

Estudio bibliométrico del impacto académico de las publicaciones indexadas de Universidades Colombianas.

NELSON VANEGAS ARBELÁEZ[†]

22 de marzo de 2024.

ÍNDICE

1. Introducción	2
2. Selección de indicadores de impacto de la investigación.	2
3. Metodología.	5
4. Análisis.	5
5. Discusión.	8
6. Conclusiones y recomendaciones.	10
7. Anexo técnico.	12

ABSTRACT

Usando fuentes abiertas de información, se estudia el impacto académico que tienen las publicaciones académicas en las que al menos un autor declara está basado en Colombia. La revisión hecha es anónima (no se tiene en cuenta el autor) y se agrupa por institución. Debido a la complejidad se hace un análisis que solo comprende una selección de Universidades, por ello no es exhaustivo. El objetivo es tener una caracterización y una medida de la fracción de publicaciones que logran alto impacto (o son susceptibles de lograrlo). El análisis es cuantitativo y se hace una comparación con algunas instituciones latinoamericanas reconocidas. Se sacan algunas conclusiones basado en los resultados obtenidos.

[†] Instituto de Física, Universidad de Antioquia, Medellín, Colombia.
nelson.vanegas@udea.edu.co

1 INTRODUCCIÓN

Colombia pasó de tener muy pocas publicaciones académicas, científicas o técnicas, a varios miles anuales en los últimos 20 años. Las razones son diversas pero una de ellas es el incentivo a aumentar la productividad científica propiciado por políticas públicas tales como las normas que fijan la remuneración de los profesores de las Universidades públicas. Esto se ha propagado a Universidades de carácter privado y de ahí un impulso a trabajar más en función de publicar.

Igualmente, las medidas de aseguramiento de la calidad de la Educación Superior, específicamente las establecidas en procesos obligatorios tales como la obtención del Registro Calificado o voluntarias como la Acreditación de programas o Instituciones, han motivado a las Instituciones a promover la creación de grupos e investigación.

Numéricamente el aumento de los productos tipo artículo es enorme y se puede revisar en las plataformas de Minciencias. Sin embargo, es importante discutir otros aspectos no solo numéricos con respecto a las publicaciones hechas en Colombia al menos parcialmente. En este escrito se da una idea del impacto de estas publicaciones. Y se buscarán algunas comparaciones con Latinoamérica buscando aclarar e insinuar caminos hacia mejoras en el sistema de investigación y modelo colombianos.

No es el propósito de este artículo encontrar relaciones causa efecto, es decir, concluir sobre cuál práctica o cultura académica de cual Institución es mejor en Colombia. Nuestro análisis no llega ahí, como tampoco llega a elucidar o intentar encontrar la relación de publicaciones y sus características con los *rankings* de universidades.

Este trabajo entonces busca más caracterizar nuestro trabajo académico, técnico y científico (repetimos que la creación artística queda bastante fuera de nuestro alcance). Esto esperamos que contribuya a buscar políticas acordes con nuestra realidad.

De la misma manera el análisis que hacemos usará solo fuentes públicas de información. Es decir, bases de datos de publicaciones, artículos y producción académica que sean de libre acceso. Estas fuentes son diversas y no siempre arrojan los mismos resultados. Con lo cual es también importante decir que las conclusiones posibles tienen esa limitación como es apenas obvio: de dónde llega la información.

Finalmente hay una motivación no menor para este estudio, buscar relaciones y medidas que permitan mejorar la productividad científica nacional. Hay una relación entre Investigación, medida en publicaciones científicas, y desarrollo, ver p.ej. [10], [12] y [13]. Y la economía colombiana, diríamos la industria y el sector productivo, necesitan de mejoras en innovación. En 2022 Colombia obtuvo 29.2 de 100 puntos según el *Global Innovation Index* de acuerdo a [11] que corresponde al puesto 63 entre las 132 economías estudiadas. Vale decir, la economía colombiana ocupa aproximadamente el puesto 44 en tamaño de economía en el mundo.

2 SELECCIÓN DE INDICADORES DE IMPACTO DE LA INVESTIGACIÓN.

Medir el impacto de una publicación académica es bastante difícil. Publicaciones que hoy no parecen tener mucho eco en algún momento futuro pueden terminar consiguiendo mucha

visibilidad internacional. Es apenas obvio que esto depende de que la publicación produzca o proponga resultados o métodos o ideas que las hagan ganar prestigio y reconocimiento. Es así como se pueden usar medidas del impacto de las citaciones logradas por el trabajo de un autor, como el *H-index*, pero con precaución.

Sin embargo, las medidas de impacto *individual* al ser muy desagregadas no permiten estudiar tendencias o aspectos macro que son relevantes y en los que nos vamos a concentrar en este trabajo. En ese sentido, para este análisis vamos a usar información agregada por Institución y nuestra búsqueda será hacia artículos producidos y en los que la afiliación de al menos uno de los autores es en Colombia. Buscamos con ello visibilizar tendencias o características generales de las publicaciones colombianas y de ahí visualizar qué las diferencia.

Para este trabajo entonces seleccionamos:

1. La base de datos (dataset) de producción académica de OpenAlex que es abierta al público y que contiene más de 250 millones de artículos indexados, incluyendo no solo los más conocidos títulos de revistas sino muchas fuentes de artículos en idiomas distintos al inglés y basados en lugares diversos del mundo [1]. Adicionalmente se han detectado numerosas ventajas al uso de OpenAlex con respecto a Scopus o Web of Science [2].
2. Hemos usado *Scimago Journal and Country Rank* [3], para extraer información de citaciones. En este trabajo se van a usar dos indicador de esta organización: SJR y *H-index*. El primero de ellos es un indicador de *prestigio* de una revista especializada. Este indicador usa las citaciones que ha tenido un artículo pero también el reconocimiento de la revista donde dicho artículo fue citado. Es decir, si un artículo que es citado en una revista de mucho reconocimiento dará lugar a un mejor indicador SJR para esa revista que una cita que proviene de una revista poco conocida. Es una transferencia de prestigio. Se trata también de información abierta y accesible por el público en general [4] y [5]. No toda revista que aparece en OpenAlex tiene un SJR, algunas revistas no lo tienen (algunas razones pueden ser que no son indexadas, que tienen muy poca visibilidad, algunas son subsecciones de otras publicaciones, etc.)
3. Igualmente el índice *h* fue propuesto como una medida del impacto en citas de los autores. Sin embargo, se ha ido agregando y ahora se cuenta con el Parámetro de Impacto *H* o *H-index* que tiene en cuenta las citas en general de los artículos publicados por una revista. También puede ser consultado abiertamente y es accesible por el público [6]. Casi toda revista tiene algún valor para este indicador.

Los indicadores mencionados pueden ser usados libremente y se cuenta con las (*API – Application programming Interface*) de forma abierta. En esta selección de indicadores se pueden escapar trabajos o artículos por múltiples razones. Sin embargo, afecta a todas las instituciones por igual así que no se espera que esta selección perjudique una institución en especial o agregue un sesgo. Sin embargo, el uso de información abierta añade el valor de volver reproducible el trabajo o ampliarlo a otras Universidades (en el anexo se dan las instrucciones, ver §7).

Institución	Multi-campus	Ciudad Ppal.	Prof.	Preg.	Posg.	I.D.
U. Nacional de Colombia - UN	Sí	Bogotá DC	2900	46000	9000	I36243813
U. de Antioquia - UdeA	Sí	Medellín	1800	35000	4000	I3018704611
U. de los Andes - Uniandes	No	Bogotá DC	740	14000	4500	I162096671
U. del Valle - Univalle	Sí	Cali	1470	30000	2750	I91732220
U. Industrial de Santander - UIS	No	Bucaramanga	x	x	x	I115684694
U. EAFIT	No	Medellín	x	x	x	I862322245
U. Pontificia Bolivariana - UPB	No	Medellín	x	x	x	I181028852
U. Javeriana	No	Bogotá	x	x	x	I233745408
Escuela de Ingeniería de Antioquia - EIA	No	Medellín	x	x	x	I3018704611
U. Externado	No	Bogotá	x	x	x	I59986765
U. del Norte - Uninorte	No	Barranquilla	x	x	x	I142879360
U. del Rosario	No	Bogotá	x	x	x	I90803817
U. de Medellín	No	Medellín	x	x	x	I4210133585
Inst. Tecn. Metropolitano - ITM	No	Medellín.	x	x	x	I4210141572
U. de Buenos Aires - UBA	No	Buenos Aires	x	x	x	I24354313
Instituto Politécnico Nacional - IPN	No	México DF	x	x	x	I59361560
U. Estadual Paulista - UNESP	Sí	Sao Pablo	x	x	x	I879563668
U. Católica de Chile	No	Santiago	x	x	x	I162148367

Cuadro 1: Todas las cifras redondeadas y aproximadas, tomadas de los sitios oficiales de cada institución, presentadas como referencia al tamaño de cada una.

3 METODOLOGÍA.

Se seleccionaron universidades colombianas y en el exterior usando el criterio de incluir instituciones:

- tanto públicas como privadas.
- que sean una muestra incluyendo las regiones.
- que haya algunas en *rankings* internacionales y otras que no.
- con diferentes vocaciones, unas más inclinadas a lo técnico y otras más generales o más a las humanidades o ciencias sociales.
- de diferentes tamaños, medidas por número de estudiantes y profesores.

Lo anterior para tener un corte de Instituciones (en realidad todas son Instituciones de Educación Superior – IES) que sea una muestra del sistema de educación superior. Infortunadamente, debido al gran volumen de IES existentes no es viable adelantar en este punto un análisis de todas y cada una. Sin embargo, se provee el *script* utilizado y se explica el software necesario para que este análisis pueda hacerse por parte de cualquiera, ver anexo §[7].

Con respecto a las Instituciones del exterior, se seleccionaron buscando que tengan características con alguna similitud a las colombianas de la muestra. Es decir, que fueran de latonamérica y economías comparables, de tamaños también comparables (descartando algunas demasiado grandes como la Universidad Nacional Autónoma de México) y estructuras también similares (por ejemplo, con campus en varias ciudades). Finalmente se seleccionaron Universidades latinoamericanas que tienen un buen posicionamiento en términos de prestigio internacional. Usando esos criterios se seleccionaron las Universidades de la tabla [1].

Usando los *scripts* provistos ya mencionados, se procede a estudiar la información para cada institución a lo largo de los últimos 11 años. De este primer corte se hizo una comparación de las Universidades de la selección más significativas y se procede a hacer una comparación para esa selección en comparación con las instituciones de latinoamérica estudiadas.

4 ANÁLISIS.

Un primer resultado que surge luego de revisar los datos *sumados* de todas las instituciones muestra que en promedio hay un descenso en la productividad científica y académica de Colombia en los últimos años. Una posible causa podría ser la pandemia de 2020-2021, sin embargo hay que considerar que los procesos de publicación implican tiempos en ocasiones de meses entre envío de un artículo y publicación del mismo. Observando las gráficas de la figura 1. Aunque no es el propósito de este estudio, notamos el descenso y queda la hipótesis de que sea una tendencia más extendida.

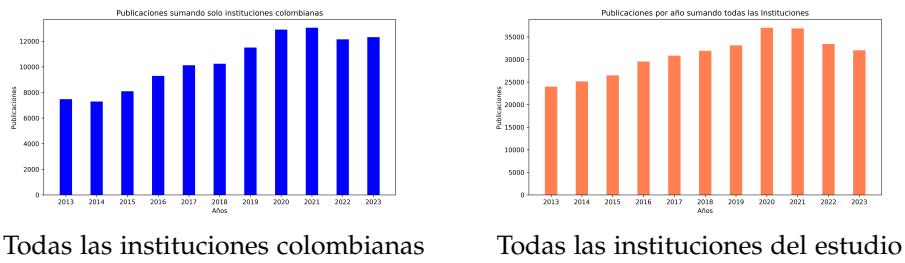


Figura 1: Agregado de todas las publicaciones e instituciones en el período 2013-2023.

Las columnas de la figura 1 deberían bajar un poco cuando se considera que en ellas se suman artículos que son producidos en colaboración entre investigadores de universidades seleccionadas y pueden estar dos veces. Una estimación de esto, por ejemplo, puede hacerse notando la cuadro 2 que la U. de Antioquia comparte con la U. de los Andes, excepto pequeñas diferencias, el número de artículos publicados debido a que ambas pertenecen a la misma colaboración – CMS. No así la U. Nacional que pertenece a la colaboración Atlas. Sin embargo, los totales de la figura 1 solo no bajarían muy significativamente.

Institución	2019	2020	2021	2022	2023
Uniandes	98	74	78	87	42
U.Nacional	49	89	107	113	96
U .de Antioquia	69	72	78	85	35

Cuadro 2: Publicaciones de Universidades Colombianas en colaboraciones de la Organización Europea para Investigación Nuclear – CERN. Fuente: <https://inspirehep.net>

A paso seguido queremos ver algunas instituciones (todas están disponibles en el repositorio, pero se han seleccionado unas pocas), para establecer diferencias y características generales de los datos. En la figura 3 podemos ver algunas generalidades.

La diferencia de altura entre cada columna de indicador, izquierda para JSR y derecha para H-index, se debe a que no toda revista tiene asignado un indicador JSR. En efecto, la reputación de una revista se gana al ser citada repetidamente por artículos en revistas con JSR, lo cual requiere tiempo. En ese sentido es notorio que las publicaciones de nuestras instituciones colombianas sucede frecuentemente en publicaciones sin índice JSR (solo H). En las fuentes dadas de scimago es claro que la probabilidad de tener citaciones aumenta con JSR.

Sin embargo, se observa en la misma figura que las Universidades del exterior presentadas, la diferencia entre estas columnas es indetectable. Es decir, la tasa de publicación en revistas sin JSR es casi nula y en el caso colombiano puede llegar a ser notoria, dependiendo de la institución.

Lo segundo que notamos de las gráficas y datos es que las publicaciones en Colombia se centran en revistas con bajo índice JSR. Esto se nota en que en la columna de la izquierda de los gráficos predominan el primer segmento de $0 \leqslant \text{JSR} < 2$ y en el segundo $5 \leqslant \text{JSR} < 5$. De ahí

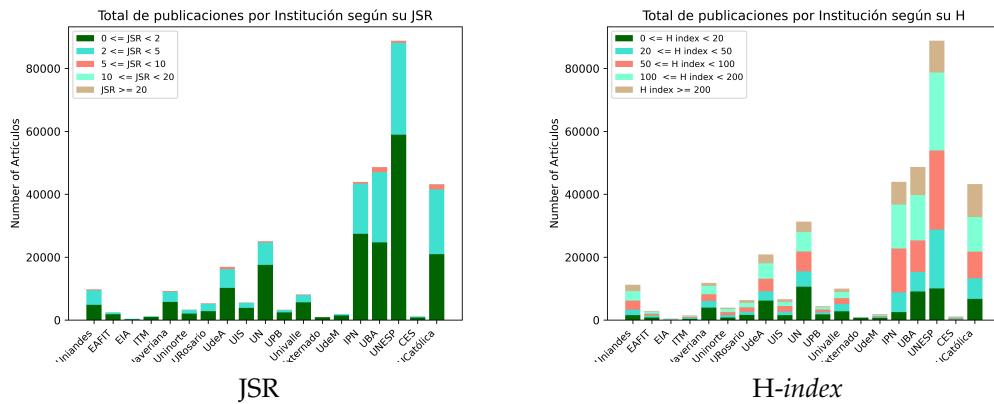


Figura 2: Producción para las instituciones en el período 2013-2023.

en adelante muy pocas instituciones tienen producción. Podemos decir que prácticamente toda la producción está concentrada en revistas sin JSR o menor que cinco.

Analizando la otra columna, es también claro que la producción se centra en revistas en los primeros tres rangos de *H-index*. Aunque mejor distribuida y estos gráficos no hay una conclusión aparente.

En términos de valores brutos de las publicaciones podemos agregar por institución las publicaciones, en la figura 2 se muestra la suma de todo el período por institución y su distribución en rangos. Es notorio el factor de escala con las instituciones del exterior escogidas. En particular La UNESP. Sin embargo, todas tienen mucha más producción que incluso la mejor ubicada de Colombia, la Universidad Nacional. En especial, es importante comparar esta diferencia con la configuración de cada Universidad en la tabla 1 que muestra número de profesores y estudiantes de pre- y posgrado. No hemos considerado pertinente retirar a la UNESP del estudio aun cuando la escala se distorsiona para mantener una institución de Brasil.

Notando entonces las diferencias de escala y buscando centrarnos en las instituciones colombianas que más aportan en términos de publicaciones, se seleccionan las de mayor número de publicaciones para el siguiente paso.

El análisis final tiene solo se usa el *H-index* toda vez que lleva más información. En la figura 4 izquierda se puede identificar para cada Universidad qué tanto aporta cada rango de *H* al total.

Lo que podemos ver en la misma figura es que en nuestras universidades, incluso de las de mejor producción, esta se concentra en rangos de $H\text{-index} \leq 100$. Que es algo que podemos verificar en la figura a la derecha. Para las instituciones colombianas en la selección la fracción promedio con rango mayor que 100 es 65.4 % mientras que en las extranjeras es 53.7 %. Si excluimos la Universidad de los Andes el porcentaje de publicaciones con $H\text{-index} \leq 100$ aumentaría más.

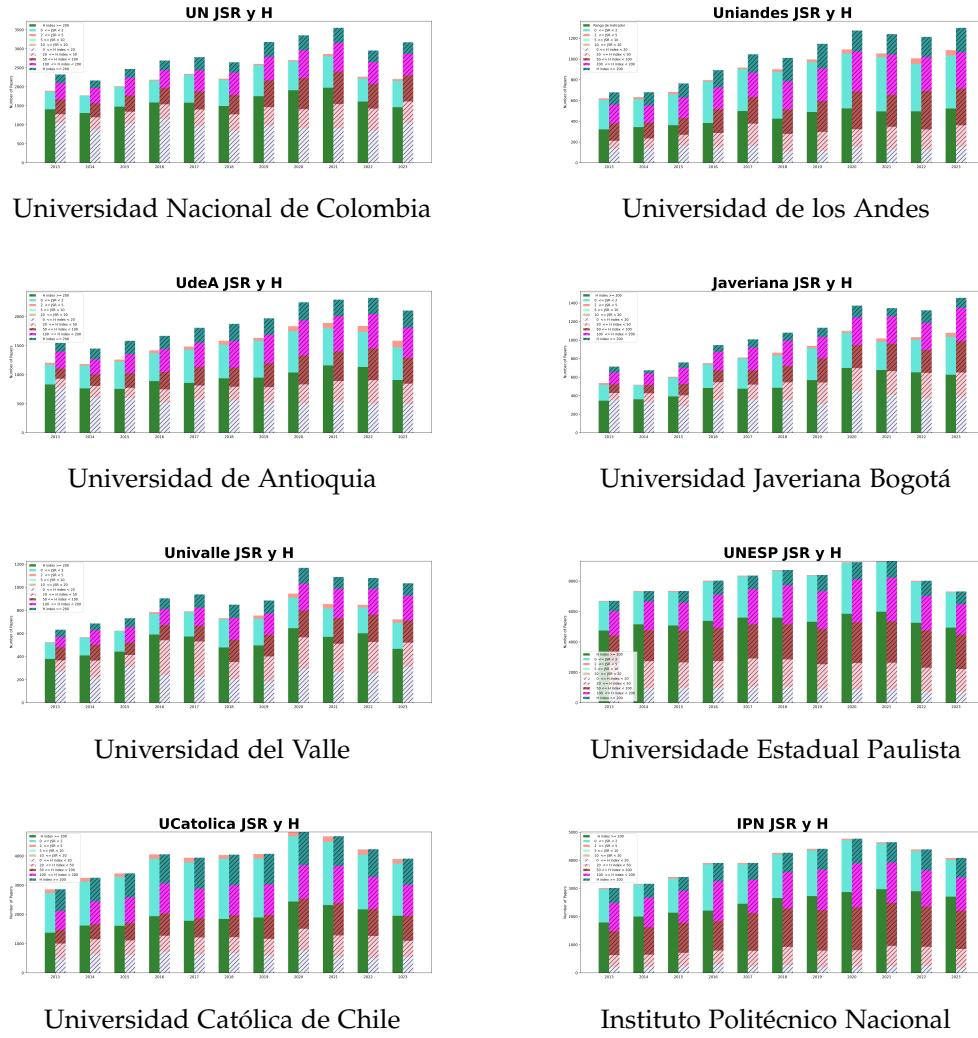


Figura 3: Producción por rango del indicador H y JSR para varias instituciones en el período 2013-2023.
Notar la diferencia de escala vertical en cada gráfico.

5 DISCUSIÓN.

Los resultados presentados pueden tener algunos errores por cuanto hemos usado bases de datos y no hemos verificado registro por registro lo encontrado. Sin embargo, los errores en general son aleatorios y afectan a cada institución por igual. Pueden deberse a artículos que deletrean mal el nombre de la institución o errores similares. Dado que los autores generalmente necesitan mostrar esta producción como propia, típicamente esto se corrige.

Igualmente hay un error al contabilizar dos veces las colaboraciones. Las Universidades citan como propia los artículos de sus investigadores y por ende un mismo artículo puede ser contado dos veces (esto afecta especialmente los agregados nacionales). Sin embargo, aun en los casos de colaboraciones prolíficas como las del CERN, el número no es en suma muy importante para el total.

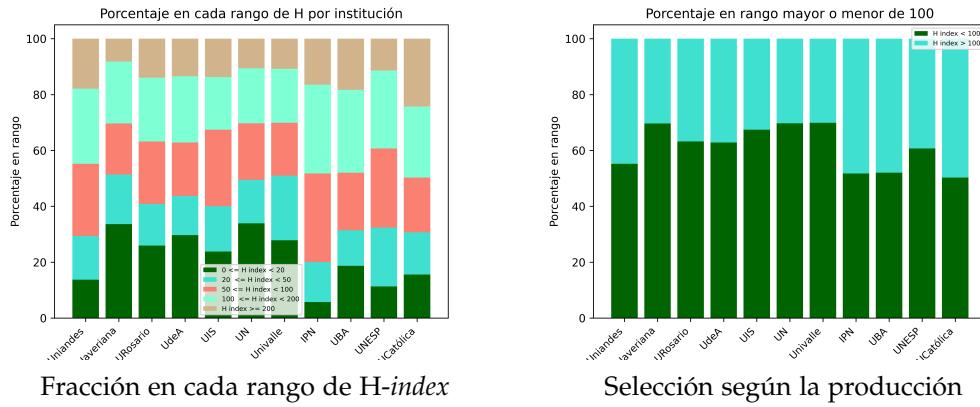


Figura 4: Porcentaje de la producción en rangos de H-index en el período 2013-2023.

Es sí importante ese número al considerar la producción con los mejores H-index. Esto por cuanto colaboraciones internacionales tienden a publicar en revistas de alto impacto (en el caso del CERN suelen ser revistas de un alto impacto). Una consulta a [14] y a [15] muestra que las revistas usuales de las colaboraciones son Journal of High Energy Physics ($H=247$), Physical Review Letters ($H=674$), Physical Review D ($H = 378$), Physics Letters, section B ($H= 275$) y Europhysics letters ($H=170$). Con lo cual la fracción agregada como país a publicaciones de alto impacto se ve reducida según la tabla 2.

Un ruido imposible de filtrar completamente es la práctica de algunas instituciones de contratar personal de investigación en el exterior que en la práctica no está basado en Colombia. aunque aparece en las bases de datos como tal. Por asuntos de comunidades académicas fuertes las Instituciones seleccionadas se minimiza este riesgo. Es algo que existe en el sistema colombiano debido a la presión de tener investigación y acreditarla como sea para procesos con Minciencias o Mineducación y algunas IES toman atajos para mostrárla.

Igualmente algunos investigadores adoptan la práctica de repetir las publicaciones en diferentes revistas. Esto es delicado pero suele afectar revistas de poco impacto así tomamos como que este ruido contribuye más al sector de publicaciones de bajos indicadores y no afecta las conclusiones en general.

Finalmente, no toda investigación o trabajo creativo es susceptible de publicaciones en revistas de alto impacto. Algunos trabajos son importantes solo a nivel local, por ejemplo, y no llevan necesariamente a publicaciones internacionales. Lo mismo sucede con producción tecnológica o aplicada. En este sentido este artículo se abstiene de hacer juicios al respecto y como es claro en título este trabajo es bibliométrico, debemos abstenernos de sacar conclusiones en otras dimensiones.

6 CONCLUSIONES Y RECOMENDACIONES.

Se sientan bases y métodos para indagar y ahondar en la caracterización de la producción académica basada en investigación que se hace en Colombia, aunque sería de interés establecer una metodología que incluya otros tipos de producción académica.

Los resultados dan evidencia de que se necesitan ajustes en la investigación en Colombia que incentiven la producción académica de alto impacto internacional (medido en número de citaciones y visibilidad) que nos lleven a los niveles al menos de las economías comparables de la región. Es indudable que para el Estado colombiano es costoso mantener la investigación que se hace y se deben buscar formas de mejorar los resultados de la investigación con los recursos disponibles o invertir recursos frescos con un propósito más estratégico.

REFERENCIAS

- [1] OpenAlex, <https://openalex.org/>, Ingreso en: 2024-03-07.
- [2] Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus. Jack Culbert, Anne Hobert, Najko Jahn, Nick Haupka, Marion Schmidt, Paul Donner and Philipp Mayr. 2024. <https://arxiv.org/abs/2401.16359v1>.
- [3] <https://www.scimagojr.com/aboutus.php>. Ingreso en: 2024-03-07.
- [4] <https://libguides.daemen.edu/c.php?g=1239513&p=9072137>.
- [5] <https://lib.guides.umd.edu/bibliometrics/SJR>. Ingreso en: 2024-03-07. Ingreso en: 2024-03-07.
- [6] <https://answers.library.westernsydney.edu.au/faq/269743>. Ingreso en: 2024-03-07.
- [7] Python Software Foundation. Python Language Reference. <https://python.org>
- [8] Kluyver, T. Ragan-Kelley, B. Pérez, F. *et. al.* "Jupyter Notebooks – a publishing format for reproducible computational workflows". Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press. 2016. pp. 87-90.
- [9] <https://www.mongodb.com/docs/manual/administration/install-on-linux/> Ingreso en 2024-03-20.
- [10] Ekolu, Stephen. "Relationship between Research, Innovation and Development-a Review." EAI International Conference for Research, Innovation and Development for Africa. 2018.
- [11] <https://www.statista.com/> Ingreso en 2024-03-20.
- [12] Luong, H.M. Hewitt-Dundas, N. The interrelationship between R&D, Innovation and Productivity: Evidence for micro-enterprises. Queen's Management School and UK Enterprise Research Centre. Apr. 2020.

- [13] Committee on Assessing the Value of Research in Advancing National Goals; Division of Behavioral and Social Sciences and Education; National Research Council; Celeste RF, Griswold A, Straf ML, editors. Furthering America's Research Enterprise. Washington (DC): National Academies Press (US); 2014 Oct 28. <https://www.ncbi.nlm.nih.gov/books/NBK253897/>; doi:10.17226/18804
- [14] Revistas según su H-index. https://www.scimagojr.com/journalrank.php?order=h&ord=desc&area=3100#google_vignette Accedida en 2024-03-22.
- [15] Base de datos de publicaciones en física. <https://inspirehep.net/literature?sort=mostrecent&size=25&page=1&q=find%20collaboration%20cms> Accedida en 2024-03-22.

7 ANEXO TÉCNICO.

Para el presente estudio se utiliza y es necesario disponer del siguiente software:

1. Python 3, con bibliotecas numpy, matplotlib y panda [7].
2. Jupyter notebooks [8].
3. Mongodb community edition, [9].

El *script*, los *notebooks*, tablas, figuras y análisis están disponibles en <https://github.com/nvanegas/ImpactoDeInvestigaciones> excepto los archivos .csv que se crean. Estos deben ser bajados usando el *script* que se provee para cada institución.

Las instrucciones para reproducir los resultados son:

1. Copiar el repositorio de *github* mencionado arriba.
2. Iniciar *jupyter* y *mongodb*.
3. En cada subdirectorio (por institución) verificar el *script* para bajar la información de OpenAlex que lleva la palabra *download* en el nombre y verificar que en la tercera celda la "I.D." de la institución de interés coincida con la dada en la tabla 1.
4. Ejecutar completo el *notebook* de *jupyter* de bajar información (puede llevar algún tiempo). Esto debe repetirse para cada institución.
5. Inmediatamente después entrar y ejecutar el *notebook* de análisis en cada institución. Los diccionarios de *python* que resultan se llevan a tablas. Cada que se ejecuta uno de los *notebooks* de bajar la información la base de datos cambia y no se debe mezclar el análisis de una universidad con la información bajada para otra.
6. Verificar que la tabla dada en el repositorio de *github* llamada "nuevo-todos-3.csv" coincide con los datos encontrados en el punto anterior. Es posible pequeñas diferencias pero no deben ser significativas.
7. Ejecutar el *notebook* de *jupyter* llamado "analisis-global-completo.ipynb" y verificar tablas y gráficos.