

2.2. VÍAS DE CAMBIO EN LA EVALUACIÓN DE LA PRODUCCIÓN CIENTÍFICA: HACIA UN PORTAFOLIO EQUILIBRADO

DANIEL SAMOILOVICH* Y PAOLA ANDREA RAMÍREZ**

“No todo lo que puede ser contado cuenta, y no todo lo que cuenta puede ser contado”¹
William Bruce Cameron

Desde la declaración de San Francisco sobre la evaluación de la investigación (DORA, 2012) y el Manifiesto de Leiden para la métrica de la investigación (2015), pocos sistemas nacionales de investigación y pocas universidades han logrado romper con las prácticas precedentes. Pese al amplio consenso que suscitan y al sentido común que animan ambas declaraciones, los cambios de este tipo requieren tiempo, inteligencia y flexibilidad. La introducción y el desarrollo de nuevos sistemas de evaluación concierne a múltiples actores —investigadores, universidades y organizaciones de investigación, agencias de financiamiento, editores de revistas científicas, etc—, todos con contextos, expectativas y funciones específicas. Es, por lo tanto, la combinación entre políticas nacionales, estrategias institucionales y actitudes personales, las que pueden promover los cambios necesarios en los diferentes aspectos de la evaluación.

La iniciativa internacional de Ciencia Abierta ha facilitado la discusión sobre los problemas actuales de la investigación científica en sus sistemas y procesos, y ha promovido el consenso general sobre la necesidad de cambios en la evaluación como aspecto fundamental para el logro de las metas comunes propuestas para el avance de la ciencia abierta en todas sus dimensiones y con el consiguiente beneficio en la calidad y relevancia de la investigación. Para facilitar la comprensión de las relaciones entre los propósitos reunidos en Ciencia Abierta y la evaluación, este artículo reproduce algunos fragmentos del *policy*

brief elaborado por los mismos autores y presenta las principales conclusiones y recomendaciones para nutrir el debate actual y proponer acciones.

En primer lugar, se enumeran cuáles son los propósitos comunes entre ciencia abierta y evaluación de la investigación. A continuación, se observan dimensiones en las que el cambio no solo es deseable, sino también posible, para luego indagar en las razones que dificultan la evolución de las prácticas institucionales. Se propone un enfoque dinámico para alcanzar un portafolio de evaluación equilibrada, seguido de un análisis de algunos casos institucionales que ilustran posibles vías de cambio. En las conclusiones se resumen los principales puntos y proponen transformaciones a nivel de sistema.

1. CIENCIA ABIERTA Y EVALUACIÓN DE LA INVESTIGACIÓN: ¿CUÁLES SON LOS PROPÓSITOS COMUNES?

Para comenzar, algunas definiciones: Ciencia Abierta se presenta, no como un fin, sino como medio para promover y coordinar la adaptación de los sistemas de investigación e innovación al nuevo entorno; desde las iniciativas promovidas por los diferentes integrantes del sistema para diferentes procesos del ciclo de transferencia de la información y del conocimiento: creación, publicación, uso, acceso, divulgación, evaluación y apropiación de la ciencia,

* Director Ejecutivo, Asociación Columbus

** Bibliotecóloga. Investigadora independiente

1. Informal Sociology: A casual Introduction to sociology thinking / W. Cameron. Random, 1963.

2. Ciencia Abierta, reporte para tomadores de decisiones. 2. ed. Paola Andrea Ramírez, Daniel Samoilovich. Paris, Montevideo : Unesco, 2019. Disponible en: <http://forocilac.org/wp-content/uploads/2019/05/2019-Policy-PapersCILAC-CienciaAbierta-29-04-2019-Final.pdf>

con sus estrategias en acceso abierto a las publicaciones y los datos, evaluación abierta y ciencia ciudadana.

Entre ellos, la evaluación se convierte en el nudo gordiano que todos queremos desatar, porque cierra el proceso con la medición de resultados e impacto, define los propósitos que enmarcan el hacer de los investigadores y determina en conjunto la valoración de las instituciones y el prestigio de los investigadores, como así también la financiación de la investigación y el desarrollo de las carreras investigativas y académicas.

Evaluación de la Ciencia. Promueve la transición de métodos exclusiva o principalmente concentrados en la publicación tradicional de resultados y las mediciones bibliométricas cuantitativas -convertidas en gold estándar por las bases de datos comerciales- a una combinación de enfoques multidimensionales: cuantitativos y cualitativos aplicados no sólo a los resultados y productos de la investigación; y adaptados a las nuevas condiciones de la comunicación científica, basados en la apertura de los insumos, procesos y resultados usados en las mediciones. Actualmente esto incluye: desarrollo y validación abierta de métricas confiables de publicaciones y datos, evaluación abierta de pares, ampliación y actualización de las mediciones de calidad e impacto de los sistemas de investigación, y adopción de nuevas modalidades de reconocimiento y valoración de los investigadores y grupos de investigación.

44 **Condiciones.** Se requiere un cambio significativo en la perspectiva actual con la que investigadores y las instituciones -no solo universidades, sino también gobiernos y financiadores- asumen los procesos de evaluación de la investigación. Este es un tema política y financieramente sensible, pero más allá de importantes diferencias de contexto regional, los documentos de referencia consultados presentan un amplio consenso sobre el cuestionamiento a los modelos actuales de evaluación y las ventajas que las acciones de Evaluación Abierta ofrecen para emprender el análisis y actualización de los sistemas actuales.”

2. ¿QUÉ ES LO QUE PODEMOS CAMBIAR? ALGUNOS PUNTOS DE PARTIDA

“Hay cosas que pueden ser medidas. Hay cosas que valen la pena medir. Pero lo que puede medirse no siempre es lo que vale la pena medir, lo que se mide puede no tener relación alguna con lo que realmente queremos saber. Los costos de medir pueden ser más grandes que sus beneficios. Las cosas que se miden pueden distraer los esfuerzos de las cosas que realmente importan. Y las mediciones pueden darnos un conocimiento distorsionado-conocimiento que parece sólido pero que en realidad es engañoso”
Jerry Z. Müller³

Un sistema maduro requiere de un conjunto variable de instrumentos. La calidad debe ser considerada como un concepto multidimensional que no puede ser capturado con indicadores y debe precisar cuál dimensión de calidad debe ser priorizada, lo que varía con el área disciplinaria, la misión de la organización evaluadora o usuaria de la evaluación, y el contexto geográfico. Parece claro en el actual debate que, la evaluación cuantitativa debe apoyar, no reemplazar la evaluación cualitativa por parte de expertos; que la necesaria homogenización de parámetros no puede impedir la evaluación específica; y que la pertinencia de los sistemas de evaluación y sus procesos dependen siempre de un delicado equilibrio entre finalidades, recursos, responsables y ámbitos de acción.

La evaluación por pares constituyó el mecanismo primario para evaluar la calidad de la investigación. Es un procedimiento utilizado en todo el ciclo de investigación: desde la revisión de propuestas de investigación, la publicación y evaluación de sus resultados, y la evaluación retrospectiva de resultados a nivel de sistema. El procedimiento tiene sus propias fortalezas y debilidades. Su principal fortaleza es que está basado en un conocimiento especializado de los temas analizados y metodologías empleadas, el juicio de expertos permite una comprensión detallada y matizada. Pero su subjetividad natural, es un riesgo que se convierte en debilidad cuando la independencia y relevancia de los expertos no es clara y se crea desconfianza. La revisión por pares además es lenta y costosa: los expertos son pocos, puede ser utilizada para perjudicar a competidores y requiere instrumentos de registro, sistematización y validación poco desarrollados y extendidos. Con respecto a las evaluaciones concentradas en indicadores cualitativos, reconociendo su capacidad y alcance, el consenso actual cuestiona su uso prevalente y aislado.

Algunas de estas debilidades pueden ser compensadas por una revisión por pares informada, es decir apoyada en una aplicación de datos específicos e indicadores, seleccionados en función del objetivo y el contexto preciso de la evaluación. Dos aspectos particularmente relevantes del contexto son la variación entre campos disciplinarios y la diversidad en las carreras de investigador. Muchos estudios recomiendan optar por una variedad de estrategias adaptadas a los contextos específicos que se desea evaluar, que mantengan la articulación con los indicadores internacionales.⁴ De esta manera, puede pensarse en grupos de métricas que incluyan indicadores cualitativos y cuantitativos, adecuadamente balanceados con relación al propósito, el contexto y los recursos disponibles.

La relación entre el uso de las mediciones en los sistemas de financiación y el reconocimiento y validación de la investigación científica vinculado a tales mediciones, será siempre un aspecto sensible. Las métricas alternativas

3. The tyranny of metrics / Jerry Z. Müller. Princeton : Princeton University Press, 2018. Disponible en : <https://press.princeton.edu/titles/11218.html>

4. Al respecto recomendamos: Bibliometrics for Research Management and Research Evaluation. A Brief Introduction / Ludo Waltman Ed Noyons. Leiden: CWTS, 2018. Disponible en : https://www.cwts.nl/pdf/CWTS_bibliometrics.pdf

han aumentado y se aplican y discuten actualmente en diferentes contextos pero los cambios propuestos deben evitar el desequilibrio disciplinar, sectorial y regional existente entre las capacidades de investigación y las exigencias de los sistemas nacionales e internacionales de reconocimiento y financiación. Y deben reconsiderar críticamente los propósitos de la evaluación y asumir que no todo lo importante es medible aunque sea observable, mientras que no todo lo medible es importante; para no replicar los problemas actuales de la fijación por las métricas en los nuevos modelos.

De esta manera puede pensarse en estrategias de triangulación de datos cuantitativos, análisis contextual y juicio de expertos combinado con una perspectiva de los stakeholders.⁵ La utilización de datos puede verse estimulada si son suficientemente contextualizados para permitir su interpretación. El impacto científico puede ocurrir de manera que no genera citas, por ejemplo cuando tiene una influencia sobre organizaciones no académicas, en medidas de gobierno, en la sociedad civil, en los medios de difusión o en el debate público. Este tipo de impacto puede y debe ser demostrado mediante evidencias y a ello apuntan las propuestas de métricas alternativas, que buscan visibilizar otros usos y aplicaciones de resultados aprovechando la generación de referencias y noticias de divulgación o apropiación de la ciencia en las diferentes redes sociales. Un ejemplo de esto es la llamada literatura gris, producida por organizaciones fuera de los canales tradicionales de publicación y distribución académica o comercial. Los tipos comunes de publicaciones de literatura gris incluyen informes (anuales, de investigación, técnicos, de proyectos, etc.), documentos de trabajo, documentos gubernamentales, informes y evaluaciones. Las organizaciones que producen literatura gris incluyen departamentos y agencias gubernamentales, organizaciones de la sociedad civil o no gubernamentales, centros y departamentos académicos, y empresas privadas y consultores,⁶ cuya divulgación y apropiación de los resultados de la investigación hace parte importante de la visibilidad y no está siendo considerada actualmente en toda su dimensión.

El cambio propuesto requiere de mediciones multifactoriales y multidimensionales, que incluyan parámetros cuantitativos y cualitativos, disciplinares y generales, más allá del problemático factor de impacto de revistas usado actualmente como norma pese a las críticas y objeciones.⁷ Tanto investigadores como financiadores y administradores precisan enfoques pertinentes y apropiados en todos los niveles, procesos y disciplinas, para evaluar productividad, calidad, pertinencia e impacto. Este desarrollo está

directamente relacionado con la transición hacia las modalidades de Ciencia Abierta.⁸

Además, si las agencias nacionales de investigación y entes financiadores comparten la información registrada sobre sus sistemas de investigación, y por su parte los servicios de indexación de revistas permiten acceso a los datos que sustentan sus métricas, sería posible aplicar evaluaciones comprensivas y específicas como las propuestas en el Protocolo de Leiden para Evaluación de la Investigación,⁹ cuya Matriz de Impacto¹⁰ incluye variables e indicadores cuya confiabilidad dependerá de los datos reunidos para aplicarla¹¹.

En la misma búsqueda de evaluaciones integrales y pertinentes que aprovechen las métricas disponibles y la capacidad de análisis de datos, y que sean aplicables a diferentes áreas del conocimiento e integrables internacionalmente, la International School on Research Impact Assessment (ISRIA) ha propuesto diez elementos guía para definir la evaluación: 1) contexto; 2) propósito; 3) necesidades de los involucrados; 4) vinculación entre los participantes; 5) marcos conceptuales; 6) métodos y fuentes de datos; 7) indicadores y métricas; 8) aspectos éticos y conflictos de interés; 9) comunicación; y 10) prácticas de comunidades.¹¹

La propuesta del International Development Research Centre (IDRC) de Canadá, reconoce que la capacidad de desarrollar evaluaciones integrales y pertinentes, pasa por la madurez de las instituciones o de los programas de investigación, y con base en esa experiencia y los valores propios del IDRC, presentó el Research Quality+ Assessment Framework, aproximación holística a la evaluación, que ofrece un “sistema informado para definir y evaluar la calidad de la investigación y su uso e impacto. Permite adecuarse al contexto, valores, mandatos y propósitos, y apoya los procesos de planeación, gestión y aprendizaje en cualquier etapa del proceso de investigación”. Incluye tres componentes principales: 1) Influencias principales; 2) dimensiones y subdimensiones (integridad, legitimidad, importancia y uso esperado de la investigación); y 3) protocolos de evaluación que incluyen indicadores cuantitativos y cualitativos, directamente relacionados con cada influencia principal.¹²

5. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management / James Wilsdon ed. Higher Education Funding Council for England HEFCE, 2015. Disponible en: https://responsiblemetrics.org/wp-content/uploads/2019/02/2015_metrictide.pdf

6. The tyranny of metrics / Jerry Z. Müller. Princeton : Princeton University Press, 2018. Disponible en: <https://press.princeton.edu/titles/11218.html>

7. Why the impact factor of journals should not be used for evaluating research. P. O. Seglen BMJ 1997;314:498–502. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf>

8. Roadmap on Research Assessment in the Transition to Open Science. European University Association. Bruselas, 2018. Disponible en: <https://www.eua.eu/component/attachments/attachments.html?id=348>

9. Leiden protocol for research assessments 2015-2021. Disponible en: <https://www.medewerkers.universiteitleiden.nl/binaries/content/assets/ul2staff/onderzoek/kwaliteit-en-integriteit/leiden-protocol-for-research-assessments-2015-2021-eng-only.pdf>

10. Leiden Impact Matrix. Disponible en: https://www.medewerkers.universiteitleiden.nl/binaries/content/assets/ul2staff/onderzoek/impact_matrix-medewerkers_website.pdf

11. ISRIA statement: ten-point guidelines for an effective process of research impact assessment / Paula Adam et. al. 2018. Disponible en: <https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-018-0281-5>

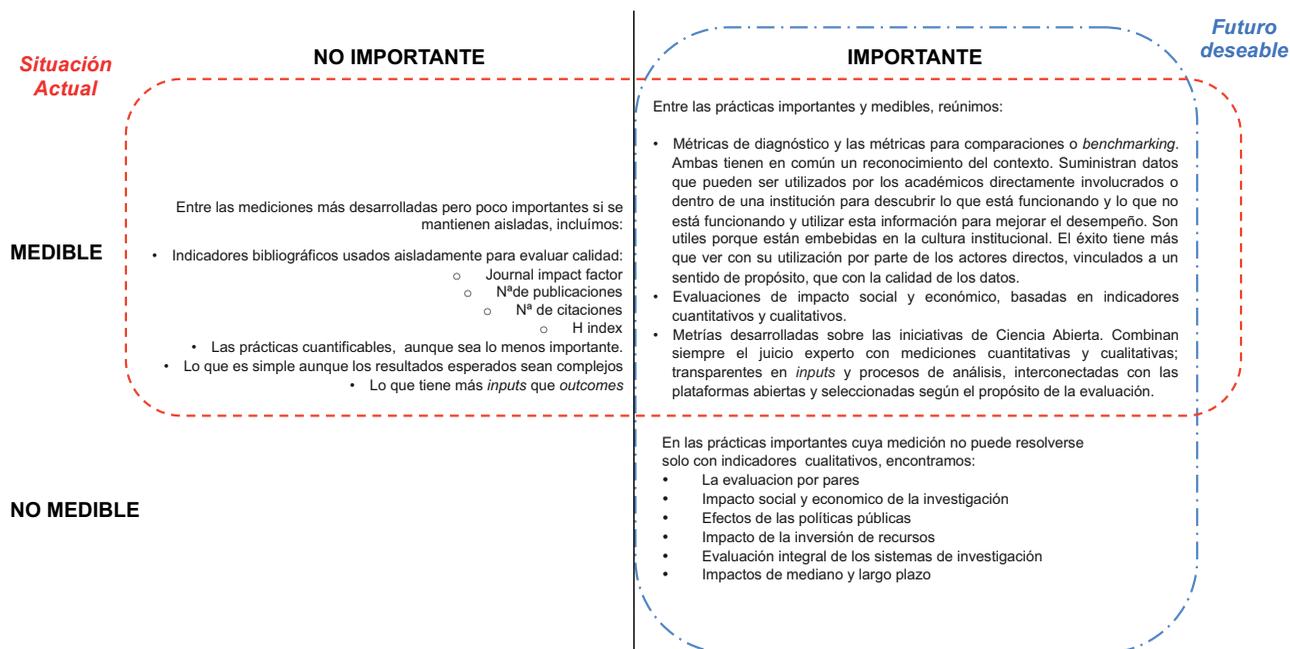
12. RQ+ Research Quality Plus. A Holistic Approach to Evaluating Research Zenda Ofir, Thomas Schwandt, Colleen Duggan, Robert McLean. IDRC, 2016. Disponible en: <https://www.idrc.ca/sites/default/files/sp/Documents%20EN/Research-Quality-Plus-A-Holistic-Approach-to-Evaluating-Research.pdf>

La evaluación del impacto económico y social de la investigación fue el tema central de la reunión anual del *Global research Council* (Sao Paulo, mayo 2019). La elección de la temática refleja el aumento de las expectativas de las organizaciones de financiación para realizar investigaciones cuyo impacto social y económico sea evidente y demostrable. Los participantes discutieron en esa oportunidad con más detalle las implicaciones de confiar en el impacto social y económico como criterio de financiación. La declaración final enumera 18 principios clave para abordar las expectativas de impacto y evaluar el impacto de la investigación desde la perspectiva de las organizaciones financiadoras.¹³

Según este enfoque, las métricas de desempeño no deben ser abandonadas, aun cuando algunas tienen consecuencias negativas por su uso actual. Saber cuándo usarlas es una cuestión de compensaciones y equilibrios. Y esto también es una cuestión de juicio. “Medir no reemplaza al juicio, medir requiere juicio, juicio sobre si medir, qué medir, cómo evaluar el significado de lo que se ha medido, si recompensar o penalizar en función de los resultados y a quién se le debe de dar acceso a las mediciones”.¹⁴

Jerry Müller ofrece una lista de chequeo para decidir la utilidad y pertinencia de la medición:

Hacia un portafolio equilibrado de evaluación



46

Nota: elaboración propia con base en “Diez razones por las cuales le conviene a un directivo desarrollar el enfoque Álamo en su equipo” de Michel Fiol (documento sin publicar)

Con base a estos elementos, postulamos una evolución hacia un portafolio de evaluación equilibrada. La matriz propuesta no debe ser leída como una taxonomía ni un diagrama estático: ilustra el movimiento que permite pasar del estado actual de la evaluación de la investigación científica a una posible dirección futura de integración que asegure el equilibrio, teniendo en cuenta, en cada contexto, lo importante, sea medible con indicadores cualitativos u observable para evaluaciones cualitativas y mixtas de mediano y largo plazo.

- Qué tipo de información se requiere medir. Cuanto más inanimado es el objeto a medir, más es posible que la medición cuantitativa resulte eficiente. Cuando los objetos a medir están influenciados por los procesos mismos de medición, la medición se vuelve menos confiable.
- ¿Cuán útil es la información disponible? Aunque haya muchos datos disponibles, es muy probable que la facilidad de medir sea inversamente proporcional a la importancia de lo que se mide.

13. ver 2019 GRC Statement of Principles Addressing Expectations of Societal and Economic Impact en https://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_2019_Statement_of_Principles_Expectations_of_Societal_and_Economic_Impact.pdf

14. The tyranny of metrics / Jerry Z. Müller. Princeton : Princeton University Press, 2018. p. 176-177. Disponible en: <https://press.princeton.edu/titles/11218.html>.

- ¿Cuál es el costo de no basarse en una medida estandarizada? ¿Hay otras fuentes de información sobre el desempeño, basadas en el juicio y la experiencia?
- ¿Con qué propósito hemos de medir? Hay una distinción clave entre la medición interna del desempeño por los actores directos versus datos utilizados por partes externas para recompensa y castigo.¹⁵

3. ¿POR QUÉ ES TAN DIFÍCIL CAMBIAR LAS PRÁCTICAS INSTITUCIONALES?

“Las métricas parecían, al principio, bastante inofensivas, pero, como los cucos en un nido, se han convertido en monstruos que amenazan a la ciencia misma. Ya han producido una ‘sociedad de auditoría’ [...] ¿Quién tiene la culpa y qué hacer? Los villanos principales son la moda, el culto al management y la política de nuestro tiempo, todos los cuales favorecen la evaluación numérica del ‘desempeño’ y recompensan el cumplimiento.

En los últimos años, dentro y fuera de los gobiernos, las personas han perdido de vista los propósitos principales de las instituciones. [...] Pero los científicos de todos los rangos, senior y junior, también tienen la culpa, ya que hemos permitido dócilmente que esto suceda. ¿Pero podemos empezar a contraatacar? Necesitamos concienciar sobre los problemas y hacer cambios a nivel local”

Peter Lawrence¹⁶

Las universidades están atrapadas en un círculo vicioso: la falta de confianza lleva a una apoteosis de métricas y la confianza en las métricas contribuye a promover su manipulación y a una confianza decreciente en los juicios expertos y en los indicadores mismos por su uso inapropiado. El resultado es una sobrerregulación y una red de reglas cada vez más densas, incluyendo la proliferación de procedimientos y formatos dentro de las organizaciones. A menudo las métricas proveen los instrumentos para tornar la red aún más densa: en un intento por detener las fallas de las métricas, debidas a usos engañosos, las organizaciones instituyen una cascada de reglas que afecta el funcionamiento de la Institución y disminuye su eficiencia.

Es lo que Jerry Müller llama la fijación métrica, cuyos componentes clave son:

- La creencia de que es posible y deseable reemplazar el juicio, adquirido con la experiencia y el talento personal, con indicadores numéricos de desempeño comparativo basados en datos estandarizados
- la creencia de que hacer públicas dichas métricas (transparencia) asegura que las instituciones van a cumplir realmente con sus propósitos (rendición de cuentas);

15. Ibid, p. 110. Para observar un ejemplo puede observarse U-Multirank (<https://www.umultirank.org/>) utiliza métricas de diagnóstico y para benchmarking, mientras que todos los demás rankings utilizan métricas estandarizadas.

16. The mismeasurement of Science / Peter Lawrence. Disponible en: [https://www.cell.com/current-biology/references/S0960-9822\(07\)01516-3](https://www.cell.com/current-biology/references/S0960-9822(07)01516-3). (Este artículo es un buen recuento de los efectos perversos de las métricas estandarizadas).

- la creencia de que la mejor manera para motivar a las personas en las organizaciones es vinculando las recompensas y penalidades a la medición del desempeño, recompensas que puede ser monetarias (complementos salariales) o de reputación (*rankings*).

Los indicadores inapropiados crean incentivos perversos. Los dos principales ejemplos son el factor de impacto de publicaciones y las citaciones de trabajos académicos. En el caso del Journal Impact Factor JIF, es sabido que no fue creado originalmente para evaluar investigadores individuales sino como referente para guiar las decisiones de compra de las bibliotecas. Su aplicación como medida de la calidad de las publicaciones y su vinculación con incentivos económicos a los investigadores, ha obligado a los investigadores a una cultura del desempeño dependiente de los JIFs o del índice-h¹⁷ y la consecuente perversión de los propósitos de la comunicación científica.

Las métricas se consideran a nivel de revista y son adscritas, por extensión, a los artículos, que son heterogéneos en sí mismos y la medición gestionada, ahora directamente por los oligopolios de las bases de datos, se presta a manipulaciones, por ejemplo a través de “club de citaciones” y el “rebanado” de publicaciones para incrementar el número de citaciones. Para la Asociación de Universidades Europeas – EUA- el predominio del factor de impacto de la revista conduce a dos problemas principales.¹⁸

- La calidad de un artículo producido por investigadores no se evalúa directamente, sino a través de un sustituto, es decir, la reputación de la revista en la que se publica, debería evaluarse en función del mérito de la investigación en sí misma.
- Esta situación refuerza la posición dominante de los editores académicos comerciales y aumenta de manera desproporcionada su poder para dar forma a la manera en que se financia y dirige la investigación.

Una de las razones que explica el uso intensivo de la métrica en contextos competitivos es que constituyen un instrumento sencillo, útil y directo para impulsar el

17. El índice H fue creado por el físico argentino-americano Jorge Hirsch en 2005. Es un indicador del impacto de las publicaciones de un investigador. Tiene en cuenta el número de publicaciones de un investigador y el número de sus citas. El índice h de un autor es igual al número más alto h de sus publicaciones que recibió al menos h citas cada una. Ver J. E. Hirsch An index to quantify an individual's scientific research output, PNAS November 15, 2005 en <https://www.pnas.org/content/102/46/16569?cited-by=yes>. Recuerda Hirsch 4 años más tarde, “Un criterio utilizado a menudo para evaluar el logro de la investigación fue contar artículos publicados en revistas de alto factor de impacto; Quería proporcionar un criterio alternativo. [...] Tuve problemas para que los artículos fueran aceptados en revistas con los factores de impacto más altos debido a la naturaleza controvertida de mi investigación. Afortunadamente, hubo revistas con factores de impacto más bajos que aceptaron mis documentos. Sin embargo, fueron bien citados, lo que significa que otros investigadores los encontraron útiles” ver Vicky Hampton Jorge Hirsch: the man behind the metric <https://www.researchtrends.com/issue14-december-2009/people-focus/>

18. Roadmap on Research Assessment in the Transition to Open Science. European University Association. Bruselas, 2018. Disponible en: <https://www.eua.eu/component/attachments/attachments.html?id=348>.

desempeño financiero de las organizaciones como parte de la competitividad de las instituciones.¹⁹ Su uso generalizado tiene poco más de 20 años, y ha sido promovido por el desarrollo de las plataformas tecnológicas y su capacidad de analizar las transacciones de las bases de datos bibliográficas; por el desarrollo de los sistemas de gestión y evaluación con datos propios de los sistemas de investigación, la demanda de auditoría y evaluación del gasto público en la educación superior y la investigación; y por la necesidad de homogenizar las mediciones para la necesaria comparación e integración regional e internacional de la evaluación.

En los países más desarrollados, el conjunto de herramientas utilizadas para identificar, analizar y evaluar la información sobre la investigación es creciente; hay un interés significativo de las plataformas en línea que pueden recurrir a sus algoritmos para dar acceso y visibilidad a la investigación, incluyendo Twitter, Facebook y sitios específicos como Research Gate,²⁰ tal y como proponen las métricas alternativas. En consecuencia aumenta la tendencia a usar esas mediciones, no porque sean realmente pertinentes o valiosas para el mejoramiento, sino porque están disponibles y son fáciles de gestionar.

La Comisión Europea incluyó el tema de métricas alternativas como uno de los tópicos principales de su Plataforma de Ciencia Abierta, y creó en 2016 el Grupo Experto de Indicadores cuyo primer informe *Next-generation metrics: Responsible metrics and evaluation for open science*,²¹ publicado en 2017, presenta un estado del arte en métricas, describe sus ventajas, cualidades y debilidades a superar; propone acciones para desarrollar sistemas que permitan una evaluación cuantitativa y cualitativa de la investigación y sus resultados. Por su parte, el Grupo de Trabajo en Recompensas de la Comisión Europea ha propuesto una

Matriz de Evaluación de la Carrera Profesional²² como posible movimiento práctico hacia un enfoque más integral para evaluar a los investigadores a través de la lente de la Ciencia Abierta.

En 2016, la Asociación de Universidades Europeas, EUA publicó la citada hoja de ruta de la Evaluación y la Liga de Universidades de Investigación Europeas, LERU – la Liga de universidades de investigación europeas- incluyó en su hoja de ruta de “Ciencia Abierta y su rol en las universidades”²³ un apartado dedicado a la evaluación.

En un nivel más amplio, la emergencia del big data combinada con la capacidad de análisis y la influencia de las métricas disponibles, aunque poco pertinentes, va en franco aumento. En una suerte de distopía similar a la de *Minority Report*,²⁴ dos investigadores del MIT, *Sloan School of Management*, inspirados por el bestseller *Moneyball* de 2003 (en el que Michael Lewis describe cómo las fortunas de un equipo de béisbol se transformaron mediante el uso riguroso de datos predictivos y modelos para identificar e invertir en talento subvaluado), sostienen que las universidades están maduras para su propio *Moneyball* para decisiones de contratación de los académicos. “El uso de la analítica para mejorar las decisiones de contratación –dicen los autores- ha transformado a las industrias del béisbol a la banca de inversión. Entonces, ¿por qué las decisiones de tenencia de los profesores todavía se toman a la antigua?”²⁵ Según esta perspectiva, al analizar un conjunto de métricas para publicaciones, citas y coautoría en una etapa temprana de una carrera del investigador, e incluyendo estas en las decisiones de contratación y promoción, el equipo del MIT sugiere que es posible predecir el rendimiento futuro con mayor precisión y confiabilidad que sólo a través de juicios subjetivos.

Como observa el Prof. James Wilsdon, coordinador del estudio “The Metric Tide”:²⁶ “Dado el papel que las citas, los índices H, los factores de impacto de la revista, los ingresos de subvenciones y las métricas convencionales ya juegan en la gestión de la investigación y la toma de decisiones (tanto explícita e implícitamente), algunos podrían aceptar el análisis predictivo como el siguiente paso lógico. Ya hay proveedores privados, como Academic Analytics, que ofrecen una versión de estos servicios a las universidades.

19. En Inglaterra y en los países nórdicos, el “Publicar o perecer” está directamente vinculado a la productividad financiera, como ilustra el caso dramático de Stefan Grimm, un biólogo del Imperial College que se suicidó en 2014 luego de que el jefe de su departamento le informó que, no obstante haber sometido el número más alto de solicitudes de financiamiento de la facultad de medicina, y no obstante haber obtenido 135.000 £ de financiamiento, la necesidad de obtener subsidios adicionales podrían dar lugar a una consideración seria sobre si su desempeño estaba al nivel esperado de un profesor del Imperial College. En 20 años, Grimm había publicado cincuenta artículos, dos libros y hecho cinco solicitudes de patentes. Su suicidio dió lugar a un cambio en los reglamentos de la institución respecto al desempeño académico. Fuente: https://en.wikipedia.org/wiki/Stefan_Grimm

20. Cuando se creó en 2008, la red social académica ResearchGate, promoción o su puntuación RG como “una nueva manera de medir la reputación científica” y de esta manera “Ayudarte a medir y mejorar tu posición dentro de la comunidad científica”. Peter Kraker, Katy Jordan y Elisabeth Lex decidieron. considerar en detalle la métrica opaca. Mediante la ingeniería inversa del puntaje, encontraron que se le daba un peso significativo a los “puntos de impacto”, una métrica similar al factor de impacto de las revistas, ampliamente desacreditado. El uso extensivo del RG Score en los correos electrónicos de marketing sugiere pretende ser una herramienta de marketing que lleve más tráfico al sitio. Fuente: <http://eprints.lse.ac.uk/70955/1/blogs.lse.ac.uk-The%20ResearchGate%20Score%20a%20good%20example%20of%20a%20bad%20metric.pdf>

21. Next-generation metrics: Responsible metrics and evaluation for open science. European Commission Expert Group on Altmetrics. 2017. Disponible en: <https://ec.europa.eu/research/openscience/pdf/report.pdf>

22. Evaluation of Research Careers fully acknowledging Open Science practices: Rewards, incentives and/or recognition for researchers practicing Open Science. Working Group on Rewards under Open Science. European Commission. July 2017. p 15-17. Disponible en: https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf

23. LERU Open Science and its role in universities: A roadmap for cultural change. League of European Research Universities. Disponible en: <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>

24. Historia corta de ciencia ficción de Philip K Dick publicado en 1956, con versión cinematográfica de Steven Spielberg en 2002.

25. ‘Moneyball’ for Professors? Frontiers blog. E. Brynjolfsson, J. Silberholz. MIT Sloan Management Review, 14 December 2016. Disponible en: <https://sloanreview.mit.edu/article/moneyball-for-professors/>

26. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management / James Wilsdon ed. Higher Education Funding Council for England HEFCE, 2015. Disponible en: https://responsiblemetrics.org/wp-content/uploads/2019/02/2015_metrictide.pdf

Otros se horrorizarían, y verían la analítica predictiva como el punto en el que la marea de la métrica ascendente rompió todas las defensas que existen alrededor de nuestra noción tradicional de carrera académica”.²⁷

En este mismo sentido se desarrollan los servicios comerciales de apoyo a los rankings universitarios, deslumbrantes por su capacidad de acumulación y análisis de datos, pero opuestos a la indispensable transparencia en la evaluación de la investigación científica porque mantienen el proceso en una «caja negra» y no permiten acceso para la verificación de sus datos y mediciones; obligando a las instituciones a participar e invertir valiosos recursos en una competencia cuyos objetivos no necesariamente están alineados con las prioridades y metas de cada comunidad académica, ni mucho menos con las diversas condiciones de la investigación científica. Sin embargo, “cuando una métrica se vuelve la moneda del reino, rehusarse a utilizarla es arriesgarse a la bancarrota”,²⁸ y ninguna institución de investigación, ni investigador puede ni quiere excluirse de los sistemas de financiación y reconocimiento.

En América Latina, los países líderes con iniciativas en Ciencia Abierta: Argentina, México, Perú, Brasil y Chile, reconocen la evaluación como un componente integral, pero aún no cuentan con políticas específicas. Aunque no se registran documentos prescriptivos de universidades o asociaciones dedicados a este aspecto, la iniciativa Conocimiento Abierto para América Latina y el Sur Global, AmeliCA,²⁹ que reúne investigadores y expertos para estudiar y desarrollar el acceso abierto en la región, cuenta con una comisión de Métricas Responsables³⁰ que ha iniciado el diseño un instrumento de medición con sus indicadores para ciencias sociales y humanidades.³¹ Paralelamente, decisiones de política laboral, como el reconocimiento monetario a los académicos con base en sus publicaciones afectan las prácticas investigativas han desdibujado las motivaciones de los investigadores para publicar.

27. UK Progress towards the use of metrics responsibly Three years on from The Metric Tide report. James Wilsdon 10 July 2018. Disponible en <https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/The%20Forum%20for%20Responsible%20Research%20Metrics/UK%20progress%20towards%20the%20use%20of%20metrics%20responsibly%2010072018.pdf>

28. The tyranny of metrics / Jerry Z. Müller. Princeton : Princeton University Press, 2018. p. 29 Disponible en: <https://press.princeton.edu/titles/11218.html>

29. Creada en 2016 con el respaldo de Unesco, el Consejo Latinoamericano de Ciencias Sociales (CLACSO), la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc), la Universidad Autónoma del Estado de México (UAEM), la Universidad de Antioquia (UdeA), la Universidad Nacional de La Plata (UNLP) y la Universidad de Panamá (UP).

30. El rol de las revistas en el apoyo a la construcción de comunidad. Métricas Responsables AmeliCA, Disponible en: <http://www.amelica.org/proyectos/metricas.html>

31. Ciencia Abierta, reporte para tomadores de decisiones. 2. ed. Paola Andrea Ramírez, Daniel Samoilovich. Paris, Montevideo : Unesco, 2019. Disponible en: <http://forocilac.org/wp-content/uploads/2019/05/2019-PolicyPapersCILAC-CienciaAbierta-29-04-2019-Final.pdf>

3.1. La avanzada en contra del *status quo*

En este contexto, es muy instructivo comprender cómo algunas instituciones han logrado liberarse del “abrazo del oso” de las métricas, y transitar la evolución propuesta en la matriz, pasando del estado actual de la evaluación de la investigación científica a una posible dirección futura, dando prioridad a lo importante —sea medible u observable— con relación a lo medible pero no importante. Algunos de estos ejemplos tienen que ver con la medicina y carreras afines. A nuestro juicio, esto no es casual: el contacto directo con la salud y la enfermedad permite dar una dimensión adecuada a los indicadores cuantitativos. Los sistemas de investigación y gestión de la salud comparten métodos validados de análisis que les permiten superar las barreras iniciales del consenso y la confianza en los datos, en los cuáles se basa la evaluación y medición de impacto.

Tratándose de políticas internas de gestión, su análisis y el de sus efectos no son fácilmente accesibles o transferibles fuera de las universidades mismas. De todos modos, los casos del INSERM, y de las universidades de Gantes, Utrecht y Manchester nos dan suficientes indicios sobre “el reclamo, el objetivo y la tarea”. Todas las instituciones aquí incluidas decidieron en algún momento cortar con las prácticas tradicionales, aunque, como veremos, los procesos pueden ser reversibles si la transformación no está suficientemente consolidada con políticas y procesos acordes.

En el INSERM de Francia, la voluntad de ir más allá de la visibilidad de un investigador, exacerbada por los sistemas de acreditación y los rankings, justificó la decisión. El INSERM es una institución pública de naturaleza científica y tecnológica, dedicada a la investigación biomédica y la salud pública. Cuenta con más de 5000 científicos, ingenieros y técnicos y es el primer depositante de patentes europeo en el sector farmacéutico y biomédico. Con lo cual tiene la masa crítica y el reconocimiento necesario para marcar una tendencia de cambio.

La Universidad de Gantes es un ejemplo de liderazgo distribuido en todas las etapas y niveles de la evaluación. El punto de partida fue una evaluación del modelo de carrera existente entre 2012 y 2017, que consolidó las críticas de la comunidad académica contra la utilización de las métricas. Las principales objeciones fueron que el proceso:

- evaluaba la producción con indicadores cuantitativos y se basaba en metas fijadas *a priori* de manera individual;
- implicaba un peso administrativo importante (descripciones de puestos anuales, informes de actividad, archivos de desempeño); y
- requería una alta frecuencia de realización de la evaluación.

El modelo llevaba a una competencia entre académicos, a una presión por el exceso de trabajo, a una insatisfacción de los empleados y hasta a un *burn out*. Era necesario dar lugar a una nueva perspectiva sobre el desempeño de los investigadores.

El cambio radical se concretó en la visión y el programa de candidatura del nuevo rector y vicerrector. En palabras del rector: “Un proceso de evaluación académica predominantemente cuantitativo e impulsado por resultados abre el camino para el desarrollo y crecimiento del talento, priorizando el desarrollo de la visión y la estrategia, tanto a nivel personal como a nivel grupal. La calidad prevalece sobre la cantidad”.³²

El nuevo equipo rectoral de la Universidad de Gantes tenía una visión y una idea inicial clara de por qué se estaba haciendo la reforma. Pero, el caso de Gantes ilustra también que es necesario ir más allá de la visión y liderazgo concentrados en el máximo nivel y desarrollar un liderazgo distribuido a todos los niveles. Se llevó a cabo una amplia consulta con los profesores a todos los niveles y con expertos en dirección de recursos humanos. El liderazgo académico a nivel de la universidad en las etapas de preparación, se replicó a nivel de las facultades y departamentos durante la preparación e implementación de la reforma. Así, la junta de gobierno aprobó los principios generales y más tarde los reglamentos para los académicos.

De manera similar, en el University Medical Center Utrecht, de Holanda, “las publicaciones que influían directamente en la atención al paciente no se ponderaban más en las evaluaciones que cualquier otro artículo, y menos si ese trabajo aparecía en la literatura gris, es decir, en informes oficiales y no en revistas científicas. Algunos investigadores desistieron de buscar publicaciones que pudieran mejorar la medicina, por los que obtendrían pocas citas. Todo esto llevó a muchos miembros de la facultad, especialmente a los más jóvenes, a quejarse de que la presión de la publicación les impedía hacer lo que realmente importaba, como fortalecer los contactos con organizaciones de pacientes o intentar que los tratamientos prometedores funcionaran en el mundo real. La institución decidió liberarse de esta mentalidad”.³³

En Utrecht, dos aspectos de su enfoque fueron cruciales: “Primero, no nos dejamos paralizar por la creencia de que sólo la acción conjunta junto con los financiadores y las revistas producirían un cambio real. Estábamos dispuestos a avanzar por nuestra cuenta como institución. En segundo lugar, nos aseguramos de que, aunque el cambio se estimuló desde arriba, los criterios fueron establecidos por los miembros de la facultad que esperan ser juzgados por esos estándares [...] Durante los últimos años de conferencias y talleres, al principio nos sorprendió lo poco que sabían los investigadores de carrera temprana y media sobre el “modelo de negocio” de la ciencia moderna

32. Para el caso de la Universidad de Gantes se tuvo en cuenta la presentación del rector Rik van de Walle en el EUA workshop - May 14, 2019 Research Assessment for Researchers' Recruitment and Career Progression, disponible en: <https://eua.eu/component/attachments/attachments.html?task=attachment&id=2162>, y en una conversación posterior con el Prof. Van de Walle.

33. Redefine Excellence: Fix incentives to fix science / Rinze Benedictus and Frank Miedema. En: Fewer numbers, better science. Oct, 2016. Disponible en <https://www.nature.com/news/fewer-numbers-better-science-1.20858>.

y sobre cómo funciona realmente la ciencia. Pero se comprometieron, aprendieron rápidamente e identificaron ideas de futuro para mejorar la ciencia”.³⁴

Ambos casos ilustran (el de Gantes en forma más explícita) una de las características con que Burton Clark define a las universidades emprendedoras: la existencia de un núcleo de dirección reforzado (*strengthened steering core*)³⁵ y también ponen el foco sobre el rol de los investigadores jóvenes, más desfavorecidos con el sistema dominante.

3.2. Del control a la confianza

El nuevo modelo desarrollado en 2018-2019 en la Universidad de Gantes supuso una nueva perspectiva sobre el desempeño de los investigadores y la carrera académica; una perspectiva integral que ya no se basa solamente en la evaluación de la investigación, ya que la educación, la divulgación y el impacto son igualmente valoradas; la diferenciación y complementariedad constituyen los fundamentos de un modelo de carrera que busca el desarrollo del talento, y que reconoce la evaluación como un proceso de aprendizaje organizacional y mejoramiento continuo.

Esta visión inspiró una reforma de la progresión en la carrera profesional muy importante, basada en la confianza en lugar del control.³⁶ De acuerdo con esta perspectiva, la confianza es entendida como un resultado de la libertad académica aunada a la responsabilidad. El talento y las ambiciones de los académicos son considerados como un impulsor natural para la progresión en la carrera, pero siempre como parte de un grupo. La motivación intrínseca de cada miembro del plantel académico garantiza que nadie necesite objetivos a priori para desempeñarse bien en las tareas centrales de la universidad: educación, investigación y compromiso institucional o social.³⁷

34. Redefine Excellence: Fix incentives to fix science / Rinze Benedictus and Frank Miedema. En: Fewer numbers, better science. Oct, 2016. Disponible en <https://www.nature.com/news/fewer-numbers-better-science-1.20858>.

35. Creating Entrepreneurial Universities: Organizational Pathways of Transformation / Burton R. Clark. Oxford: Published for the IAU Press by Pergamon Press, 1998.

36. La cuestión de la confianza es de una amplitud que sobrepasa a la gestión de talentos en un marco institucional. Frente a la complejidad de las situaciones que los ciudadanos viven cotidianamente –cambio climático, migraciones, corrupción-, el juicio de los expertos está en entredicho. La reconstrucción de la confianza en ellos es un problema central de las sociedades democráticas. “El requerimiento de una rendición de cuentas medible y transparente crece a medida que la confianza disminuye. Hay una afinidad electiva entre una sociedad democrática, con una movilidad social sustancial y una mayor heterogeneidad étnica, y la cultura de la rendición de cuentas mensurable. En sociedades con una clase alta establecida y transgeneracional, los miembros de esta clase se sienten probablemente más seguros de sus posiciones y confían unos en otros y comparten un cierto grado de conocimientos tácitos sobre cómo gobernar” (The tyranny of metrics / Jerry Z. Müller. Princeton: Princeton University Press, 2018, p. 39). Indicadores y rankings son sólo sucedáneos para reestablecer la confianza. En el momento actual, la complejidad de los problemas requiere restaurar la confianza en la ciencia a través de la reconstrucción de la confiabilidad: cuándo y bajo qué condiciones podemos confiar en alguien. No bastan actitudes genéricas, hay que establecer salvaguardas sobre la actuación de las personas y las instituciones sin que ellas distorsionen su desempeño. Ver también: *Linking trust to trustworthiness* / Onora O’Neill. Conferencia en el University College Dublin. Disponible en: <https://www.youtube.com/watch?v=A0u76tA1OyA>.

37. Ghent University is changing course with a new career model for professorial staff. Diciembre de 2018. Disponible en <https://www.ugent.be/en/news-events/new-career-model-professorial-staff.htm>

De esta manera se abordan tres cuestiones centrales de la gestión y gobernanza de las organizaciones, particularmente vitales en las instituciones académicas, que dependen principalmente de la calidad y gestión de sus recursos humanos:

- Sustituye un control basado en la dirección por objetivos, en crisis incluso en organizaciones con fines de lucro,³⁸ por una confianza asentada en una motivación intrínseca, no en objetivos fijados a priori controlados a intervalos regulares.
- Permite una simultánea integración de los académicos entre sí, y una diferenciación de sus roles. El grupo se convierte en el ámbito natural de referencia de los académicos. Todas las contribuciones a las misiones sustantivas son igualmente valoradas.
- Recompensa a individuos por el desempeño en su grupo fortaleciendo el sentido de propósito común, así como las relaciones sociales que alimentan la motivación por la cooperación y la efectividad institucional.

3.3. Indicadores en contexto y procesos manejables

Veamos cómo se concretan estos principios rectores en los casos considerados. En la Universidad de Gantes, el nuevo proceso se aplica a todos los niveles académicos, desde el *tenure track* hasta el profesor a tiempo completo, independientemente de la duración del contrato y de su dedicación. El ritmo de evaluación pasó de dos/cuatro años a cinco años (el mínimo previsto por la ley). La idea principal es que “todas las personas que tienen un buen desempeño serán promovidas con un mínimo de rendición de cuentas y esfuerzo administrativo y con un máximo de libertad y responsabilidad”.³⁹

Se aplica un enfoque basado en el mérito y la evidencia, o sea una reflexión global a posteriori al final del ciclo (en lugar de una identificación de objetivos cuantitativos *a priori*); que tiene cuenta los resultados más significativos (¿De qué está usted orgulloso?); en lugar de basarse en una lista exhaustiva de productos (*outputs*) toma en cuenta logros o realizaciones (*achievements*). El foco está puesto en el desarrollo del talento, el crecimiento personal y la orientación en la carrera, mediante un continuo diálogo (en lugar del rellenado de formularios), incluyendo la evaluación de las capacidades de liderazgo y de gestión de personas.

38. University Management: A Necessary Debate, en Leadership and Governance in Higher Education, a Handbook for decision Makers and Administrators / Michel Fiol y Daniel Samoilovich. RAABE Verlag: Berlin, 2014. Disponible en: https://www.academia.edu/34264671/University_Management_A_Necessary_Debate, version previa en español publicada en el anuario 2011 de la Fundación Conocimiento y desarrollo, disponible en https://www.columbus-web.org/images/pdf/FIOL_SAMOILOVICH%281%29.pdf

39. Para el caso de la Universidad de Gantes se tuvo en cuenta la presentación del rector Rik van de Walle en el EUA workshop - May 14, 2019 Research Assessment for Researchers' Recruitment and Career Progression, disponible en: <https://eua.eu/component/attachments/attachments.html?task=attachment&id=2162>, y en una conversación posterior con el Prof. Van de Walle.

Un aspecto central de la reforma es la inclusión de los objetivos individuales en la entidad más amplia, sea esta el grupo investigación, el departamento, el programa de estudio, la facultad o la universidad. Esta integración se da a través de un texto en el que cada académico, al inicio del ciclo de evaluación, describe sus ambiciones en relación a la entidad más amplia, cómo contribuye a sus objetivos estratégicos. La incorporación en la entidad más grande va de la mano con la diferenciación de acuerdo con los talentos de todos: no es un modelo competitivo, sino que se basa en el uso complementario de las fortalezas y cualidades de los miembros del plantel académico.

Las herramientas de apoyo son muy ligeras, pero con un propósito claro: guiar a las personas. Se utilizan tres modelos o guías: la mencionada guía de integración, el informe de reflexión (mis principales contribuciones y realizaciones); el informe de evaluación: contribuciones y realizaciones y una evaluación global (a cargo del comité de recursos humanos). Estos insumos son una piedra angular del nuevo sistema. A cada académico se le asigna un comité de recursos humanos, que tiene un doble rol: apoyar y orientar en las distintas etapas de la carrera y asesorar a la Junta de la facultad sobre la evaluación y promoción. El comité está compuesto por cinco miembros, cuya conformación lo hace suficientemente cercano y distante del individuo para permitirle una visión de lo general y lo específico, y aportar una perspectiva desde el punto de vista de la gestión del talento humano.

Dos factores de éxito sobresalen en este caso: alta pericia en los sistemas de gestión del desempeño del talento humano; y procesos de administración livianos y flexibles (el foco en las personas y no en los formularios), para dedicar más tiempo a la investigación, a la enseñanza y a otras actividades académicas fundamentales. La evaluación global se basa en cuatro criterios: muy bueno, bueno, pobre e insatisfactorio. Con una evaluación buena o muy buena se recomienda la promoción. En el caso de una evaluación pobre o insatisfactoria se propone una trayectoria remedial y se realiza una nueva evaluación luego de dos años.

Algunas de estas características se repiten en los procesos practicados en el University Medical Center Utrecht. El centro se inspiró en el enfoque utilizado en el Instituto Karolinska de Estocolmo, que solicita a los candidatos un paquete de logros científicos, docentes y de otro tipo. Junto con otros elementos, los candidatos de Utrecht ahora brindan un breve ensayo sobre quiénes son y cuáles son sus planes como miembros de la facultad. Deben discutir los logros en términos de cinco dominios, solo uno de los cuales es el de las publicaciones científicas y subvenciones. Primero, los candidatos describen sus responsabilidades gerenciales y deberes académicos (como revisar revistas y contribuir a comités internos y externos); en segundo lugar, explican cuánto tiempo dedican a los estudiantes, qué cursos han desarrollado y qué otras responsabilidades han asumido. Luego, si corresponde, describen su trabajo clínico, así como su participación en la organización de ensayos clínicos y la investigación de nuevos tratamientos

y diagnósticos. Finalmente, el portafolio cubre el emprendimiento y el acercamiento a la comunidad.⁴⁰

También renovaron el procedimiento de evaluación del solicitante. El presidente del Comité tiene la tarea formal de garantizar que se discutan todos los dominios para cada candidato. El consejo de doctorado elige a un supervisor del año, sobre la base de la calidad de la supervisión, y no solo por el número más alto de estudiantes de doctorado supervisados, como era costumbre antes.

Encontramos similar modificación del proceso de evaluación con expedientes manejables, en la experiencia del Prof. Mark Ferguson: “Hace unos 20 años, cuando era decano de ciencias biológicas en la Universidad de Manchester, Reino Unido, hice un experimento. En ese momento, evaluábamos a los candidatos que solicitaban promociones utilizando medidas convencionales: número de publicaciones, calidad de la revista, índice h, etc. En su lugar, decidimos pedir a los solicitantes que nos dijeran cuáles consideraban sus tres publicaciones más importantes y por qué, y que enviaran una copia de cada una. Hicimos preguntas simples y directas: ¿qué has descubierto? ¿Por qué es importante? ¿Qué has hecho sobre tu descubrimiento? Para hacer que los solicitantes se sintieran más cómodos con esta peculiar evaluación, también indicamos que podrían enviar, si lo deseaban, una lista de todas sus otras publicaciones científicas, todos lo hicieron. Las tres publicaciones elegidas por el solicitante me dijeron mucho acerca de sus logros y juicio. A menudo, destacaron los impactos no convencionales de su trabajo”.⁴¹

Los tres ejemplos mencionados indican que desarrollar procesos basados en expedientes manejables, concentrados en los propósitos de la carrera académica, con formatos y procedimientos flexibles, permite concentrar la atención sobre los aspectos más cualitativos del desempeño académico, y considerar los indicadores en su contexto. Tal y como resume Müller: “En un departamento académico la evaluación de la productividad de los académicos puede hacerse por parte del director del departamento o de un comité pequeño, en consulta con otros miembros académicos cuando sea necesario, basándose en su conocimiento y experiencia acumulada sobre lo que constituye significativo en un libro o un artículo. En caso de decisiones importantes, como por ejemplo una promoción, pueden ser llamados académicos en el área del candidato para dar una evaluación confidencial, una forma elaborada de la revisión por pares. Los números extraídos de las bases de datos pueden ser de algún uso en este proceso, pero los números requieren un juicio basado en la experiencia para poder evaluar su utilidad. Este juicio basado en la experiencia profesional es precisamente

lo que es eliminado por una excesiva confianza en los indicadores de desempeño estandarizados [...] El desafío es abandonar indicadores universales y descubrir qué es lo que vale la pena contar y qué es lo que realmente los números quieren decir en su contexto local”.⁴²

Mark Ferguson, Director General de la Science Foundation Ireland, comenta al respecto de los cambios adelantados: “Cuando me moví de mi posición como decano, el sistema volvió a su forma convencional. Los cambios que resultan en diferencias con respecto a una norma cultural son difíciles de sostener, particularmente cuando se basan en la pasión de un pequeño número de personas”.⁴³ Cualquier estrategia, para tener un impacto en la cultura, debe integrarse en las prácticas de la organización, sus estructuras, y procesos. “Las discusiones extendidas de la comunidad dan sus frutos. Creemos que los comités de selección y evaluación son conscientes de que la bibliometría puede ser una fuerza reductora, pero que los evaluadores pueden carecer del vocabulario para discutir dimensiones menos cuantificables. Al exigir formalmente indicadores cualitativos y una cartera descriptiva, ampliamos de lo que se puede hablar. Damos forma a las estructuras que dan forma a la ciencia; podemos asegurarnos de que no se deformen”.⁴⁴

Naturalmente, las políticas e instrumentos a nivel de sistema favorecen el tránsito hacia nuevas formas más equilibradas de evaluar la ciencia. La experiencia de Ferguson en la Universidad de Manchester, según su testimonio, influyó en su tarea como director general de la principal agencia de financiación de la ciencia en Irlanda: “Para llevar la empresa científica a mejores medidas de calidad, tal vez necesitemos un esfuerzo colectivo de un grupo de universidades internacionales líderes y financiadores de investigación. Lo que se mide es lo que se obtiene: por lo tanto, si los financiadores se enfocan en evaluar avances de investigación sólidos (con un potencial impacto económico y social), esto puede alentar un trabajo confiable e importante y desalentar los juegos bibliométricos [...] ¿Qué pueden hacer los financiadores? Al ajustar las recompensas, estos cuerpos pueden moldear las elecciones de los investigadores profundamente [...] Deben alentar explícitamente la investigación importante, incluso a expensas de la tasa de publicación. Con este fin, en Science Foundation Ireland, experimentaremos con cambios en el formulario de solicitud de subvención que son similares a mi piloto de Manchester. También presentaremos premios, por ejemplo, para la tutoría. Creemos que estos pasos concretos incentivarán la investigación de alta calidad a largo plazo, compensarán algunas de las distorsiones en

40. Sobre la conexión entre evaluación y carrera recomendamos: Assessing scientists for hiring, promotion, and tenure / D Moher. [et al]. PLoS Biol 16(3): e2004089. Disponible en : <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2004089>

41. Mark W. J. Ferguson, DO JUDGE: Treat metrics only as surrogates, en Nature, 26 October 2016, <https://www.nature.com/news/fewer-numbers-better-science-1.20858>

42. The tyranny of metrics / Jerry Z. Müller. Princeton : Princeton University Press, 2018. p. 80 y 176. Disponible en: <https://press.princeton.edu/titles/11218.html>

43. DO JUDGE: Treat metrics only as surrogates / Mark W. J. Ferguson. En: Nature, 26 October 2016, <https://www.nature.com/news/fewer-numbers-better-science-1.20858>

44. Redefine Excellence: Fix incentives to fix science Rinze Benedictus and Frank Miedema. En: Nature, 26 October 2016. Disponible en: <https://www.nature.com/news/fewer-numbers-better-science-1.20858>

el sistema actual y ayudarán a las instituciones a seguir su ejemplo. Si suficientes organizaciones internacionales de investigación y financiadores regresan a los principios básicos en promociones, citas y evaluaciones, entonces tal vez los sustitutos puedan ser utilizados adecuadamente, como información de respaldo. No son fines en sí mismos".⁴⁵

4. ¿BYE, BYE METRICS?

El sistema que rige la evaluación del desempeño de los académicos es un conjunto complejo de normas y prácticas sociales que han evolucionado en el tiempo. Reformarlo implica reemplazar la reputación acumulada a lo largo del tiempo por revistas de prestigio, socavando su poder; que las editoriales de prestigio acepten un rol más acotado, y contar con sistemas de evaluación por pares bien desarrollados y apoyados en marcadores de calidad aceptados por los investigadores. Las declaraciones de San Francisco DORA y Leiden, fueron justamente lideradas por grupos influyentes de investigadores y académicos reconocidos, a las que luego adhirieron cientos de instituciones de todo el mundo.

Sin embargo, muy pocas instituciones o sistemas de investigación logran convertir en acciones y procesos tales principios, ¿cómo convencer a los académicos, científicos, responsables de la gestión de la investigación y financiadores de que la nueva idea es mejor y lograr que transformen sus comportamientos, desarrollados durante décadas?

En ese sentido, las iniciativas de Ciencia Abierta promueven el cambio en la evaluación basado en:

- Desarrollar, contrastar e integrar mediciones disciplinares pertinentes sobre las revistas y artículos e incluir otras publicaciones como datos y resultados específicos de la investigación en ciencias sociales, arte y humanidades;
- Exigir la apertura de los datos de registro de las métricas comerciales e institucionales para facilitar su validación, integración y análisis;
- Promover la evaluación abierta de pares en las publicaciones arbitradas y en las decisiones de financiación, para aumentar la transparencia del proceso y facilitar su mejoramiento;
- Establecer criterios de evaluación basados en el contenido, la calidad y el impacto de los resultados adaptados a las políticas de ciencia, tecnología e innovación, en sus contextos específicos.

También existe un problema de coordinación: la transformación social de un sistema extraordinariamente descentralizado requiere algo más que convencer a un número significativo de agencias de financiamiento

45. Mark W. J. Ferguson, DO JUDGE: Treat metrics only as surrogates, en *Nature*, 26 October 2016, <https://www.nature.com/news/fewer-numbers-better-science-1.20858>

y directivos, estos son procesos impulsados por los académicos y requieren su confianza.⁴⁶

Treinta meses después del informe *The Metric Tide*,⁴⁷ una encuesta entre universidades británicas indicó que, de 96 respuestas, solo 4 universidades habían encarado una acción que podía ser considerada como comprehensiva.⁴⁸ El Prof. James Wildson, coordinador del informe, observó: "Cuando se publicó *The Metric Tide* [...] estábamos alentados por el grado de consenso en apoyo de nuestras recomendaciones. En el día de su lanzamiento, casi todo el mundo, desde defensores de la métrica como Elsevier hasta críticos feroces, ofrecieron su apoyo al informe. Sin embargo, en la política de investigación del Reino Unido, los problemas estaban lejos de ser resueltos [...] En un sentido estricto, la revisión tuvo éxito, en la medida en que el Research Excellence Framework mantuvo a la revisión por pares como su metodología primaria de evaluación. Pero si ganamos esa batalla, ¿qué hay de la guerra?".⁴⁹

Como presentamos, algunas universidades, han emprendido el análisis de estas reformas, y han realizado acciones específicas para promover la apropiación de las iniciativas abiertas. Tratándose de políticas internas de gestión, el análisis de este tipo de incentivos y su efecto no son fácilmente accesibles fuera de las universidades mismas. La experiencia demuestra que el margen de acción que concede la norma a las universidades, en ejercicio de la autonomía en la gestión de recursos humanos, suele ser mayor al efectivamente ejercitado.

La revisión de los sistemas de evaluación y las propuestas de actualización y adaptación avanza en varios frentes, con sus respectivas limitaciones y barreras:

- Las Publicaciones científicas comprometidas con las iniciativas de acceso abierto, sus editores y los investigadores, están desarrollando nuevos sistemas de evaluación para las publicaciones y los datos, que incluyen métricas alternativas, evaluación abierta de

46. Entrevista a Jeff Mackie-Mason en "North, South and Open Access, The view from California" Disponible en: <https://poynder.blogspot.com/2018/04/north-south-and-open-access-view-from.html>

47. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* / James Wildson ed. Higher Education Funding Council for England HEFCE, 2015. Disponible en: https://responsiblemetrics.org/wp-content/uploads/2019/02/2015_metrictide.pdf

48. Puede consultarse los casos de universidades británicas que hasta 2018 lograron iniciar reformas comprehensivas en Bristol (<http://www.bristol.ac.uk/research/environment/responsible-metrics/>), Birmingham (<https://intranet.birmingham.ac.uk/as/libraryservices/library/research/research-metrics.aspx>), Bath (<https://www.bath.ac.uk/corporate-information/principles-of-research-assessment-and-management/>), Loughborough (<https://www.lboro.ac.uk/research/support/publishing/measuring-research-visibility/>), Glasgow (https://www.gla.ac.uk/media/media_555903_en.pdf) y York (<https://www.york.ac.uk/staff/research/governance/research-policies/policy-for-research-evaluation-using-quantitative/>).

49. UK Progress towards the use of metrics responsibly Three years on from *The Metric Tide* report. James Wildson 10 July 2018. p. 22. Disponible en <https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/openscience/The%20Forum%20for%20Responsible%20Research%20Metrics/UK%20progress%20towards%20the%20use%20of%20metrics%20responsibly%2010072018.pdf>

pares y sistemas abiertos de citación que permitan el monitoreo y la transparencia. La validación y reconocimiento de estas nuevas mediciones requerirá tiempo para acumular evidencia que permita el consenso o la normalización para que el cambio supere las falacias de las mediciones actuales basadas en el factor de impacto y los índices de citación.

- Las instituciones financiadoras y los sistemas nacionales de investigación son los responsables de desarrollar nuevas normativas de reconocimiento y evaluación de los investigadores y sus grupos, y de revisar las exigencias en la carrera profesional de los investigadores y los mecanismos de financiación de la investigación. Esta transición exige un análisis cuidadoso sobre los costos actuales de la carrera profesional y el impacto de la adopción de nuevas mediciones en la promoción y reconocimientos; dadas las grandes diferencias entre sistemas y universidades, las soluciones no serán comunes.⁵⁰

El portafolio de evaluación equilibrada propuesto puede orientar debates y acciones concretas en otras instituciones: basadas en principios comunes y compartidos por el sistema de investigación, con un conjunto mixto de indicadores y mediciones cuantitativas y cualitativas adecuadas a la diversidad disciplinar e institucional, analizados en su contexto incluyendo juicios expertos, y con procesos ligeros y flexibles capaces de generar confianza y acumular aprendizajes.

54

5. ¿POR DÓNDE EMPEZAR LOS CAMBIOS AL SISTEMA?

En el marco de Ciencia Abierta, el principal desafío es desarrollar estrategias integradas y multifacéticas, capaces de afrontar la triple complejidad que caracteriza las políticas en ciencia, tecnología e innovación:

- La multiplicidad de roles y responsabilidades: entes legislativos, instituciones de CTI, agencias de financiamiento, universidades e institutos de investigación, sistemas y unidades de información, redes de investigadores, editores de revistas, industrias y emprendedores, organizaciones sociales y, por supuesto los investigadores mismos;
- La multiplicidad de niveles de acción: internacional y nacional (entes supranacionales, ministerios nacionales, agencias de CTI, órganos legislativos, etc.); institucional y sectorial (entes públicos y privados de CTI, sectores económicos y sociales); organizacional y profesional (redes y grupos de trabajo, observatorios, academias, gremios),
- La multiplicidad de procesos del sistema de investigación: desde la regulación y el financiamiento, hasta la evaluación y difusión, pasando por el proceso mismo de investigación y sus actividades.

La interacción entre estos niveles da lugar a tensiones que requieren delicados equilibrios: entre la colaboración y competencia de sectores animados por distintos intereses; entre la investigación de largo plazo y la innovación; entre la protección de la propiedad intelectual y el derecho a la información; entre el acceso libre y gratuito y la sostenibilidad de las plataformas; entre las expectativas de los productores y las demandas de los usuarios de los resultados de la investigación.

La modificación de los sistemas de investigación, las normas de propiedad intelectual, los sistemas de registro de información en CTI, sumadas a las modificaciones de criterios de evaluación, si se hicieran simultáneamente desatarían una reacción contraria a las ventajas del proceso, impidiendo el cumplimiento de sus propósitos de mejora. Por ello, es conveniente que las acciones enfocadas en sectores específicos asuman planes integrales que combinen bloques temáticos específicos y fases de desarrollo con base en acuerdos progresivos.

Las políticas nacionales y estrategias institucionales pioneras (mandatos o promoción de acceso y datos abiertos, incentivos y reconocimiento para los investigadores, servicios e infraestructuras de apoyo) presentan avances que justifican su promoción como medio y marco de acción para la mejora de la efectividad y productividad del sistema científico:

- Permiten externalidades positivas basadas en recursos comunes.
- Facilitan una evaluación comprensiva de los resultados de la investigación.
- Promueven el diálogo y la deliberación sobre la comunicación científica.
- Estimulan la interacción con los sistemas de innovación
- Desarrollan nuevos productos y servicios basados en la investigación.
- Alientan la participación y vinculación de los ciudadanos, y
- Multiplican las oportunidades de colaboración nacional e internacional.

La investigación comparada puede permitir reconocer y anticipar los efectos sistémicos y potenciales de los indicadores. Una estrategia efectiva, es impulsar la producción de estudios de caso para analizar el impacto de los distintos tipos de indicadores.⁵¹ Este tipo de investigación puede permitir también comprender mejor la relación entre la excelencia académica y su impacto social y económico. En este sentido, las métricas alternativas de impacto social pueden ser consideradas como una transición hacia sistemas de investigación más abiertos y dispuestos a una rendición de cuentas como fundamento de su legitimidad.

50. Ciencia Abierta, reporte para tomadores de decisiones. 2. ed. Paola Andrea Ramírez, Daniel Samoilovich. Paris, Montevideo : Unesco, 2019. Disponible en: <http://forocilac.org/wp-content/uploads/2019/05/2019-Policy-PapersCILAC-CienciaAbierta-29-04-2019-Final.pdf>

51. A título de ejemplo, en España, la utilización por parte de ANECA - Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) – de indicadores estandarizados y descontextualizados, con criterios de puntuación juzgados por algunos científicos como poco transparentes - ha dado lugar al término de *anecados*. Ver: El País, Científicos de élite rechazados por la universidad española. Investigadores con ayudas europeas millonarias son descartados por la agencia evaluadora del Ministerio de Ciencia, 4 Agosto 2019. Disponible en: https://elpais.com/elpais/2019/08/02/ciencia/1564743477_599879.html

Ejemplos de impacto económico y social pueden ser observados en la salud de la población, en el desempeño de empresas existentes o en la creación de nuevas, en la formación de personas altamente capacitadas, en la atracción de inversiones en investigación al desarrollo, en la mejora de políticas públicas o de servicios públicos. Estas medidas pueden dar lugar a enfoques más sofisticados sobre la contribución y limitaciones de los indicadores cuantitativos.

En el contexto de una revisión independiente sobre el rol de la métrica en la evaluación y gestión de la investigación, se recomendó otorgar cada año una distinción a la Mala Métrica como ejemplo del uso inapropiado de indicadores cuantitativos en la gestión de la investigación.⁵² Una decisión importante pasa por definir qué peso deben tener los indicadores cuantitativos en la perspectiva de un mejoramiento de los sistemas de investigación a nivel nacional e institucional.

En particular, para los cambios propuestos en evaluación, es necesario, también, llevar a cabo más investigación sobre la investigación. En ese sentido hay un potencial para que la comunidad dedicada al análisis de indicadores juegue un papel más estratégico informando cómo los indicadores cuantitativos son utilizados a nivel de sistema y en las instituciones por quienes toman las decisiones.

Algunas de las preguntas más importantes a considerar son:

- El mérito relativo de los distintos tipos de métrica para analizar las cualidades académicas y los diversos tipos de impacto de la investigación.
- Las ventajas y desventajas de utilizar métricas, en comparación con la revisión por pares, en la creación de un entorno que estimule y aliente una investigación de calidad y relevante.
- Cómo la evaluación basada en la métrica se relaciona con las misiones de la universidad.
- El equilibrio adecuado entre la revisión por pares y las métricas en la evaluación de la investigación en los distintos tipos de disciplina.
- Qué es lo que no puede ser medido con métricas cuantitativas.
- El impacto de la evaluación basada en la métrica en los investigadores individuales.
- El uso de las métricas en aspectos más amplios del gobierno de la ciencia, la innovación y las políticas industriales.⁵³

En resumen, el Reporte del Grupo Experto de la Comisión Europea, Future of scholarly publishing and scholarly Communication, cuya lectura recomendamos para conocer

en detalle los principales argumentos y recomendaciones sobre la comunicación científica en general y la evaluación de las publicaciones en particular, insiste en la importancia de las agencias financiadoras como líderes del proceso de cambio en curso: “Obviamente, el liderazgo asumido por las agencias de financiación deberá contar con el apoyo de los demás responsables. Los financiadores controlan algunas fases estratégicas de la evaluación de la investigación, y colaborar con los investigadores, las universidades y los centros de investigación debería ser bastante directo. Con los editores, está claro que también se necesita cooperación y alentamos a los editores a informar la mayor variedad de evidencia posible para contribuir con información útil para la toma de decisiones informada. Trabajar con el público en general en todas sus formas debe incluir imaginar y crear canales de comunicación que permitan a la población en general ejercer su influencia en las prioridades y orientaciones de la investigación. Por su parte, algunos editores pueden encontrar dificultades para diseñar modelos de negocio que no tengan en cuenta la evaluación de la investigación, y todos los editores tendrán que adaptarse cada vez más a las normas y mandatos que excluyen algunos modelos de negocio.”⁵⁴

Para concluir y confirmar que la coincidencia entre las diferentes fuentes analizadas en este artículo demuestra la fortaleza del consenso actual sobre la necesidad de la reforma a los sistemas de evaluación y las vías de cambio, retomamos un fragmento de nuestro policy brief: Las iniciativas de Ciencia Abierta ofrecen la posibilidad de mejorar los actuales modelos de evaluación de la investigación, necesaria en todas sus facetas. Además de la necesidad de actualizar y diversificar la medición de las revistas científicas, cada sistema o institución, en su contexto particular, puede desarrollar enfoques multifactoriales de evaluación que reconozcan, integren y recompensen todas las contribuciones de los investigadores, que no se limiten a las publicaciones.

En todos estos aspectos, los países de América Latina tienen la enorme oportunidad de compartir logros e innovaciones de las iniciativas de Ciencia Abierta, optimizando la inversión en desarrollo de capacidades y estructuras para mejorar, monitoreando al mismo tiempo, políticas y acciones relevantes a nivel local e internacional.

Para la evaluación de los proyectos que financia, la *Science Foundation Ireland* declara que «Cada proyecto es diferente y, por lo tanto, no existe una metodología única de evaluación para todos. [...] Todos los eventos y proyectos son únicos, por lo que depende de usted aplicar los métodos de evaluación más apropiados para su proyecto». Tomado de: <http://sfi.ie/engagement/sfi-discover/guidance-and-best-practice/>.

Para otros ejemplos de prácticas innovadoras de evaluación de proyectos por parte de agencias de financiamiento de la investigación consultar el registro de buenas prácticas en copiados por la plataforma de trabajo de la Declaración de San Francisco DORA, en: <https://sfdora.org/good-practices/funders/>.

54. Future of scholarly publishing and scholarly Communication. Report of the Expert Group to the European Commission. European Commission, Luxembourg: Publications Office of the European Union. 2019. p.64. Disponible en: <https://publications.europa.eu/s/kIza>.

52. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management / James Wilsdon [et al]. HEFCE, 2015. Disponible en: https://responsiblemetrics.org/wp-content/uploads/2019/02/2015_metrictide.pdf

53. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management / James Wilsdon [et al]. HEFCE, 2015. Disponible en: https://responsiblemetrics.org/wp-content/uploads/2019/02/2015_metrictide.pdf