a altos funcionarios de diferentes áreas del gobierno británico vinculados con las políticas de innovación que venían siendo promovidas desde el gobierno.² Esto fue en el marco de una investigación mayor sobre las estrategias de innovación en siete países,³ encargada por el *Ministério do Desenvolvimento, Indústria e*

Comércio Exterior del gobierno federal brasileño y la *Agência Brasileira de Desenvolvimento Industrial al Centro Brasileiro de Análise e Planejamento* (CEBRAP), en São Paulo, Brasil, en donde me desempeñaba como investigador, siendo responsable en ese momento por la parte de esa investigación que se llevó a cabo en el Reino Unido (Arbix, Salerno, Toledo, Miranda y Álvarez, 2010).

Vale la pena referir aquí brevemente algunos pasajes de una de las entrevistas porque, a nuestro entender, resume e ilustra bien lo que podríamos llamar “el espíritu de época” que se vivía hace una década y media en el ámbito científico-académico del Reino Unido. La entrevista en cuestión fue al Director Ejecutivo del *Arts and Humanities Research Council* (Consejo de Investigación en Artes y Humanidades, AHRC) quien estaba a cargo, por entonces, de la presidencia temporal de los siete RCs.⁴ Estos RCs cubren las diferentes áreas de conocimiento a través de los cuales se organiza el sistema científico-tecnológico británico para su financiamiento. El entrevistado pidió ser acompañado por una de sus asistentes quien se desempeñaba como jefa de un área técnica vinculada con el uso de métricas de evaluación y gestión de la investigación (*Director of Knowledge & Evaluation*). Un tema tomó primacía rápidamente en la entrevista y tenía que ver con que la presión del gobierno venía aumentando para que todos los RCs mostraran el impacto económico de lo que financiaban. No sólo era menester demostrar la mayor transferencia de conocimiento que se hacía, sino que debía expresarse en posibilidades de comercialización. Con una clara expresión de incomodidad, el presidente de los RCs británicos –quien era profesor de teología y estudios bíblicos– hablaba de los problemas de evaluación, de las diversas formas de medir el impacto y de cómo justificar los financiamientos. A continuación, le pidió a su asistente que nos diera algunos ejemplos de lo que habían hecho para cumplir con estos objetivos en el AHRC. Ella, en contraste, con una expresión de gran entusiasmo –semejante al que trasmitten ciertas formas de fervor religioso–, nos contó algunas de las estrategias que habían ideado para las artes y humanidades. Para mostrar, por ejemplo, la utilidad de la filosofía, habían ofrecido consultorías a grandes empresas (como *British Telecom*) para que especialistas formados en el área pudieran analizar y discutir con el personal conceptos como los de “lealtad” y pensar juntos en maneras en las que se pudiera establecer e incrementar la lealtad de sus consumidores. De forma semejante habían promovido experiencias con la industria farmacéutica en la que habían explorado el concepto “ética”. También, con el mismo objetivo, habían propuesto estrategias y metodologías de trabajo específicas para llevar, desde el mundo del arte, *insights* que pudieran ayudar a estimular la creatividad y la innovación en el mundo de los negocios. Todo el énfasis estaba puesto en encontrar los indicadores y las métricas que pudieran dar cuenta, de manera cuantitativa, de la transferencia y comercialización de los resultados de los proyectos financiados. “Si logramos demostrar el impacto económico, habrá más dinero para estas áreas” –concluía la directora de evaluaciones con la apasionada vehemencia de quien enuncia un postulado cuya lógica es irrefutable.

Lo que la entrevista dejó claro para quienes participamos de la misma en 2007 fue que en el Reino Unido, después de más de dos décadas de difusión de la doctrina del *New Public Management*, se había impuesto sobre el *ethos* científico-académico – con sus criterios de evaluación y gestión interna; es decir, definidos por la propia comunidad científica–, una racionalidad burocrático-*managerial* con criterios de gestión externos basado en una concepción productivista y cuantitativa de la ciencia. La forma en que el uso generalizado de indicadores y métricas llegaba a las artes y humanidades da, a nuestro entender, un ejemplo casi caricatural –y por ello muy elocuente– de cómo la gestión académico-científica había sido imbuida por esta lógica *managerial*. Como ya señalamos, nuestros entrevistados, si bien enfocaron sus ejemplos en su área específica, a la vez representaban a los siete consejos de investigación del Reino Unido de cuya presidencia estaban temporariamente a cargo.

3. 2015: *The metric tide* ... y su resaca

En julio de 2015 es publicado el informe de la “Revisión independiente sobre el papel de las métricas en la evaluación y gestión de la investigación”. Con el título de “*The metric tide*” (La marea métrica, traducción propia), aparece este exhaustivo informe en tres volúmenes encargado a un grupo multidisciplinario de expertos independientes por el entonces Ministro de Universidades y Ciencia. El

informe examina de cerca los posibles usos y limitaciones de las métricas e indicadores empleados para evaluar y gestionar la investigación, explorando a fondo su utilización dentro de las instituciones y entre las distintas disciplinas. Para su elaboración, este grupo, constituido en comisión directiva independiente, contó con el apoyo del equipo de políticas de investigación del *Higher Education Funding Council for England*, (Consejo de Financiamiento de la Educación Superior de Inglaterra, HEFCE),⁵ organismo estatal con la función de regular y financiar la educación superior y la investigación. Los autores del informe, un panel de especialistas en cienciometría, financiamiento, políticas de investigación, industria editorial y gestión universitaria, estuvo compuesto por doce miembros más un presidente. Ellos son profesores de reconocidas universidades, miembros de asociaciones científicas como la *British Academy* y la *Royal Society*, representantes de la Asociación de gestores y administradores de la investigación, el jefe de evaluaciones del Consejo de Investigación en Medicina (en representación de los RCs) y el editor en jefe de la Revista *Nature*.⁶

El Profesor James Wilsdon (*Sussex University*), quien ejerció la presidencia de la comisión directiva independiente que preparó el informe, abre el prefacio refiriéndose a las reacciones mixtas que evocan las métricas en la comunidad de investigadores. Junto con otros colegas, aluden, por un lado, a las simpatías e incluso al entusiasmo que despiertan las posibilidades de uso del *big data* dentro del sector científico y, por el otro, a los inconvenientes que ha ocasionado la utilización obtusa de métricas tales como los factores de impacto de las revistas, los índices-h y de las metas de ingresos por subvenciones (*grant income targets*).⁷ En este sentido afirman:

(...) los indicadores individuales tienen dificultades para hacerle justicia a la riqueza y pluralidad de nuestra investigación. Demasiado a menudo, criterios de evaluación mal diseñados están 'dominando las mentes, distorsionando los comportamientos y determinando carreras'.⁸ En su extremo, las métricas pueden contribuir con lo que Rowan Williams, el ex Arzobispo de Canterbury, llamó una 'nueva barbarie' en nuestras universidades (Wilsdon *et al.*, 2015, p. iii, traducción propia).

A continuación, remiten al caso trágico de Stefan Grimm, investigador del *Imperial College*, cuyo suicidio en setiembre de 2014, llevó a esa institución a iniciar una revisión del uso dado a las métricas de rendimiento. Vale la pena agregar que el caso de este prestigioso biólogo de 51 años a cargo de la Cátedra de Toxicología, sometido a enormes presiones para conseguir mantener el nivel de financiamiento de sus proyectos, tuvo gran repercusión pública y abrió una polémica que se extendió a los medios de comunicación. En particular, un artículo publicado en un *blog* por David Colquhoun, Profesor Emérito de Farmacología del *University College London*, con el título "*Publish and perish at Imperial College London: the death of Stefan Grimm*", fue visitado 196.000 veces durante el año que siguió a la muerte de Grimm. Allí, Colquhoun describe las metas de rendimiento del *Imperial College* como "una receta a corto plazo para la investigación poco original. Un incentivo a la búsqueda de atajos" y concluye afirmando, respecto de esta lógica de funcionamiento, que "es una prostitución de la ciencia" (Colquhoun, 2014). Para Wilsdon *et al.* (2015), en el prefacio de *The metric tide*, el caso Grimm es un impactante recordatorio de que lo que está en juego en estos debates es más que apenas el diseño de sistemas manageriales efectivos, frente a lo que afirma: "Las métricas tienen poder: ellas son constitutivas de valores, de identidades y de medios de subsistencia" (Wilsdon *et al.*, 2015, p. iii, traducción propia). Y concluyen enfatizando que el foco del informe es cómo ejercer ese poder con fines que sean positivos, proponiendo, después de quince meses de recolección de evidencias, análisis y consultas, un marco para el uso de las métricas de manera responsable.⁹

La metodología de trabajo que siguió la comisión a cargo de esta revisión del uso de las métricas implicó recurrir a fuentes y estrategias diversas. En primer lugar se hizo una convocatoria abierta para el envío de evidencias a instituciones de educación superior, individuos interesados, asociaciones científicas, "proveedores" (editores, consultores y empresas de estrategia), entre otros, para que enfocaran sus respuestas sobre cuatro cuestiones claves, a saber: a) la identificación de métricas útiles para la evaluación de la investigación, b) cómo se deben utilizar las métricas en la evaluación de la investigación, c) la manipulación ("gaming") y el uso estratégico de métricas, d) las perspectivas internacionales. A su vez,

fueron organizados tres talleres con la participación de las distintas partes interesadas que abordaron tópicos de preocupación general: el primero de ellos, en octubre del 2014 en la *Sussex University*, bajo el título: “*In metrics we trust? Prospects & pitfalls of new research metrics*”. El segundo, en diciembre de 2014 en la *University of Sheffield*, “*Metric for all? Equality and diversity workshop*”. Y el último, en enero de 2015, en la *Warwick University*: “*Metrics and the assessment of research quality and impact in the art and humanities*”. A su vez, distintos miembros de la comisión fueron designados para participar en eventos y reuniones de diverso tipo con las partes interesadas (incluso en otros países), presentando luego informes a la comisión en sus reuniones plenarias.

Paralelamente, la comisión encomendó dos informes complementarios para dar fundamentos a sus análisis y conclusiones. El primero de ellos consistió en una revisión bibliográfica encargada a un grupo interdisciplinario de autores especialistas en estudios sobre la ciencia y la tecnología, la bibliometría y la estadística, liderados por Paul Wouters. Los siete autores de este informe pertenecen a dos instituciones diferentes: el *Center for Science and Technology Studies* (CWTS) de la *Leiden University* en Holanda y el *Statistical Cybermetrics Research Group* de la *University of Wolverhampton* en el Reino Unido. El objetivo expreso fue asegurar que las recomendaciones de la revisión independiente estuvieran basadas en investigaciones académicas relevantes. Por eso, allí se discuten los más variados aspectos concernientes a la bibliometría y al uso de indicadores, la revisión por pares y al uso de métricas alternativas, teniendo como marco el estado del arte más actual de la literatura específica –el que remite a un listado de referencias de casi setenta páginas (Wouters *et al.*, 2015, pp. 108-175). Por su parte, el segundo informe complementario estuvo a cargo de la Dirección de Servicios Analíticos del HEFCE y consistió en un análisis de correlación entre las puntuaciones del Marco de Excelencia en Investigación 2014 (*Research Excellence Framework*, REF)¹⁰ a nivel de artículos por autor, y un conjunto de 15 indicadores bibliométricos y altmétricos (nuevas métricas alternativas). Los datos fueron proporcionados por la empresa Elsevier y el análisis cubrió 149.670 productos individuales, constituyendo hasta esa fecha el análisis más exhaustivo de correlación entre estas variables (Higher Education Funding Council for England (HEFCE), 2015a).¹¹ El objetivo de este trabajo fue evaluar hasta qué punto los indicadores bibliométricos correlacionaron o predijeron el resultado de la evaluación de revisión por pares hecha por el REF.

La principal conclusión de esta revisión es que las métricas no pueden y no deben reemplazar la revisión por pares (*peer review*). La evaluación por pares, debe ser la base fundamental para la evaluación de la investigación. Las métricas deben apoyar, no suplantar la opinión de los expertos. A pesar de sus defectos y limitaciones, la evaluación por pares sigue suscitando un amplio apoyo a través de las diferentes disciplinas como la base primaria para la evaluación de los resultados de la investigación, los proyectos y los individuos. Como hace notar Ellen Hazelkorn en una presentación hecha para el Banco Mundial en enero de 2015, citada en el informe:

la calidad académica es una noción compleja que no puede ser fácilmente reducida a la cuantificación –el uso de variables que la representen [proxy variables] corre el riesgo de malinterpretar las cualidades de las contribuciones de la investigación y puede llevar a consecuencias indeseadas (Hazelkorn, 2015, p. 13, traducción propia).

La autora señala que existe una dificultad considerable para obtener indicadores significativos y datos comparativos (a niveles nacionales e internacionales). La adopción de clasificaciones como los *rankings* sirve para incorporar una cultura de métricas que reduce la importancia de otras características de la investigación o de la enseñanza de calidad, las que no se pueden capturar fácilmente con números. Esta es una perspectiva compartida por el grupo de expertos de la Comisión Europea sobre evaluación de la investigación para quienes “consecuencias indeseadas pueden ocurrir cuando los indicadores son tomados en forma aislada y son hechas correlaciones simples” (Wilsdon *et al.*, 2015, p. 75, traducción propia). Es por ello que el informe enfatiza en que un sistema de investigación maduro necesita de una geometría variable que reúna tanto criterio experto (que provea de juicios sustantivos) como indicadores cuantitativos y cualitativos. Subraya, así, que la evaluación de la investigación debe ser hecha teniendo en cuenta el contexto y la diversidad disciplinaria:

Es poco probable que un solo tamaño se ajuste a todos. (...) La calidad académica es altamente específica del contexto, y es sensato pensar en términos de calidad de investigación, en lugar de esforzarse por [alcanzar] una sola definición o medida de calidad (Wilsdon et al., 2015, p. viii, traducción propia).

Como se señala más adelante, parte del problema está en que se considera que las métricas eximen a los administradores de la investigación (*research managers*) de la responsabilidad de realizar evaluaciones basadas en información más precisa y completa. Esto contribuye a producir una serie de deformaciones en las orientaciones que el sistema pretende dar y contribuye a generar desconfianza sobre la administración de la investigación en general (Wilsdon et al., 2015). Como vienen argumentando algunos autores citados en el informe: “el control de calidad basado en la auditoría –afirma Hoecht– ha reemplazado, en la academia del Reino Unido, a la forma de control basada en la confianza, lo que puede afectar negativamente las prácticas innovadoras de enseñanza e investigación”. O, como afirma Willmott: “los mecanismos de auditoría empujan a los académicos hacia temas establecidos [*mainstream*] que tienen mayor probabilidad de ser publicados en las revistas mejor clasificadas” (Wilsdon et al., 2015, p. 85, traducción propia), lo que desestimula la creatividad y la innovación.

Por su parte, los informes complementarios muestran la capacidad limitada de los indicadores cuantitativos para predecir los resultados de la evaluación por pares. Como afirman los autores de la revisión bibliográfica: “la evaluación por pares sigue siendo considerada el principal mecanismo de control de calidad debido a la falta de alternativas serias” (Wouters et al., 2015, p. 46). A lo que más adelante agregan:

La literatura no apoya actualmente la idea de reemplazar la revisión por pares por la bibliometría. (...) No todos los campos muestran una correlación fuerte entre los datos bibliométricos y los de la revisión por pares. (...) La literatura apoya la idea de suplementar la revisión (revisión por pares informada) (Wouters et al., 2015, pp. ix, 66-67, traducción propia).

Por revisión por pares informada los autores refieren al “concepto básico de que una aplicación juiciosa de datos bibliométricos específicos puede informar el proceso de revisión por pares, dependiendo de la meta exacta y del contexto de la evaluación” (Wilsdon et al., 2015, p. 64; Wouters et al., 2015, p. 61, traducción propia). En otras palabras, no se trata de no usar indicadores, sino de usarlos cuando son adecuados y confiables, y hacerlo en forma criteriosa y como una de las variables que dan apoyo al juicio sustantivo de quien está evaluando.

A su vez, a pesar de que durante muchos años las métricas se han considerado como una posible alternativa a la revisión por pares en la evaluación de la excelencia de la investigación en el Reino Unido, el análisis de correlación entre las puntuaciones del REF 2014 –considerando artículos por autor– y las métricas –tomando en consideración un conjunto de 15 indicadores– parece demostrar lo contrario: “la correlación general entre los indicadores métricos y el perfil de calidad REF fue generalmente baja con todas las estadísticas de correlación por debajo de 0,4” (HEFCE, 2015a, p. 10). A pesar de la amplitud del análisis realizado –que consideraron casi 150.000 artículos individuales–, solamente se encontraron correlaciones débiles, siendo éstas significativamente más bajas para trabajos publicados más recientemente y mostrándose una cobertura altamente variable de métricas a través de las distintas áreas temáticas. Como afirman los autores del informe:

Este trabajo ha demostrado que las métricas individuales proporcionan resultados significativamente diferentes del proceso de revisión por pares del REF, lo que prueba que las métricas no pueden proporcionar un sustituto equivalente a la revisión por pares REF (HEFCE, 2015a, p. iii).

Otra de las conclusiones importantes de la revisión es que el uso de indicadores inapropiados crea incentivos perversos y consecuencias negativas para el sistema de investigación, para las instituciones y para los investigadores. Especialmente son señalados como inapropiados los factores de impacto de las revistas (*journal impact factors*, JIFs), los índices de citaciones (como el índice-h) y los rankings y tablas de posiciones de las universidades. Existe en toda la comunidad científica británica una preocupación legítima de que algunos de los indicadores cuantitativos que son utilizados para respaldar las decisiones en torno a la

excelencia y la calidad de la investigación pueden ser manipulados y conducir a consecuencias no deseadas. Como el peor de los ejemplos es mencionado el uso generalizado de los JIF cuando es tomado como indicador de la calidad de los artículos. Otro ejemplo prominente en un sentido similar es la contabilidad del número de citaciones que acaban estimulando prácticas perversas como los “clubes de citaciones” o las tácticas de fragmentación (*salami-slicing*) de los artículos para aumentar el número de citas y la posición de los autores. Es por eso que poner demasiado énfasis en indicadores estrechos y mal diseñados puede tener consecuencias negativas en todos los niveles (tanto institucionales como individuales). Por otra parte, como afirman los autores de la revisión bibliográfica (Informe complementario I), la opinión popular de que la tasa de citaciones es una medida de la calidad científica no está respaldada por la comunidad de expertos en bibliometría. La calidad es vista como un concepto multidimensional que no puede ser capturado por ningún indicador. Además es necesario tener en cuenta que la dimensión de la calidad que debe ser priorizada en las evaluaciones de la investigación varía según el campo y el objetivo de la investigación (Wouters *et al.*, 2015). Es por ello –remarcan los revisores– que quienes participan en la evaluación y gestión de la investigación deben actuar de manera responsable, considerando y previniendo las consecuencias negativas siempre que sea posible, particularmente en lo que concierne a la igualdad y a la diversidad entre los investigadores, las formas de investigar y el resultado de lo que se investiga.

Otros de los hallazgos que destaca el informe es que la evidencia de una relación robusta entre nuevas métricas y calidad de investigación sigue siendo muy limitada, por lo que es necesaria mayor experimentación. Además, existe preocupación en relación con la definición de impacto a través de indicadores cuantitativos porque es probable que “restrinja el pensamiento” en función de la búsqueda de temas que puedan tener mayor aceptación y potencialmente “limate la diversidad” de la base de investigación del Reino Unido. Finalmente, el informe concluye que en la actualidad el uso de indicadores cuantitativos en la evaluación de la investigación y la gestión “no puede depender de la necesidad de reducir costos o cargas administrativas” (Wilsdon *et al.*, 2015, p. viii-x, 136-140, traducción propia).

Propuesta y principales recomendaciones

El informe destaca que en los últimos cinco años hubo un debate cada vez más sofisticado en el Reino Unido, en toda Europa e internacionalmente sobre la gobernanza de la investigación y la innovación. Con esto, lo que ha quedado de manifiesto es la necesidad de contar con una mejor evidencia para respaldar las políticas en esta área. Tanto para la Comisión Europea, para los RCs del Reino Unido y para quienes financian la investigación en países como Holanda, Alemania, los Estados Unidos y Japón viene ganando popularidad el concepto de “investigación e innovación responsable” (“*responsible research and innovation*”, RRI) como marco para la gobernanza de la investigación. En su forma más simple, RRI puede ser definida como “cuidar el futuro a través de la administración colectiva de la ciencia y la innovación en el presente” (Wilsdon *et al.*, 2015, p. 134, traducción propia). Según ellos, esta perspectiva nutre a la política de investigación con un enfoque en términos de anticipación, reflexividad, deliberación, inclusividad y receptividad. Es por ello que, a partir de estas discusiones en torno a RRI, los expertos a cargo de la revisión del uso de las métricas en el Reino Unido proponen la noción de “métricas responsables” como forma de enmarcar los usos de los indicadores cuantitativos en la gobernanza, la gestión y la evaluación de la investigación. Este concepto, “métricas responsables”, se expresa en cinco dimensiones:

1. Robustez: basar las métricas en los mejores datos posibles en términos de precisión y alcance;
2. Humildad: reconocer que la evaluación cuantitativa debe apoyar, pero no suplantar la evaluación cualitativa de los expertos;
3. Transparencia: mantener las colecciones de datos y procesos de análisis abiertos y transparentes, de manera tal que aquellos que están siendo evaluados puedan probar y verificar los resultados;
4. Diversidad: dar cuenta de las variaciones por campo de conocimiento, y usar una serie de indicadores que reflejen y apoyen la pluralidad de investigaciones y de trayectorias de carrera de los investigadores en el sistema;

5. Reflexividad: reconocer y anticipar los efectos potenciales y sistémicos de los indicadores, y revisarlos en respuesta a ello (Wilsdon *et al.*, 2015, pp. x, 134-135).

El informe va a formular de manera fundamentada veinte recomendaciones para el uso de las métricas en la evaluación y gestión de la investigación a partir de los hallazgos que derivaron de todo este trabajo de revisión. Por razones de espacio, destacamos aquí sólo las de carácter más general.

En primer lugar, se recomienda apoyar el liderazgo efectivo, gobernanza y gestión de las culturas de investigación. Son las culturas de investigación y sus propios objetivos y fines los que deben prevalecer con sus lógicas de funcionamiento y organización sobre los criterios administrativo-burocráticos o *manageriales*. Por ello, es la propia comunidad de investigación la que debe desarrollar un enfoque más sofisticado y matizado sobre la contribución y las limitaciones de los indicadores cuantitativos. Por otra parte, son los dirigentes de las instituciones de educación superior los responsables de formular una declaración de principios clara, que dé cuenta de su forma de abordaje de la gestión de la investigación y la evaluación, el papel que cumplen los indicadores cuantitativos, los *rankings* y las tablas de posiciones dentro de su contexto institucional y destacando manifiestamente –en esa declaración de principios– que el contenido y la calidad de un artículo es mucho más importante que el factor de impacto de la revista en que fue publicado.

En relación con los gerentes de recursos humanos y a los paneles de reclutamiento o promoción, la recomendación es que deben ser explícitos acerca de los criterios utilizados para la contratación, permanencia y promoción. Estos criterios deben basarse en juicios de expertos y pueden reflejar tanto la calidad académica de lo producido como las contribuciones más amplias a la política, la industria o la sociedad. Los juicios pueden a veces ser guiados por métricas, si son relevantes para los criterios en cuestión y se usan de manera responsable y con la debida conciencia de sus limitaciones. Métricas como JIF no deben utilizarse. Además, los investigadores individuales deben ser conscientes de las limitaciones de los indicadores cuando presentan sus propios *CVs* y cuando evalúan el trabajo de sus colegas. Cuando los indicadores estándares son inadecuados, los investigadores individuales deben buscar otras fuentes de datos que sirvan para documentar y respaldar sus argumentos sobre el impacto de su trabajo.

La recomendación a los editores es que estos deberían reducir el énfasis en el factor de impacto de las revistas como una herramienta de promoción. Los editores deben alentar las prácticas de autoría responsable y el suministro de información más detallada sobre las contribuciones más específicas de cada autor y, a su vez, fomentar un cambio hacia la evaluación basada en la calidad académica de los artículos en lugar de los JIF. Con relación a los proveedores de datos, analistas y productores de rankings de universidades, la recomendación es que deben esforzarse por una mayor transparencia e interoperabilidad entre diferentes sistemas de medición, evitando la utilización de rankings, que permanecen opacos o que funcionan como “cajas-negras”.

Finalmente, se propone establecer un Foro para el uso de Métricas Responsables que reúna a quienes financian la investigación, a las instituciones de educación superior y organismos representativos, editores, proveedores de datos y otros para trabajar en temas como estándares de datos, interoperabilidad, apertura y transparencia. El argumento es que la comunidad de investigación del Reino Unido necesita un mecanismo para llevar adelante la agenda establecida en este informe. Se sugiere también que sea este foro el que se encargue de coordinar las respuestas del Reino Unido a las numerosas iniciativas en trono a las métricas de investigación, los estándares y la infraestructura de datos en Europa e internacionalmente. De esa forma se espera garantizar que el sistema del Reino Unido se mantenga a la vanguardia y continúe progresando efectivamente en este tema, apoyando la investigación de la manera más inteligente y coordinada posible, para poder así influir en los debates en Europa y en los estándares que se seguirán en otros países.

Sobre la resaca de la marea métrica

Curiosamente, si bien Wilsdon, el presidente de la comisión independiente encargada de la revisión del uso de las métricas afirma, al final del prefacio que citamos al comienzo, que “la marea métrica está sin duda subiendo” (Wilsdon *et al.*, 2015, p. iii, traducción propia), la imagen elegida para ilustrar el informe es la de

la resaca que ha dejado la marea, con sus residuos, una vez que las olas del mar han retrocedido y dejado la playa vacía después de la crecida. Si se ponderan los hallazgos, las propuestas y las recomendaciones del informe, esta imagen no parece, sin embargo, deberse a un descuido que resultó parojo. De hecho, en el informe, se reconoce explícitamente que “hay corrientes poderosas que azotan la marea métrica” (Wilsdon *et al.*, 2015, p. 136, traducción propia) y se incluye un listado que va desde las presiones crecientes para la auditoría y evaluación del gasto público en educación e investigación; las demandas de los responsables de la formulación de políticas para el uso de una estrategia más inteligente en relación a la calidad de la investigación y su impacto; la necesidad de que las instituciones gestionen y desarrollen sus estrategias de investigación; la competencia dentro y entre las instituciones de prestigio, los estudiantes, el personal y los recursos; y el aumento en la disponibilidad de “big data” en tiempo real. En el actual contexto es claro que el número y la variedad de las herramientas para identificar, analizar y evaluar la información de la investigación está aumentando. Es, innegablemente, un área en rápido movimiento. Sin embargo, como ha dejado en evidencia esta revisión, existe el peligro de apresurarse y sobre-interpretar los datos disponibles causando así graves perjuicios y distorsiones para los objetivos del sistema, las instituciones que lo componen y los propios individuos que son sus actores directos. Esta es, entendemos nosotros, la resaca que es necesario evitar que deje la marea métrica si ésta, efectivamente, va a continuar subiendo.

En un pasaje de una entrevista concedida con motivo del lanzamiento del informe, Wilsdon habla sobre los peligros de la gestión a través de las métricas y el creciente uso de los *rankings* y tablas de posiciones en los siguientes términos:

Es un ejemplo clásico de cómo el sector [científico-académico] ha, en un sentido, tercerizado algunos de los temas fundamentales sobre la estrategia, la dirección y la gestión (...). El uso extendido de los Factores de Impacto de las Revistas [JIF] está creando grandes *efectos negativos y perversos* en el sistema de investigación en términos de forzar a los investigadores a seguir determinadas vías de publicación, ahogando el sistema de investigación al reforzar el poder de las revistas de alto impacto en detrimento de otras y los efectos a largo plazo de esto van a ser todavía más corrosivos para el tipo de sistema de investigación vibrante y diverso que queremos tener. (...) Sí, efectivamente, los Factores de Impacto de las Revistas [los JIFs] se han convertido en *un tipo de métrica realmente tóxica*, como una metáfora mezclada con la de la marea métrica. Las métricas han llegado a jugar un papel tóxico en muchos aspectos en el sistema de investigación y queremos que esto termine (HEFCE, 2015b –nuestro énfasis, traducción propia).

Lo que deja claro esta revisión es que después de la aplicación y de la utilización, durante años, entusiasta y hasta fervorosa, como mostramos al inicio– de métricas e indicadores, se produjo una suerte de “intoxicación” por el uso excesivo, y esto ha producido innegables efectos “corrosivos” para el sistema de investigación. Llamativamente, se converge aquí en otra de las acepciones de la palabra “resaca” para describir los resultados del mismo proceso.

4. 2018: *Turning culture, la contramarea y el concepto de “métricas responsables”*

En febrero de 2018 tuvo lugar un evento con el título “The turning tide: a new culture of research metrics”, organizado por el *Forum for Responsible Research Metrics* (Foro del Reino Unido para Métricas de Investigación Responsable, FFRRM). Este foro fue creado en setiembre de 2016 en cumplimiento de las recomendaciones hechas por el informe *The metric tide*. El mismo contó con la participación de más de 140 asistentes, delegados de una amplia variedad de instituciones e interesados vinculadas con esta temática.¹³ El foco del evento fue la necesidad de reflexionar sobre “la cultura existente del uso de métricas” debido a “los usos y los abusos” de ellas en la evaluación de la investigación en los últimos años. La conclusión general fue la necesidad de “un cambio cultural”: “A partir del evento, queda claro que el sector de educación superior del Reino Unido desea cambiar la cultura existente, pero que la implementación del cambio es multifacética y desafiante” (Forum for Responsible Research Metrics, 2018a, pp. 4, 10).

Las conclusiones de “*The turning tide*” fueron lanzadas por el FFRRM en un informe, en julio de 2018, en el Euroscience Open Forum 2018 (ESOF) que tuvo lugar en Toulouse, Francia. Bajo el título: “*UK*

Progress towards the use of metrics responsibly. Three years on from The metric tide report”, allí se señala que:

Las discusiones en el evento demostraron que, aunque las instituciones del Reino Unido están dispuestas a cambiar la cultura existente (eliminando la dependencia de las métricas de las revistas y utilizando las métricas de manera responsable), muy pocas de ellas saben cómo implementar el cambio en la práctica (FFRRM, 2018a, p. 8).

Los temas que se trataron giraron en torno a entender y medir mejor lo que es realmente valorado en la investigación; la importancia de la transparencia en los procesos evaluativos; incentivar el “buen” comportamiento de los investigadores; desafiar la cultura del prestigio basada principalmente en la práctica de publicación y garantizar que sean consideradas las distintas etapas de la carrera de investigación, así como las diferencias entre las disciplinas.

En uno de los paneles, “Desafíos y soluciones: creando una cultura que use las métricas de manera responsable”, se destacó que:

Las instituciones y los financiadores necesitan incentivar y apoyar a los académicos para cambiar la forma en que piensan acerca de las publicaciones. (...) Las competencias de aquellos que implementan e interpretan las métricas necesitan ser probadas y mejoradas. (...) Los investigadores deben ser reconocidos por participar en la revisión por pares para incentivar a las personas a tomar el tiempo necesario para considerar los elementos cualitativos y cuantitativos de la evaluación de manera integral (FFRRM, 2018a, p. 10).

En otro de los paneles, titulado “La perspectiva de los investigadores”, se señaló que las partes interesadas deben ser sensibles al impacto potencial que tienen las métricas sobre los individuos. A este respecto se enfatizó que:

Las métricas son influyentes y tienen el potencial de afectar a las personas, sus carreras, su salud mental y su bienestar. Si se utilizan como un atajo a la evaluación de la investigación y de los investigadores, tienen el potencial de ser dañinas para los individuos, por lo que la evaluación holística sigue siendo importante. (...) Las métricas tienen que medir lo que valoramos. El sector [científico-académico] necesita entender lo que se valora en la investigación y en los investigadores, para que las métricas se puedan centrar en medir lo que importa en lugar de centrarse en lo que se puede medir. Investigar es una tentativa de hacer algo, y necesitamos apoyar el “esfuerzo exploratorio” en la forma en la que medimos la investigación (FFRRM, 2018a, p. 10).

Sino, se argumenta más adelante, “las métricas pueden (...) alentar a los investigadores individuales a centrarse en el trabajo incremental de menor riesgo” (FFRRM, 2018a, p. 19). Y se hace referencia aquí a un artículo publicado en la Revista Nature por dos miembros del University Medical Center Utrecht, en Holanda, con el muy elocuente título de “*Fewer numbers, better science*” en el que relatan el proceso de cómo en ese centro emprendieron acciones para “moverse más allá de las métricas” y “liberarse de la ‘mentalidad bibliométrica’” (Benedictus, Miedema & Ferguson, 2016, p. 454).

Finalmente, en julio de 2018, el Foro para Métricas de Investigación Responsable publicó un documento con recomendaciones para el Marco de Excelencia en Investigación (REF) 2021 (FFRRM, 2018b). Todas las recomendaciones van en la misma línea de destacar la necesidad de producir un cambio de cultura en el uso de las métricas para evitar los efectos perversos que, su mal uso, ha ocasionado en el sector científico-académico británico.¹⁴

5. 2022: *The Metric Tide Revisited* y la necesidad de reducir la burocracia

En mayo de 2022, *UK Research and Innovation* anunció una revisión del papel de las métricas en la evaluación y gestión de la investigación. Con el título “*The metric tide revisited*”, esta revisión busca analizar en forma breve, precisa y basada en evidencias los usos actuales y potenciales de las métricas. Para ello ha convocado a un panel de expertos para examinar las conclusiones y recomendaciones dadas en 2015 y para que se evalúen los progresos obtenidos en estos siete años. También se espera que esta revisión ofrezca asesoramiento actualizado a *UK Research and Innovation* y a los organismos de financiación de la educación superior del Reino Unido sobre formas más eficaces de apoyar e incentivar la evaluación

responsable de la investigación y los usos responsables de las métricas. La conclusión de esta revisión estaba prevista para mediados de setiembre de 2022, pero hasta la fecha (14 de noviembre) no han sido publicados aún sus resultados. Algunos comentarios que han trascendido señalan por ejemplo, que en opinión de Wilsdon –que forma parte del panel de expertos a cargo de la revisión– *The metric tide* tenía un enfoque excesivamente *managerial*. Para él, hoy es necesario centrarse en cambiar la cultura hacia procesos de investigación más saludables.

Otra importante iniciativa en curso en el Reino Unido tiene que ver con la reducción de la burocracia en procesos vinculados con la evaluación y gestión de la ciencia. Ha quedado comprobado que la carga administrativa que recae sobre los investigadores es cada vez más elevada, lo que resta tiempo, desvía la atención, obstaculiza la investigación y, en última instancia, hace un uso ineficiente de los recursos destinados a la misma. A este respecto, fue publicada en julio pasado una revisión independiente de la burocracia en la investigación con sugerencias para liberar a los investigadores de la carga burocrática innecesaria y apoyarlos para que puedan centrar su actividad en la investigación (Tickell, 2022).

6. La experiencia británica no es sólo británica: de las “métricas responsables” a la “evaluación responsable de la investigación”

¿Qué enseñanzas podemos extraer de la experiencia británica? En primer lugar, que la experiencia británica no es la única ni la primera en iniciar una reflexión profunda respecto de los efectos causados en la ciencia por el uso de las métricas. Existe en la actualidad un debate en todo el mundo respecto de cómo evaluar la investigación sin que con esto se afecte su calidad y se distorsionen sus objetivos. En diciembre de 2012, un grupo de editores de revistas científicas se juntaron durante la Reunión Anual de la Sociedad Americana de Biología Celular en San Francisco y firmaron la Declaración sobre Evaluación de la Investigación conocida como *Declaration on Research Assessment DORA 2013*. Los firmantes piden que la investigación se evalúe por sus propios méritos y que se ponga fin al uso de los factores de impacto de las revistas en las decisiones de financiamiento, contratación y promoción. Para esto, apoyan la adopción de 18 recomendaciones que van dirigidas a los diferentes actores del sistema: las agencias de financiamiento, las instituciones de investigación, los editores de revistas científicas, las organizaciones que proveen las métricas y los investigadores.¹⁵ En junio de 2018, *DORA* publicó un plan estratégico por dos años para el avance global de la evaluación de la investigación a nivel institucional, nacional y de los agentes financieros (*Declaration on Research Assessment*, 2018). A la fecha, el número de adherentes a *DORA* en todo el mundo es de más de 22.000 individuos y organizaciones en 159 países. Justamente, una de las recomendaciones enfatizadas en *The metric tide* para las instituciones del Reino Unido es adherir a *DORA*, lo que los Consejos de Investigación (RCs) de ese país hicieron en febrero de 2018. A partir de noviembre de 2022, *DORA* pide a las organizaciones signatarias que compartan con sus comunidades una declaración pública en la que detallen su compromiso con *DORA* y con la evaluación responsable de la investigación, así como que escuchen a sus comunidades y las mantengan al día a medida que avanza la aplicación de los principios de la declaración.¹⁶

En 2013 también tomó forma un movimiento de investigadores de los Países Bajos con el nombre de “Ciencia en Transición” (*Science in transition*). Sus iniciadores creen que la ciencia necesita de una reforma fundamental. Por eso se agruparon con el objetivo de abordar los problemas sistémicos de la investigación porque la cultura académica se ha convertido en un sistema autoreferencial en el que la calidad es medida principalmente en parámetros bibliométricos y donde se subestima la relevancia social (Dijstelbloem, Huisman, Miedema & Mijnhardt, 2013).¹⁷

Por otra parte, un grupo de destacados expertos en cienciometría liderados por Diana Hicks (*Georgia Institute of Technology*) y Paul Wouters (Leiden University) propusieron 10 principios para el uso de indicadores cuantitativos en la evaluación de la investigación que fueron publicados en 2015 en la Revista *Nature*. Allí destacan que la evaluación de la investigación se ha convertido en una rutina y, a menudo, se basa cada vez más en métricas y no en las valoraciones de expertos. Como resultado de ello, los

procedimientos que fueron diseñados para aumentar la calidad de la investigación ahora amenazan con dañar el sistema científico, por lo que “no debe permitirse que la información cuantitativa se convierta en un fin en sí misma”. Y agregan: “un solo indicador es susceptible de crear comportamientos estratégicos y substitución de objetivos” (Hicks, Wouters, Waltman, de Rijcke y Rafols, 2015, pp. 429-430). El conocido como “Linden Manifesto for Research Metrics” ha sido traducido a 25 idiomas y existe una versión en video que resume en pocos minutos los 10 principios del manifiesto.¹⁸

En 2016 fue creado el Grupo de Expertos de la Comisión Europea en Altmetrics. En 2017, este grupo publicó un informe con el título “*Next-generation metrics: responsible metrics and evaluation for open science*”, en el que incorpora explícitamente el concepto de “métricas responsables” y da una serie de 12 recomendaciones específicas para la Agenda Europea de Ciencia Abierta (European Commission, 2017). En enero de 2019, un grupo de expertos sobre el futuro de la publicación y la comunicación académica, presidido por Jean-Claude Guédon, publicó un informe en donde se insta a las instituciones de investigación, a la comunidad científica y a los investigadores a incorporar las recomendaciones de *DORA* y del Manifiesto de Leiden. Además, concluyen que las deficiencias actuales en el sistema de publicación académica (estructurada en torno de ránquines e indicadores de impacto) sólo puede cambiar de manera significativa si los organismos de financiamiento toman la iniciativa e inician el cambio (European Commission, 2019). En este sentido, *Science Europe*, la asociación de financiadores públicos de la investigación científica en Europa que cuenta con 40 miembros provenientes de 30 países europeos, ha adoptado como una de sus prioridades el tema de la evaluación científica. Después de impulsar un estudio sobre las prácticas de investigación en 2019¹⁹ y promover en 2020 un conjunto de recomendaciones destinadas a las organizaciones de investigación de los países miembros,²⁰ en el primer semestre de 2022 llevó a cabo un ejercicio que recogió los intereses y perspectiva de más de 350 organizaciones de más de 40 países. Esto condujo a la publicación de un “Acuerdo sobre la Reforma de la Evaluación de la Investigación” que sienta las bases para una coalición de organizaciones dispuestas a llevar adelante estos cambios.²¹

Como podemos ver, para mantener la confianza en la ciencia, se ha vuelto imperioso cambiar la cultura del “*publish or perish*”, es decir, dejar de valorar la cantidad por encima de la calidad. Con este objetivo fueron formulados los “Principios de Hong Kong para la evaluación de investigadores” (Plackett, 2020). En junio de 2019, en esa ciudad, tuvo lugar el 6to. Congreso Mundial sobre integridad en la investigación. En ese ámbito, fueron discutidos y definidos cinco principios con la intención de ayudar a las instituciones de investigación a minimizar los incentivos perversos que incitan prácticas de investigación cuestionables y, a la vez, de reconocer y premiar la investigación fiable. Estos principios son: 1) valorar prácticas de investigación responsables, 2) valorar reportes completos y transparentes de las investigaciones, 3) recompensar las prácticas de ciencia abierta (investigación abierta), 4) valorizar la diversidad de tipos y actividades de investigación y 5) reconocer todas las demás contribuciones esenciales a la investigación y la actividad académica tales como la revisión por pares y la enseñanza.²² Hasta el momento han recibido el aval de 25 instituciones y 196 individuos de distintas partes del mundo.

Otra institución relevante a nivel internacional que ha adherido a la causa de promover una evaluación responsable de la ciencia es el *Global Research Council* (GRC). Creado en 2012, es una organización virtual formada por los responsables de las agencias nacionales de financiación de la ciencia y de la ingeniería. Su objetivo es promover las mejores prácticas y la colaboración entre organismos de financiamiento de todo el mundo. Está formado mayoritariamente por organismos públicos que financian la investigación y hay en la actualidad aproximadamente 120 organizaciones que participan activamente en el GRC. Su Junta Directiva la integran 12 directores de consejos nacionales de investigación.²³ En 2020 el GRC organizó una conferencia virtual sobre “evaluación responsable de la investigación” (*Responsible Research Assessment, RRA*). Previo a ese encuentro, y como disparador de las discusiones, fue presentado un documento de trabajo –cuyos autores son también algunos de los autores del informe *The metric tide*– en el que se define este concepto como “un término que engloba enfoques de evaluación que incentivan, reflejan y recompensan las características plurales de la investigación de alta calidad, en apoyo a las culturas

de investigación diversas e inclusivas” (Curry *et al.*, 2020, p. 4). En 2021, fue creado el grupo de trabajo del GRC sobre Evaluación Responsable de la Investigación con el objetivo de defender la importancia del RRA y dar orientación y apoyo a las organizaciones participantes para la incorporación de los principios del RRA en sus prácticas, así como en las de las organizaciones que financian.

Finalmente, a nivel regional fue constituido en 2019 el Foro Latinoamericano sobre Evaluación Científica (FOLEC) en el ámbito del Consejo Latinoamericano de Ciencias Sociales (CLACSO). En los últimos años el FOLEC viene publicando una serie de diagnósticos, herramientas y declaraciones con el fin de ofrecer directrices específicas para la región sobre evaluación de la investigación.²⁴ En junio de 2022 fue aprobada una declaración de principios en la XXVII Asamblea General de CLACSO, en México, con el significativo título de “Una nueva evaluación académica y científica para una ciencia con relevancia social en América Latina y el Caribe.” También desde 2022, el FOLEC integra la Junta Ejecutiva de *DORA*.

Así, podemos apreciar que existen varias experiencias semejantes a la británica en otras partes del mundo. Un caso también notable es el de China donde se ha iniciado recientemente una reforma radical de la evaluación de la investigación. Ésta, en primer lugar, le da “el adiós al ‘culto al SCI’” (*Science Citation Index*) de Web of Science y se propone “reestablecer el espíritu científico”, “promover el regreso de las universidades a sus objetivos académicos originales” y, en definitiva, “restablecer los valores científicos y sociales originales para que guíen la investigación en China” (Zhang & Silvertsen, 2020, pp. 1, 7). Esto contrasta fuertemente con el caso de España donde el sistema de evaluación científica imperante todavía premia preferentemente las publicaciones en revistas bien posicionadas en los *Journal Citation Reports* y/o *Journal Ranks* y los investigadores reclaman por “un cambio radical del sistema de evaluación científica” (Delgado López-Cózar & Martín-Martín, 2022, p. 26; Delgado López-Cózar, Ràfols & Abadal, 2021).

En términos generales, puede apreciarse a través de este rápido y no exhaustivo recorrido cómo el término “métricas responsables”, acuñado por *The metric tide* en 2015, se ha sido difundido y adoptado a nivel global. Lo llamativo, sin embargo, es que en este proceso el concepto se amplió al de “evaluación responsable de la investigación” (RRA), poniendo énfasis en la importancia de que la ciencia no pierda fiabilidad –por lo que su calidad más que su cantidad deben ser privilegiadas– y, fundamentalmente, que ésta tenga relevancia social, que admita la diversidad en sus prácticas y enfoques y que estimule la colaboración y la solidaridad entre quienes producen conocimiento.

Consideraciones finales

El cambio cultural es un cambio fundamental de creencias,

no sólo un cambio en las reglas del juego.²⁵

Vimos entonces que, en primer lugar, la experiencia británica no es solamente británica y que la preocupación por el mal uso de las métricas se extiende a muchos otros lugares del sistema científico mundial. En segundo lugar, que la propuesta del concepto de “métricas responsables” trajo consigo un claro reconocimiento de la forma irresponsable con que han venido siendo utilizadas las métricas para la evaluación de la investigación. Esto ha conducido, por otra parte, a reconocer que la propia evaluación científica se ha estado llevando a cabo de manera irresponsable, haciendo uso de artefactos metodológicos no consistentes ni adecuados para ponderar la calidad de lo producido y su relevancia social por sobre la mera cantidad. Esto, como notamos al inicio, comenzó hace años con el auge de las culturas de auditoría, la difusión del concepto de *accountability* y el uso fervoroso de indicadores propiciados por la implementación del *New Public Management* en el sector científico académico. La evidencia sobre los efectos dañinos que esto ha tenido para el sistema, las instituciones y los individuos (comportamiento estratégico, desplazamiento de los objetivos, desestímulo a las actividades de enseñanza y a las prácticas innovadoras de investigación, etc.) es, justamente, sobre la que se basa, como hemos podido mostrar, la revisión independiente llevada a cabo en el Reino Unido en 2015. Y, como hemos señalado, de la noción más acotada de “métricas responsables” acuñada allí, se ha pasado, a nivel global y en años recientes, al

concepto más amplio de “evaluación responsable de la investigación” (*RRA*), que muestra una toma de conciencia sobre otras dimensiones importantes a considerar al momento de evaluar como son la calidad, la integridad, la diversidad, la equidad y la relevancia social.

En un plano más general, lo que vemos es que se ha venido imponiendo por encima del *ethos* científico-académico una racionalidad *managerial* (o *managerialismo*) que privilegia sus objetivos de “eficiencia” y “productividad” (medidos en números y, por tanto, fácilmente cuantificables) por sobre los fines últimos de la ciencia –siempre más complejos y diversos– en sus diferentes dominios. Dicho en otros términos, las culturas de investigación han quedado subsumidas a criterios administrativo-burocráticos que siguen una lógica *managerial* importada desde el sector privado. Justamente, la preocupación de fondo que manifiestan todas estas iniciativas recientes a las que hemos hecho alusión es cómo proteger a las culturas de investigación de esa racionalidad *managerial*. ¿Cómo hacer para que las necesidades concretas de la gestión y de la evaluación que tienen a su cargo los organismos responsables de promover la investigación creativa, innovadora y de calidad, no acaben conspirando contra sus propios objetivos y su propia razón de ser? ¿Cómo hacer para que la ciencia no se vuelva una simple reproducción de si misma que termine no haciendo aportes significativos para la sociedad? O, en otras palabras: ¿cómo hacer para que la práctica de la evaluación académica y el uso de términos como “culturas evaluativas” no acaben convirtiéndose en eufemismos de las “culturas de la auditoría permanente” promovida por doctrinas neoliberales como la del *New Public Management* que tanto daño le han hecho al *ethos* científico-académico? ¿Cómo hacer, entonces, para que los organismos que regulan el sistema científico-tecnológico no se conviertan en una maquinaria esterilizante de lo que deberían promover, proteger y amparar?

Estas son las preguntas que quedan en abierto a partir de la experiencia británica y la de varios otros países que hemos revisado aquí. Lo que está claro es que si no hay un cambio de cultura en el uso que se le da a las métricas para la evaluación de la investigación y en el sentido que se le da a los propios proceso de evaluación –un cambio de cultura que implique una verdadera transformación en lo que se cree y en lo que se valora–, y lo que se plantea son apenas cambios más o menos “cosméticos” en las reglas del juego, la marea métrica amenaza dejar una resaca con efectos nocivos y consecuencias potencialmente devastadoras para la ciencia y también para la sociedad.

Agradecimiento:

quiero agradecer a Maximiliano Salatino no apenas por su atenta lectura de este manuscrito y sus sugerencias, sino por la rica interlocución que se inició cuando un libro sugerido por él (Beer, 2016) me llevó al hallazgo fortuito de la referencia al informe *The metric tide*.

Referencias

- Arbix, G., Salerno, M. S., Toledo, D., Miranda, Z. & Álvarez, R. (2010). *Inovação: estratégias de sete países*. Brasília: ABDI.
- Beer, D. (2016). *Metric power*. London: Palgrave Macmillan.
- Benedictus, R., Miedema, F. & Ferguson, M. W. (2016). Fewer numbers, better science. *Nature*, 538, 453-455. <https://doi.org/10.1038/538453a>
- Colquhoun, D. (1 de diciembre de 2014). Publish and perish at Imperial College London: the death of Stefan Grimm [mensaje en el blog *DC's Improbable Science*]. Recuperado de <http://www.dcsscience.net/2014/12/01/publish-and-perish-at-imperial-college-london-the-death-of-stefan-grim>
- Curry, S., de Rijcke, S., Hatch, A., Pillay, D., van der Weijden, I. & Wilsdon, J. (2020). *The changing role of funders in responsible research assessment: progress, obstacles & the way ahead*. RoRI Working Paper N° 3. <http://doi.org/10.6084/m9.figshare.13227914>
- Declaration on Research Assessment (DORA). (2018). *DORA roadmap: a two-year strategic plan for advancing global research assessment reform at the institutional, national, and funder level*. Recuperado de <https://sfdora.org/2018/06/27/dora-roadmap-a-two-year-strategic-plan-for-advancing-global-research-assessment-reform-at-the-institutional-national-and-funder-level/>
- Delgado López-Cózar, E. & Martí-Martín, A. (2022). Detectando patrones anómalos de publicación científica en España: más sobre el impacto del sistema de evaluación científica. *ResearchGate* [Preprint, Septiembre 2022]. Recuperado de <https://www.researchgate.net/publication/363535388>
- Delgado López-Cózar, E., Ràfols, I. & Abadal, E. (2021). Letter: A call for a radical change in research evaluation in Spain. *El profesional de la información*, 30(3), 2-30. <https://doi.org/10.3145/epi.2021.may.09>
- Dijstelbloem, H., Huisman, F., Miedema, F. & Mijnhardt, W. (2013). *Why science does not work as it should? And what to do about it*. Recuperado de <http://www.scienceintransition.nl/app/uploads/2013/10/Science-in-Transition-Position-Paper-final.pdf>
- European Commission, Directorate General for Research and Innovation. (2019). *Future of scholarly publishing and scholarly communication: report of the Expert Group to the European Commission*. Publications Office. Recuperado de <https://data.europa.eu/doi/10.2777/836532>
- European Commission, Directorate-General for Research and Innovation, Peters, I., Frodeman, R. & Wilsdon, J. (2017). *Next-generation metrics: responsible metrics and evaluation for open science*. Publications Office. Recuperado de <https://data.europa.eu/doi/10.2777/337729>
- Forum for Responsible Research Metrics (FFRRM). (2018a). *UK Progress towards the use of metrics responsibly. Three years on from The metric tide report*. Recuperado de <https://dera.ioe.ac.uk/31945/>
- Forum for Responsible Research Metrics (FFRRM). (2018b). *FFRRM's advice to the Research Excellence Framework (REF) 2021 panels*, 26 jul. 2018.
- Hazelkorn, E. (2015). *The Obsession with rankings in tertiary education: implications for public policy*. Recuperado de https://hepru.files.wordpress.com/2015/01/the-obsession-with-rankings-in-tertiary-education_wb_0115.pdf
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520, 429-431. <https://doi.org/10.1038/520429a>
- Higher Education Funding Council for England (HEFCE). (2015a). *The metric tide: correlation analysis of REF2014 scores and metrics. Supplementary report II to the independent review of the role of metrics in research assessment and management*. <https://doi.org/10.13140/RG.2.1.3362.4162>

- Higher Education Funding Council for England (HEFCE). (2015b). *The metric tide: report of the independent review of the role of metrics in research assessment* [entrevista de Steven Hill a James Wilsdon, 9 jul. 2015]. Recuperado de <https://www.youtube.com/watch?v=hhci90s-WPw>
- López Ruiz, O. J. (2010). As inovações institucionais no Reino Unido. En G. Arbix, M. S. Salerno, D. Toledo, Z. Miranda & R. Álvarez (Eds.). *Inovação: estratégias de sete países* (pp. 246-280). Brasília: ABDI.
- 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), 1-14. <https://doi.org/10.1371/journal.pbio.3000737>
- Muller, J. (2018). *The tyranny of metrics*. Princeton: Princeton University Press.
- Plackett, B. (2020). Five better ways to assess science. Hong Kong principles seek to replace “public or perish” culture. *Nature index*. Recuperado de <https://www.nature.com/nature-index/news-blog/five-better-ways-to-assess-science-research-metrics>
- Salatino, M. & López Ruiz, O. (2021). El fetichismo de la indexación. Una crítica latinoamericana a los regímenes de evaluación de la ciencia mundial. *Revista iberoamericana de ciencia, tecnología y sociedad (CTS)*, 16(46), 73-100. Recuperado de <http://www.revistacts.net/contenido/numero-46/el-fetichismo-de-la-indexacion-una-critica-latinoamericana-a-los-regimenes-de-evaluacion-de-la-ciencia-mundial/>
- Shore, C. (2008). Audit culture and liberal governance: universities and the politics of accountability. *Anthropological theory*, 8(3), 278-298. <https://doi.org/10.1177/1463499608093815>
- Shore, C. & Wright, S. (2000). Coercive accountability: the rise of audit culture in higher education. En M. Strathern (Ed), *Audit cultures: anthropological studies in accountability, ethics and the academy* (pp. 57-89). London: Routledge.
- Strathern, M. (Ed.). (2000). *Audit cultures: anthropological studies in accountability, ethics and the academy*. London: Routledge.
- Tickell, A. (2022). *Independent review of research bureaucracy*. Recuperado de <https://www.gov.uk/government/publications/review-of-research-bureaucracy>
- Wilsdon, J., Allen, L., Belfiore, E. Campbell, Ph., Curry, S.,... 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., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S.,... Franssen, T. (2015). *The metric tide: literature review (Supplementary report I to the independent review of the role of metrics in research assessment and management)*. <https://doi.org/10.13140/RG.2.1.5066.3520>
- Zhang, L. & Sivertsen, G. (2020). The new research assessment reform in China and its implementation. *Scholarly assessment reports*, 2(1), 1-7. <https://doi.org/10.29024/sar.15>

Notas

1 Sobre el *New Public Management* en Gran Bretaña y el auge, con este, de la cultura de la auditoría vinculada a lo que se ha dado en llamar “nuevo *managerialismo*”, cf. Shore y Wright (2000, pp. 63-67) y Shore (2008, pp. 287-289).

2 Entre otro, entrevistamos a funcionarios del gobierno (*Prime Minister's Strategy Unit, Cabinet Office, Department of Trade and Industry*), representantes de universidades y organismos de investigación (*Begbroke Science Park, Oxford University; Manchester Institute of Innovation, Manchester University; Arts and Humanities Research Council; etc.*) y a representantes de la industria (*Confederation of British Industries*).

3 Estos fueron: Estados Unidos, Canadá, Finlandia, Francia, Irlanda, Japón y Reino Unido.

4 La entrevista en cuestión fue realizada en la ciudad de Bristol el 5 de junio de 2007. El 1º de abril del 2018, los siete consejos de investigación del Reino Unido se unieron a otros dos consejos, *Research England y Innovate UK*, para formar *UK Research and Innovation*.

5 Desde el 1 de abril de 2018, sus funciones fueron absorbidas por el nuevo ente regulador de la educación superior en Inglaterra, *Research England*, dentro del *UK Research and Innovation*.

6 Para la lista completa de los miembros de esta comisión independiente, véase Wilsdon *et al.* (2015, pp. v-vi).

7 Los objetivos, en término de cantidad de dinero, que se le fijan a cada investigador, programa o departamento de fondos que debe conseguir atraer para el financiamiento de sus investigaciones.

8 Wilsdon cita aquí a Lawrence, P. A. (2007). The mismatch of science. *Current biology*, 17(15), 583-585.

9 Sobre el poder que ejercen las métricas, cf. Beer (2016) y Muller (2018), dos libros recientes que analizan este fenómeno y toman entre sus referencias al informe *The metric tide*.

10 Ejercicio de evaluación nacional de la investigación en las instituciones británicas de educación superior realizado cada siete años. El último de estos ejercicios fue realizado en 2021 y tomó las recomendaciones de *The metric tide*.

11 Es interesante notar que Elsevier es una de las cinco empresas oligopólicas editoriales que orientan a través de criterios *mainstream* la producción y circulación del conocimiento científico a nivel global. Sobre el negocio de la comunicación científica y cómo éste impacta en las formas en que la ciencia produce y comunica su conocimiento, cf. Salatino y López Ruiz (2021).

12 Para la lista completa, véase Wilsdon *et al.* (2015, pp. 142-147).

13 Entre los asistentes había encargados de la formulación de políticas, vicerrectores de universidades, académicos, gerentes de investigación, especialistas en bibliometría, secretarios de investigación, proveedores de datos, periodistas e interesados en el uso de métricas en la investigación (FFRRM, 2018a, p. 8).

14 El adjetivo “perverso” aparece utilizado varias veces en *The metric tide*: 1) en el sentido de la creación de “incentivos perversos” a través del uso de métricas e indicadores inapropiados (Wilsdon *et al.*, 2015, pp. viii, 132, 138), 2) en el sentido del riesgo de fomentar “comportamientos perversos” dentro y entre las instituciones por el énfasis excesivo en un pequeño conjunto de indicadores (Wilsdon *et al.*, 2015, p. 76), 3) en el sentido de la necesidad de evitar “efectos perversos” a través de utilización de rankings que no sea totalmente transparentes (Wilsdon *et al.*, 2015, p. 76) y, finalmente, 4) en el sentido de “culturas de auditoría perversas” (Wilsdon *et al.*, 2015, p. 87).

15 Se puede encontrar la declaración con las recomendaciones traducida a 28 idiomas y ejemplos de buenas prácticas en evaluación de la investigación en <https://sfdora.org>

16 Sobre la nueva política de *DORA* sobre el compromiso y la participación de las organizaciones firmantes, https://sfdora.org/wp-content/uploads/2022/11/Engagement-and-outreach-policy-for-DORA-organizational-signers_Approved-2022.pdf

17 En el sitio web de *Science in transition* pueden encontrarse ejemplos de otras iniciativas que van en el mismo sentido en otras partes del mundo: <https://scienceintransition.nl/en/about-science-in-transition>

18 Las versiones del manifiesto en las diferentes lenguas puede encontrarse en: <http://www.leidenmanifesto.org/translations.html>. La versión del video de poco más de 4 minutos puede verse en: <https://vimeo.com/133683418>

19 “*Science Europe study on research assessment practices. Final report*”, 27/12/2019 (<https://www.scienceeurope.org/our-resources/science-europe-study-on-research-assessment-practices/>).

20 Estas recomendaciones buscan enfatizar la importancia de la evaluación cualitativa y apoyar y complementar otras iniciativas como *DORA* y Leiden. “*Position statement and recommendations on research assessment processes*”, 9/07/2020, p. 9 <https://www.scienceeurope.org/our-resources/position-statement-research-assessment-processes/>

21 “*Agreement on reforming research assessment*”, 20/07/22 (<https://www.scienceeurope.org/our-resources/agreement-reforming-research-assessment/>).

22 La página oficial de “*World Conferences on Research Integrity*” donde se presentan estos principios puede accederse a través de <https://wcrif.org/guidance/hong-kong-principles>. La versión final de los “*Hong Kong principles for assessing researchers*” fue publicada en *PLOS biology* en julio de 2020 (Moher *et al.*, 2020).

23 Actualmente la Argentina forma parte de la Junta Directiva y está representada por la Presidenta del CONICET, la Dra. Ana Franchi: <https://globalresearchcouncil.org/about/governing-board/>

24 Como partes de la serie “Para una transformación de la evaluación de la ciencia en América Latina y el Caribe”, pueden encontrarse en <https://www.clacso.org/folec/clacso-ante-la-evaluacion/>

25 “*Cultural change is a fundamental change of beliefs, not just a change in the rules of the game*”, según Kim Huijpen (Recognition and Rewards Programme, <https://recognitionrewards.nl>, Holanda), en Webinar *DORA: “A systems approach for a systems problem: rethinking research assessment reform”*, 11/02/2022.