METODOLOGÍA PARA EL CÁLCULO DE LOS INDICADORES PRODUCCIÓN CIENTIFICA BIBLIOMÉTRICA

Contenido

1. INTRODUCCIÓN 4

2. ANTECEDENTES 5

2. DISEÑO DE LA OPERACIÓN ESTADISTICA 7

2.1.1 DISEÑO TEMÁTICO/METODOLÓGICO 2

2.1.2 Necesidades de Información 2

2.1.3 Objetivos 3

Objetivo general 3

Objetivos Específicos 4

2.1.3. Alcance 4

2.1.4 Marco de referencia 8

I. Marco teórico 8

II. Marco conceptual 10

III. Marco legal 14

IV. Referentes internacionales 16

V. Referentes nacionales 18

2.1.5 Definición de variables y construcción de indicadores estadísticos 21

2.1.6 Plan de resultados 30

2.1.7 Diseño de cuadros de salida o de resultados 31

2.2 Diseño del formulario o cuestionario 31

2.3 Normas, especificaciones o reglas de validación, consistencia e imputación . 32

2.4 Nomenclaturas y clasificaciones utilizadas 32

3 DISEÑO ESTADÍSTICO 33

3.1 Universo de estudio 33

3.2 Población Objetivo 33

3.3 Cobertura geográfica: Nacional 33

3.4 Desagregación geográfica: Se presenta a nivel nacional 33

3.5 Desagregación temática: Tipo de producto, tipo de autoría, área OCDE, promedio de citas 33

3.6 Fuente de datos (tipo de operación) 34

3.7 Comprobación de datos fuentes secundarias 34

3.8 Unidades estadísticas 35

3.9 Período de referencia: Ventana de observación de 4 años, partiendo desde el año respectivamente anterior, es decir, si la operación estadística se ejecuta en el año 2023, la ventana de observación será 2019 a 2022. 35

3.10 Período de recolección/ acopio: dos meses finales del año de referencia. 35

4 DISEÑO DE LA RECOLECCIÓN/ACOPIO 36

4.8 Métodos y estrategias de recolección o acopio de datos 36

4.9 Estructura organizacional del operativo y definición del equipo requerido 37

4.10 Esquema de entrenamiento del personal 39

4.11 Proceso de sensibilización y acuerdos de intercambio 39

4.12 Elaboración de instructivos 39

4.13 Diseño de la estrategia de comunicación y plan de contingencias 40

4.14 Diseño de la estrategia de seguimiento y control 40

4.15 Diseño de sistemas para la obtención de datos 40

4.16 Transmisión de datos 41
5 DISEÑO DEL PROCESAMIENTO 41
5.1 Consolidación de archivos de datos 41
5.2 Revisión y validación 42
5.3 Diseño para la generación de cuadros de resultados 43
5.4 Seguridad y almacenamiento de la información 43
5.5 Diseño del plan de pruebas 44
6. DISEÑO DEL ANÁLISIS 44
5.6 Métodos de análisis de resultados 44
5.7 Anonimización 46
5.8 Verificación de anonimización 47
7. DISEÑO DE LA DIFUSIÓN Y COMUNICACIÓN 47
7.1 Diseño de productos de difusión y comunicación 47
7.2 Entrega de productos 48
7.3 Estrategia de servicio 48
8. DISEÑO DE LA EVALUACIÓN DE LAS FASES DEL PROCESO 48
8.1 Mapa de procesos 51
9. DOCUMENTACIÓN RELACIONADA 52
10. BIBLIOGRAFÍA 54
11. CONTROL DE CAMBIOS 56
12. APROBACIÓN 56
13. ANEXOS 57
Anexo 1. Glosario 57
Anexo 2. Especificaciones de validación 59

1. INTRODUCCIÓN

La publicación es un punto importante en el proceso de la investigación. Es a través de este medio que los científicos e investigadores socializan sus hallazgos, descubrimientos y reflexiones. El poder de comunicación que tiene la publicación ha ocasionado que esta se convierta en un elemento básico en la valoración de las capacidades de investigación que existen en un país, en una institución o en una temática específica.

Sin embargo, el gran crecimiento de las publicaciones y la tendencia hacía una ciencia cada vez más global requieren de procesos de organización, clasificación y distinción de la organización. Los servicios de indexación se convierten en una alternativa de organización y clasificación de las publicaciones de acuerdo con criterios de calidad, visibilidad y relevancia de las revistas en las que se encuentran incluidas.

Los servicios de indexación de mayor prestigio son el Web of Science (WoS) y Scopus. Su buen nombre está basado en parámetros técnicos de cobertura, tanto temática como regional, y en los criterios que utilizan para seleccionar las revistas que incluyen en sus bases, puesto que la mayoría de las revistas incluidas tanto en el Web of Science como en Scopus son editadas en Estados Unidos o Europa, y existe una predominancia de publicaciones en inglés.

Por esta razón, indicadores basados en las revistas incluidas en el SciELO Citation Index y en la Red de Revistas Científicas de Acceso Abierto de América Latina, y el Caribe (Redalyc) son importantes. Para el caso exclusivo de Colombia, existe el índice de publicaciones científicas seriadas, Publindex. Sin embargo, este no permite la construcción de indicadores a nivel de publicación, mientras que Scielo.

Los indicadores bibliométricos aquí presentados se construyen a partir de la identificación de los investigadores adscritos a instituciones colombianas y que se encuentran en Web of Science (WoS), Scopus, en SciELO Citation Index, identificadas como las bases con características técnicas apropiadas para esta operación estadística y que permiten identificar las características de las publicaciones científicas, como analizar las prioridades de investigación, procesos

de difusión y transmisión del conocimiento generado en la investigación, entre otros.

El presente documento metodológico describe cada una de las fases del proceso estadístico aplicado a la operación estadística denominada ***Estadística de la Producción Científica Bibliométrica*** que se desarrolla siguiendo la guía para la elaboración del documento metodológico de operaciones estadísticas establecida por el Departamento Administrativo Nacional de Estadística (DANE).

2. ANTECEDENTES

En los primeros años de la década de 1990 nace la Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICyT), auspiciada por la Organización de los Estados Americanos (OEA) y el Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED). Para ese entonces, los países de la región, entre ellos Colombia, manifestaban la necesidad de contar con indicadores de ciencia y tecnología (CyT) confiables, como resultado en el año de 1995 se realiza en el país una reunión preparatoria para el Taller Iberoamericano llamada "**Hacia la Construcción de un Observatorio de CyT**" promovida por el Ministerio de Ciencias, Tecnología e Innovación (MinCiencias), anteriormente llamada Colciencias, y el DNP y en 1996 se realizael Segundo Taller Iberoamericano sobre Indicadores de CyT, organizado por Minciencias, RICyT y OEA, llevándose a cabo en Cartagena.

Como resultado, MinCiencias encargó un estudio de factibilidad técnica y financiera y llevó a cabo un prelanzamiento del Observatorio Nacional en Medellín en julio de 1998. La formalización oficial de la creación del Observatorio Colombiano de Ciencia y Tecnología (OCyT) se realizó en agosto de 1999. La creación del OCyT fue claramente impulsada por MinCiencias, con el objetivo de fortalecer el sistema nacional de ciencia y tecnología, con el propósito de apoyar los procesos de toma de decisiones en materia de investigación e innovación a través del diseño, producción y difusión de indicadores de ciencia, tecnología e innovación (CTeI). La misión del OCyT en el marco del Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) es fundamental para el seguimiento de su dinámica, pues consiste en brindar información y conocimiento que permita a los principales actores del Sistema, la toma de decisiones y la definición de políticas y estrategias.

MinCiencias delegó la producción de indicadores al OCyT, otorgándole autonomía e independencia.

Desde su creación, el OCyT ha dado una especial importancia a la construcción de representaciones sobre la dinámica de la producción científica bibliográfica colombiana, reconociendo la importancia de no sólo medir, sino además mostrar lo relacionado con los procesos investigativos por medio de indicadores que presentan las tendencias con relación a las publicaciones colombianas donde es importante el proceso de organización, clasificación y distinción.

Fue así, como una de sus primeras tareas fue el apoyo a la construcción de las primeras fases de implementación del Sistema de Indexación de Revistas Científicas Nacionales y el Sistema de Homologación de revistas que en la actualidad son administrados por Minciencias los cuales son conocidos bajo el nombre de Publindex.En un principio fue necesario atender la construcción conceptual de lo que debe ser entendido como producción científica bibliográfica, lo que conllevó a un trabajo de definición de los estándares que debían cumplir los objetos informacionales que apoyaría el diseño y construcción de indicadores. Esto permitió identificar que, si bien era posible construir representaciones a partir de la producción en revistas internacionales (Web of Science y Scopus), existía un vacío de información sobre la producción científica en otras revistas científicas que no se encontraban indexadas en estas bases. Debido a la preocupación por considerar distintos niveles de circulación de los conocimientos producidos en el país, se realizó una alianza con el Sistema de Información Científica Redalyc, para estudiar la producción científica de autores asociados a instituciones colombianas.

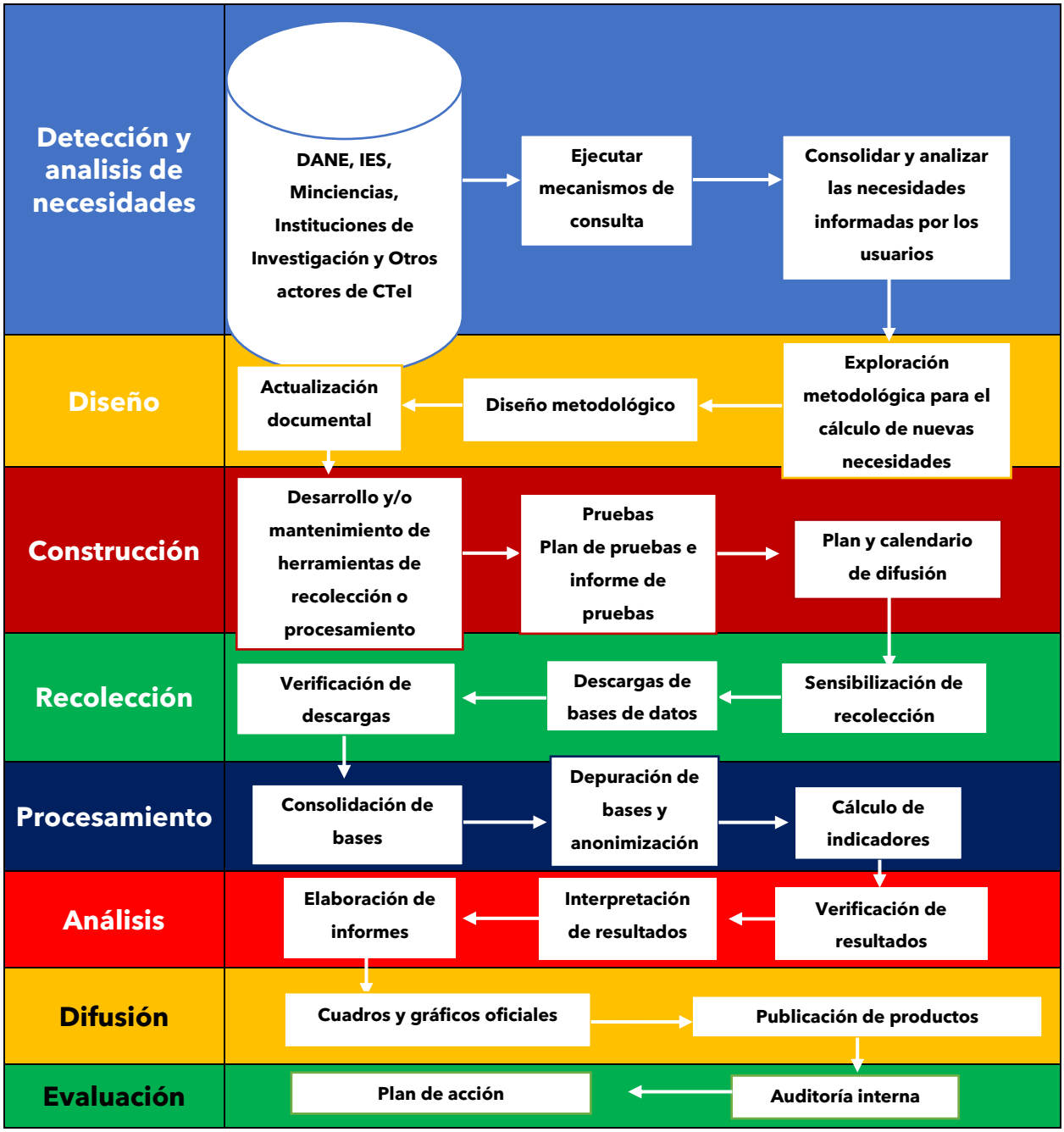
Para identificar la producción científica de los investigadores adscritos a las instituciones colombianas que tienen visibilidad internacional, se consultan las bases bibliográficas de WoS Clarivate y Scopus. Así, se recopila la información a partir de la afiliación institucional reportada por los autores en el momento de su publicación. Las bases WoS Clarivate y Scopus recogen distintos tipos de documentos que incluyen: artículos y artículos en conferencia (proceedings), abstracts de conferencias, correcciones y fe de erratas, material editorial, cartas, reseñas y capítulos de libros y software.

Para la construcción de los indicadores provenientes de estas bases de datos, el OCyT en compañía de expertos bibliométricos y estadísticos, construyó un algoritmo que permite generar, sistematizar y unificar las bases de datos, con el objetivo de contar productos únicos, e identificar las categorías o desagregaciones necesarias, tales como: tipología del documento, tipo de coautoría, áreas de conocimiento y conteo de citas.

2. DISEÑO DE LA OPERACIÓN ESTADISTICA

En este capítulo se describen la estructura y los métodos que se pretenden desarrollar con el objetivo de satisfacer las necesidades de información del usuario. En el gráfico 1, se muestra de manera esquemática el flujo de trabajo que se desarrolla en la operación estadística.

Figura 1. Flujo de trabajo



Fuente. Observatorio Colombiano de Ciencia y Tecnología

2.1.1 DISEÑO TEMÁTICO/METODOLÓGICO

2.1.2 Necesidades de Información

En la operación "Estadística de Producción Científica Bibliométrica", se han identificado y confirmado diversas necesidades de información en interacción con los usuarios pertenecientes a diferentes sectores, tanto nacionales como internacionales. Estas necesidades se han determinado a través de consultas con actores clave del Sistema Nacional de Ciencia y Tecnología, incluyendo entidades gubernamentales, instituciones de educación superior, centros de investigación y desarrollo tecnológico, así como asociaciones profesionales y organismos internacionales.

Requerimientos Sectoriales de Información:
Las entidades del poder ejecutivo nacional y territorial requieren estadísticas confiables sobre la producción científica para la formulación y evaluación de políticas públicas en ciencia y tecnología. Estas estadísticas son fundamentales para medir el cumplimiento de las metas establecidas en los planes de desarrollo y en políticas específicas como el CONPES 3582. Además, la información es crucial para la construcción de instrumentos que faciliten la asignación de recursos públicos y la caracterización de las entidades involucradas en el proceso científico y tecnológico.

Por otro lado, las instituciones de educación superior y los centros de investigación utilizan los resultados de esta operación estadística como herramientas para la certificación de centros de investigación, seguimiento de resultados de investigación, análisis comparativo entre instituciones y fortalecimiento de redes de conocimiento.

La operación estadística proporciona una base sólida para la toma de decisiones informadas en la política pública. La información generada permite a las entidades del gobierno y a otros actores del Sistema Nacional de Ciencia y Tecnología evaluar el desempeño del país en términos de generación de conocimiento y su posicionamiento en la arena científica global. Esto es particularmente relevante para la definición de estrategias que buscan mejorar la visibilidad internacional y la calidad de la investigación realizada en Colombia.

Criterios y Mecanismos Utilizados para la Priorización de Necesidades:

La priorización de las necesidades de información se ha llevado a cabo mediante un proceso de consulta continua con los usuarios de la operación estadística, incluyendo reuniones y talleres de retroalimentación con expertos y representantes de las diferentes instituciones involucradas. Se han utilizado criterios como la relevancia para el diseño de políticas públicas, la necesidad de información para el seguimiento de metas de desarrollo, y la capacidad de las estadísticas para reflejar la realidad del entorno científico y tecnológico colombiano.

Además, se ha implementado un mecanismo de soporte y consulta a través de un formulario de contacto disponible en la página web del Observatorio Colombiano de Ciencia y Tecnología (OCyT)[1]. Este formulario permite a los usuarios plantear sus necesidades, dudas o sugerencias de manera directa, facilitando una interacción continua y efectiva con el equipo encargado de la operación estadística. Esta herramienta ha sido clave para asegurar que las necesidades de información de los usuarios sean identificadas y priorizadas de manera oportuna y eficiente.

Las principales conclusiones de este proceso indican que la operación estadística es fundamental para diversos sectores, y su información es considerada crítica para la planificación y gestión de la ciencia y la tecnología en Colombia. Los usuarios han valorado especialmente la capacidad de la operación estadística para ofrecer datos desagregados a nivel nacional y departamental, así como para permitir el análisis comparativo entre diferentes áreas del conocimiento y tipos de coautoría.

2.1.3 Objetivos

Objetivo general

Proveer información estadística anual, nacional y departamental, sobre la producción científica (productos únicos) que tienen los autores vinculados a las instituciones colombianas, en revistas de alta visibilidad internacional y que se encuentran indexadas en las principales bases de datos Web of Science de Clarivate, en Scopus de Elsevier y en Scielo.

[1] https://ocyt.org.co/contacto/

Objetivos Específicos

- ✓ Brindar información sobre la producción bibliográfica de autores vinculados a la institución por tipo de producto a nivel nacional y departamental y que se encuentran indexadas en las principales bases de datos Web of Science de Clarivate, en Scopus de Elsevier y en Scielo.

- ✓ Producir información estadística acerca de la producción bibliográfica de autores vinculados a la institución por tipo de coautoría a nivel nacional y departamental y que se encuentran indexadas en las principales bases de datos Web of Science de Clarivate, en Scopus de Elsevier y en Scielo.

- ✓ Generar información estadística acerca de distribución de la producción bibliográfica de autores vinculados a la institución por área OCDE a nivel nacional y departamental y que se encuentran indexadas en las principales bases de datos Web of Science de Clarivate, en Scopus de Elsevier y en Scielo.

- ✓ Producir información estadística sobre el promedio de citas de documentos publicados por autores vinculados a la institución a nivel nacional y departamental y que se encuentran indexadas en las principales bases de datos Web of Science de Clarivate, en Scopus de Elsevier y en Scielo.

2.1.3. Alcance

La operación estadística "Producción Científica Bibliométrica" se enfoca en medir y analizar la producción científica de los autores afiliados a instituciones colombianas. Específicamente, se investigará la producción de documentos científicos publicados en revistas de alta visibilidad internacional, indexadas en las principales bases de datos: Web of Science, Scopus y SciELO. Los aspectos que se investigarán incluyen:

- **Producción bibliográfica:** Se medirán el número y tipo de documentos (artículos científicos, artículos cortos, notas editoriales, reseñas, memorias de conferencias, entre otros) publicados por autores colombianos.

- **Tipo de coautoría:** Se analizarán los documentos según la modalidad de coautoría, clasificándolos en coautoría internacional, coautoría nacional y autoría única. Del número de documentos construidos en cooperación en tres modalidades:

  i) Coautoría Internacional: Un documento se considera en esta categoría si cuenta con más de un autor y por lo menos uno de los autores cuenta con una afiliación institucional del extranjero. ii) Coautoría nacional: Un documento se considera en esta categoría, si cuenta con más de un autor y todos los autores están vinculados a instituciones colombianas. iii) Autoría única (Solo autor): Un documento se considera en esta categoría, si cuenta con un único autor vinculado a una institución colombiana.

- **Distribución por áreas de conocimiento:** Se realizará una homologación de las subáreas de Scopus y categorías de Web of Science a las áreas según la clasificación OCDE, evaluando la distribución de la producción científica por estas áreas.

De la distribución porcentual de la producción científica por área de conocimiento, para su construcción, el primer paso consistió en la homologación de las subáreas de Scopus y categorías de WoS a las áreas según clasificación OCDE. Lo anterior permite el conteo de productos científicos únicos en cada una de las áreas, siempre y cuando se trate de un producto con clasificación en una única área de la ciencia OECD. Sin embargo, se debe aclarar que algunos productos reportados pueden pertenecer a una o varias áreas de conocimiento, dependiendo del número de áreas de conocimiento que el producto reporto en la revista indexada, estas se clasifican en como ciencias multidisciplinarias.[2]

- **Citas recibidas:** Se calculará el promedio de citas que han recibido los documentos desde su publicación.

[2] Revisar las bases de homologación

Temas No Incluidos y Justificación de su Exclusión:

Publicaciones en revistas no indexadas: No se incluirán publicaciones en revistas que no estén indexadas en Web of Science, Scopus, o SciELO, ya que estas no cumplen con los criterios de visibilidad y calidad requeridos para esta operación.

Publicaciones técnicas o divulgativas: Documentos de naturaleza técnica, divulgativa o de menor rigor científico no serán considerados, dado que el objetivo es medir la producción científica que tiene un impacto reconocido internacionalmente.

Publicaciones en lenguas locales fuera del ámbito académico internacional: Las publicaciones en lenguas locales que no están presentes en las bases de datos mencionadas se excluyen por no tener el alcance y reconocimiento internacional necesario.

Delimitación Geográfica

La operación estadística tiene una cobertura geográfica que abarca todo el territorio nacional de Colombia, con desagregación a nivel departamental y municipal, cabe aclarar que esta desagregación no es publicada, pero está disponible para remitirla ante cualquier necesidad de información de un usuario.

Justificación de Exclusiones Geográficas

No se consideran exclusiones geográficas dentro de Colombia en la operación estadística. Sin embargo, las publicaciones que no pueden ser asociadas claramente a una ubicación geográfica específica en Colombia o que están vinculadas a instituciones no colombianas no serán parte de los análisis desagregados por departamento.

Delimitación temporal

Período de Observación por frecuencia de ejecución del operativo. La operación estadística abarca una ventana de observación de los últimos 4 años, lo cual permite analizar tendencias a mediano plazo en la producción científica nacional, incluir ajustes.

Período de recolección

La recolección de datos se lleva a cabo durante los dos últimos meses del año de referencia, asegurando que se capture la mayor cantidad de publicaciones posibles dentro del período de observación.

Justificación de la delimitación temporal

El período de observación de 4 años fue seleccionado para proporcionar un análisis robusto y evitar las fluctuaciones temporales que podrían distorsionar los resultados si se considerara un período más corto. Un período más largo, en cambio, podría incluir cambios estructurales en el sistema científico que no reflejen la situación actual con precisión. El período de recolección en los dos últimos meses del año asegura que la mayoría de las publicaciones del año de referencia hayan sido indexadas y estén disponibles para su inclusión en el análisis.

- ✓ De la distribución porcentual de la producción científica por área de conocimiento, para su construcción, el primer paso consistió en la homologación de las subáreas de Scopus y categorías de WoS a las áreas según clasificación OCDE. Lo anterior permite el conteo de productos científicos únicos en cada una de las áreas, siempre y cuando se trate de un producto con clasificación en una única área de la ciencia OECD. Sin embargo, se debe aclarar que algunos productos reportados pueden pertenecer a una o varias áreas de conocimiento, dependiendo del número de áreas de conocimiento que el producto reporto en la revista indexada, estas se clasifican en como ciencias multidisciplinarias.

- ✓ Del promedio de veces que los documentos han sido citados por los otros documentos dentro de la misma base de datos, desde el inicio de su publicación. Este periodo permite capturar datos de nuevas revistas fuentes que son indexadas en los SIRes de referencia. Para los documentos más antiguos, al estar disponibles por un lapso mayor, aumentan las posibilidades de ser citados.

2.1.4 Marco de referencia

I. Marco teórico

A continuación, se describen los aspectos conceptuales que soportan parte del análisis propuesto para la medición de la producción de conocimiento de investigadores vinculados a las universidades en el país**.** Lo anterior, en la medida que el conocimiento de carácter científico se inscribe en el desarrollo de los sistemas de innovación a nivel nacional o regional; comprenden la configuración de arreglos institucionales entre entidades del orden gubernamental, las empresas y el sector académico, donde los sistemas de investigación en las universidades cumplen un papel estratégico(Millar et al., 2016)

- Enfoque del modelo insumo - producto de las actividades científicas

La investigación por parte de las universidades contiene procesos de entrada y de salida, donde los procesos de entrada son entendidos como los insumos con los que las universidad cuentan para la realizar investigación, estos comprenden los recursos humanos, la inversión y su interacción para generar nuevo conocimiento, siendo este último el proceso de salida (Vargas-Quesada & de Moya Aragón, 2007). Al respecto, Benjam Godin citado por la OCDE (Organización para la Cooperación y el Desarrollo Económico) indica que: "este modelo lineal promovido por la OCDE y sus países miembros, con el objetivo de medir las actividades de ciencia, tecnología y sociedad, en las que el conocimiento se reconoce como elemento principal para el desarrollo económico, sin embargo no permite observar la interacción de las relaciones entre la ciencia y la sociedad"(Organización para la Cooperación y el Desarrollo Económicos (OCDE), 2017).

Los procesos de salida se relacionan con los resultados obtenidos, es decir la producción científica y tecnológica. Este proceso debe ser visto como un conjunto de factores sociales y disciplinarios que interactúan entre sí. Según Diets y Bozeman citados por (Arango-Botero et al., 2019) estas relaciones permiten el fortalecimiento de las competencias incorporadas en cada

individuo, que se traducirán en la generación de nuevo conocimiento. Sin embargo, para este estudio la producción científica se abordará solo como el insumo final generado por el investigador.

Dentro del análisis realizado por Rivera et al 2013, en el cual articulan los insumos del SNCyT y sus actores involucrados, se expone una aproximación a la interacción entre ellos, a partir de dos momentos diferentes pero complementarios.

- Modo 1 y 2 de la producción de conocimiento científico y tecnológico

En la literatura, a partir del trabajo de Gibbons et al, citados por (Ferreira & Ramos, 2015) coexisten dos modos de producción de conocimiento, como lo definen estos autores, el modo 1, consiste en investigación básica y aplicada, que genera impactos directos en la demanda del mercado. Este tipo de investigación fue aplicada principalmente por las universidades, generando que las investigaciones realizadas estuvieran dentro de un marco lógico de sus intereses. El modo 2 se desarrolla dentro de un contexto de aplicación con mayor inclusión interdisciplinaria, permitiendo la inserción de nuevos actores que garantiza mayor diversidad en la manera producción científica.

Como lo indican (Salazar et al., 2014) la investigación científica en Colombia por diferentes mecanismos aún avanza en realizar el tránsito del modo 1 al modo 2 de producción de conocimiento, con el que aparecen nuevos actores que interactúan entre sí, diversificando las problemáticas de investigación y las aproximaciones disciplinares propuestas para abordarlas, lo que configura formas y practicas heterogéneas de producir conocimiento.

En este orden de ideas, las actividades de producción de conocimiento se asocian al rol de los científicos bajo la tutela de las universidades, que en el pasado evidenciaron baja interrelación con otros actores de la sociedad aunado a la autonomía con que contaban los científicos para definir las líneas de trabajo y las agendas de investigación, esperando que luego de un proceso de apropiación y uso del conocimiento generado, dieran tránsito a innovaciones para el conjunto de la sociedad.

Aunque el tránsito del modo 1 al modo 2 debe involucrar distintos actores, el rol de la universidad en la generación de conocimiento científico y tecnológico, se mantiene y en otras palabras, ha sido robustecido, gracias al fortalecimiento de la investigación y los procesos de formación científica y tecnológica que se adelanta a nivel de maestría y doctorado, así como los niveles de diversificación y especialización de la investigación desarrollada. Los factores antes señalados, dan cuenta en alguna medida, de la expansión de las capacidades de investigación que, en distinto grado, han alcanzado las universidades en Colombia.

II. Marco conceptual

El análisis de los resultados de investigación visto desde la perspectiva de productos científicos, han presentado en las últimas décadas avances pertinentes con el objetivo de buscar garantizar la calidad, la globalización del conocimiento y la pertinencia de las publicaciones científicas, por medio del uso de nuevas tecnologías de información que han facilitado la forma de interactuar con estos productos.

Para comprender estas dinámicas, es importante analizar las fuentes y los orígenes de estos sistemas de indexación y resumen. La mayoría de ellos nacen como modelos de catalogación para organizar y visibilizar un conocimiento específico, como es el caso de la ISI ahora Web of Science –WoS, el cual nació con el propósito de organizar y visibilizar la ciencia(Garfield, 1972), que con los años, los avances tecnológicos y nuevos modelos de análisis de información que ayudaron al desarrollo de la cienciometría y la bibliometría; este modelo de base citacional fue tomando mayor prestigio y reconocimiento entre las comunidades académicas; un modelo similar es la base de datos citacional de Scopus Elsevier

Estos sistemas han sido líderes en el desarrollo de la organización y evaluación de la producción del conocimiento, pero igual también han hecho parte de los vertiginosos cambios, un ejemplo de ello es la venta de la ISI WoS al grupo Clarivate.

Algunos modelos de indexación de información han surgido desde un enfoque disciplinar a partir de áreas de conocimiento especializadas donde prima el prestigio y los criterios específicos de cada área de conocimiento, ejemplo de ello

es Philosopher Index, Chemical Abstract, Medline, Psyinfo, entre otras; que son curadores del conocimiento y abogan por su producción, indexación y difusión con criterios específicos de cada área de conocimiento.

Las nuevas tecnológicas han creado nuevos modelos de producción científica colaborativa bajo conceptos de ciencia abierta, ciencia 2.0, e-science, sciencecommon, ejemplos de ello son: https://arxiv.org/, Public Library of Science PLOS http://www.plos.org/, CogPrints, CitesserX, Dimensions, Lens, OpenAlex. Otro modelo que ha tomado fuerza es el de acceso abierto, sobre todo en Latinoamérica casos como Redalycy La Referencia ,Scielo, Clacso, los cuales abogan por el acceso libre y sin costo del conocimiento (Fishman et al., 2015). De acuerdo con Orduna Malea et al. Citado por (Uribe-Tirado & Alhuay-Quispe, 2017), en los último años Google Scholar ha comenzado a mejorar su modelos de análisis de conocimiento científico usando criterios de citación y relevancia que junto con herramientas como Publisher or Perish ha facilitado la forma de analizar la producción académica. A pesar de esto se sigue considerando Scholar un sistema de caja negra pues no brindan información sobre políticas de cosecha de datos.

Para tal fin, el OCyT observa sistemas de indexación que tiene en cuenta las transformaciones digitales que han generado el desarrollo de los nuevos sistemas de información y modelo de evaluación de recursos digitales en línea. Por ello, se han tomado desde las ciencias de la información dos conceptos claves: recuperación y consumo de información como base para definir los criterios de evaluación y la aplicación de la metodología.

De igual manera, se tienen en cuenta otros criterios basados en los análisis disciplinares como son las ciencias de la información (cienciometría, bibliometría, altimetría); filosofía de la ciencia, sociologías de la ciencia, políticas científicas (evaluación de I+D). A continuación, se presentan los principales criterios que se tienen en cuenta:

i) Pertinencia geográfica (local, regional, internacional): Para ello, se analizan las revistas indexadas en los principales sistemas de información, a nivel internacional (WoS, Scopus- Scimago) y regional (Scielo), para identificar donde se encuentran indexadas las mejores revistas en cada uno de estos niveles verificando los Sistemas de Indexación y Resumen -SIRES-de

mayor indexación y que permiten evidenciar diferencias geográficas en la medida que estas se generen.

ii) Criterios disciplinares: Desde la perspectiva de la comunicación científica cada área de conocimiento tiene unas formas de producir, comunicar y difundir sus propios procesos disciplinares. Por ello se ha tenido en cuenta este criterio tomando las áreas del conocimiento propuestas por la OECD para productos con única área reportada y la clasificación ciencias multidisciplinarias para productos con múltiples áreas reportadas.

iii) Visibilidad y accesibilidad: Estos criterios están acompañados con el uso de sistemas e información que permiten el aumento de la visibilidad y accesibilidad de la producción científica, la metodología desarrollada por webmetrics (http://www.webometrics.info/en) o la forma de indexación de la información que permita la eficiencia en la gestión de la información y el conocimiento que lleven a realizar o el uso de software para el análisis de vigilancia científica son la base en este punto. En la matriz de caracterización se identifican estos puntos, que incluyen variables de indexación de información que permitan el uso y consumo de información acorde a las características de los gestores de información y del conocimiento que se reflejan principalmente en la eficiencia de los buscadores en las bases de datos.

De otra parte, los conceptos básicos que enmarcan la operación estadística son:

- **Bases de datos Bibliométrica**: Recursos de información que suministran no solo información bibliográfica sino indicadores relativos a distintas facetas (impacto, citación, obsolescencia, etc.) de las publicaciones científicas recogidas, tanto referidos a autores, como a instituciones, publicaciones, países, materias u otros agentes intervinientes. Entre las principales podemos señalar Web of Science, Scopus, Scielo.

- **Bibliometría**: Disciplina que emplea métodos y técnicas cuantitativas para medir y evaluar la publicación de la investigación científica y las relaciones entre los agentes que la producen(Pritchard, 1969).

- **Filiación institucional:** denominación por la cual un autor indica expresamente en su publicación a qué institución pertenece.

- **Indicadores bibliométricos:** miden distintos aspectos cuantitativos sobre la generación, uso e impacto de la producción académica dentro de un área temática o un espacio geográfico determinado (Larivière & Sugimoto, 2019).

- **Indicadores de colaboración:** miden las relaciones que han existido entre los productores o agentes científicos y que han finalizado con la publicación conjunta de resultados científicos. Se basan en los datos de autoría (Newman, 2004).

- **Indicadores de producción:** se basan en el recuento del volumen de publicaciones académicas o científicas. Son los más fáciles de utilizar ya que nos proporcionan información sobre las características de las unidades analizadas. El número de publicaciones es una medida razonada de la actividad científica, pero no del progreso científico. Estos indicadores están condicionados por factores políticos y sociales. Miden la productividad científica, que es la cantidad de publicaciones producida por un autor, país o institución durante un período determinado (Zárate Rueda et al., 2017).

- **Indicadores de visibilidad o impacto:** Se basan en los recuentos del número de citaciones que reciben los documentos durante un período determinado de tiempo o a partir de su publicación(Chubin & Garfield, 1980)

- **Publicación académica:** Publicación resultado de actividades de investigación, cuyo objetivo es aportar al desarrollo de un área científica, técnica o humanística determinada. Este tipo de publicación se suele difundir a través de artículos, libros, actas de congresos, etc.

- **Áreas de conocimiento**: son estructuras en las que se organiza gran parte del saber humano. En estas áreas suele haber métodos socialmente establecidos para producir conocimiento, así como normas sobre lo que se considera un hecho o una buena explicación

III. Marco legal

El observatorio de Colombiano de Ciencia y tecnología como organismo responsable de la producción de las estadísticas de producción científica bibliométrica, se enmarca en la normativa establecida para el Sistema Nacional de Ciencia, Tecnología e Innovación. En particular a continuación se mencionan algunos aspectos normativos que de manera indirecta se relacionan con la importancia y el respaldo que se tiene para la generación de las estadísticas arriba mencionadas:

Acuerdo 008 de 2008 de la Junta Directiva del OCyT: que de acuerdo con lo establecido en su parágrafo 2 del artículo 1 "MinCiencias como principal socio promotor, fijará para cada vigencia el monto del aporte de acuerdo a su disponibilidad presupuestal"

Estatutos del OCyT - agosto de 1999 y reformas de mayo de 2000 y mayo de 2015, dentro de sus objetivos específicos están:

- Constituirse en instrumento fundamental de apoyo para la formulación de políticas, tanto públicas como privadas.
- Dinamizar la visibilidad nacional e internacional de la actividad científica y de innovación tecnológica del país.
- Desarrollar modelos analíticos para la interpretación de indicadores y el relacionamiento de indicadores cualitativos y cuantitativos.
- Apoyar al Sistema de Educación Superior en la construcción de indicadores de actividades científicas y tecnológicas, de recursos humanos y financieros en Ciencia y Tecnología y de investigación científica, desarrollo e innovación tecnológicos, así como en el seguimiento de impacto.

- **Constitución Política de Colombia (1991):** La Constitución establece principios fundamentales para la promoción del acceso a la información y la participación ciudadana, incluyendo la producción de estadísticas confiables.

- **Decreto 1781 de 1994:** Reglamenta la creación del Observatorio Colombiano de Ciencia y Tecnología (OCyT), definiendo su rol en la recolección y difusión de indicadores de ciencia y tecnología.

- **Ley 1286 de 2009:** Transformó a Colciencias en el Departamento Administrativo de Ciencia, Tecnología e Innovación (Minciencias) y fortaleció el Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI).

- **Ley 1581 de 2012 - Ley de Protección de Datos Personales:** Establece disposiciones para la protección de datos personales, asegurando que la información recolectada sea tratada de manera confidencial y usada únicamente con fines estadísticos.

- **Decreto 4170 de 2011:** Crea el Departamento Administrativo Nacional de Estadística (DANE) como entidad coordinadora del Sistema Estadístico Nacional (SEN).

- **Ley 1712 de 2014 - Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional:** Garantiza el derecho de acceso a la información pública, promoviendo la transparencia y la rendición de cuentas.

- **NTC 1486:2006 - Gestión Documental:** Define los requisitos para la gestión documental aplicables a la administración y archivo de datos en operaciones estadísticas.

- **NTC PE 1000:2020:** Norma técnica que establece los requisitos generales para la planificación, ejecución, control y evaluación de las operaciones estadísticas, asegurando que estas cumplan con estándares de calidad reconocidos internacionalmente.

- **Ley 2335 de 2023 - Ley de Estadísticas Oficiales:** Establece un marco jurídico e institucional para la producción y administración de estadísticas

oficiales en Colombia, reforzando la independencia técnica, la confidencialidad de los datos y la calidad de la información estadística.

IV. Referentes internacionales

En el caso de la "Producción Científica Bibliométrica" no existen claros referentes internacionales ni manuales indicativos para la producción de este tipo de indicadores bibliométricos. Sin embargo, si existen algunas entidades (con y sin ánimo de lucro) que hacen mediciones comparativas entre los diferentes países, como la RICYT (en el caso de entidades sin ánimo de lucro) o algunas empresas privadas que hacen estudios bibliométricos, a nivel de consultorías. También hay esfuerzos de algunas entidades internacionales, como el BID o la UNESCO, en la capacitación para diferente tipo de usuarios en la construcción de estos indicadores. Finalmente, vale la pena revisar algunas experiencias de países cercanos que, dentro de sus estadísticas de CTI, producen indicadores bibliométricos, España y México, entre otros.

- *Red de indicadores de ciencia y tecnología Iberoamericana e Interamericana – RICYT*

  En el marco de sus actividades permanentes, la RICYT calcula para todos los países Iberoamericanos, diferente tipo de indicadores, dentro de esos los indicadores bibliométricos; es así como anualmente se producen indicadores, por país, sobre publicaciones en Science CitationIndex (SCI) – prácticamente el mismo WoS –, en Scopus, en Medline, en Periodica, en Clase y en Lilacs. Adicionalmente presenta información, para Scopus y SCI con relación a la población, el PIB el el gasto en I+D. También calculan indicadores por disciplina científica y colaboración internacional, únicamente para el caso de Scopus. http://www.ricyt.org/indicadores

- *BID, UNESCO*

  Diferentes esfuerzos para la capacitación de personas en el cálculo de indicadores bibliométricos.

  BID – Módulo de capacitación parala recolección y el análisis deindicadores de productode las actividades deciencia y tecnología

http://docs.politicascti.net/documents/Doc%2007%20-%20capacitacion%20prat.pdf

UNESCO – Curso virtual: construcción de indicadores bibliométricos. Dictado por el Centro Redes de Argentina.
http://www.unesco.org/new/es/media-services/single-view-tv-elease/news/curso_virtual_construccion_de_indicadores_bibliometricos/

- España

Gobierno de España, Ministerio de Economía, Industria y Competitividad, FECYT

Indicadores bibliométricos de la actividad científica española 2005–2014. Edición 2016.Se publica desde 2002 (se basa solo en Scopus y ScienceDirect de Elsevier).
https://icono.fecyt.es/sites/default/files/filepublicaciones/indicadores_bibliometricos_2016_0.pdf

- México

México, Foro Consultivo Científico y Tecnológico, AC, Indicadores de Ciencia, Tecnología e Innovación. 2016. Dentro de los indicadores publicados anualmente por el CONACYT existen, entre otros: Publicaciones de mexicanos incluidos en el ISI, por año y por disciplina; Citas recibidas por disciplina; Participación de la producción mexicana en el total de producción mundial por disciplina. Básicamente son indicadores producidos de ISI (WoS).
http://www.foroconsultivo.org.mx/libros_editados/INDICADORES_CTI.pdf

- GLOBAL INNOVATION INDEX

Número de artículos de Publicaciones científicas y tecnológicas en revistas indexadas (por mil millones de PPP $ PIB) a | 2018 El número de artículos científicos y de ingeniería publicados en esos campos, que incluyen: agricultura, astronomía, astrofísica, sistemas de control de automatización, bioquímica, biología molecular, conservación de la biodiversidad, biotecnología, microbiología aplicada, biología celular, química, informática,

tecnología de construcción, odontología oral medicina quirúrgica, ingeniería, ciencias ambientales, ecología, biología evolutiva, tecnología de ciencias de los alimentos, medicina interna general, biomedicina de ciencias de la vida y otros temas, biología marina de agua dulce, ciencia de materiales, biología computacional matemática, matemáticas, metalurgia e ingeniería metalúrgica, meteorología ciencia atmosférica, microbiología, ciencia y tecnología nuclear, física, ciencias de las plantas, radiología, medicina nuclear, imágenes médicas, biología reproductiva, investigación en medicina experimental, tecnología científica y otros temas, telecomunicaciones, transporte y ciencias veterinarias. El recuento de artículos proviene de un conjunto de revistas cubiertas por el Science Citation Index (SCI) y el Social Sciences Citation Index (SSCI). Los artículos se clasifican por año de publicación y se asignan a cada país / economía en función de las direcciones institucionales enumeradas en el artículo. Los artículos se cuentan por conteo (en lugar de hacerlo de forma fraccionada), es decir, para artículos con instituciones colaboradoras de múltiples países / economías, cada país / economía recibe crédito sobre la base de sus instituciones participantes. Los datos se informan por mil millones de PPP $ PIB. Fuente: Clarivate Analytics, tabulaciones especiales de Thomson Reuters, Web of Science, Science Citation Index (SCI) e Social Sciences Citation Index (SSCI); Fondo Monetario Internacional, World Economic Outlook Database, octubre de 2018 (PPP $ PIB). https://www.globalinnovationindex.org/gii-2019-report

Como se puede observar, a pesar de existir algunos referentes internacionales, en ninguna de las experiencias presentadas existe la práctica de construir una base única sobre publicaciones científicas a partir de diferentes bases internacionales.

V. Referentes nacionales

No existen entidades que produzcan operaciones estadísticas nacionales sobre la información de los documentos de autores afiliados a instituciones colombianas publicados en revistas indexadas en el Web of Science de Clarivate, en Scopus de Elsevier y en Scielo y que se fundamenten en un algoritmo que permite la homologación de los productos (documentos de autores) presentes en las tres bases de datos con el fin de identificar productos únicos.

Sin embargo, se tienen presentes los siguientes dos actores como referentes en función a su papel en la producción científica:

Ministerio de Ciencia, Tecnología e Innovación
Lineamientos y Políticas que Afectan la Producción Científica en Colombia
El Ministerio de Ciencia, Tecnología e Innovación (Minciencias) es la entidad encargada de formular, coordinar y orientar las políticas públicas en materia de ciencia, tecnología e innovación en Colombia. Sus lineamientos y políticas juegan un papel fundamental en la producción científica del país, ya que establecen el marco normativo y estratégico que orienta la investigación científica, el desarrollo tecnológico y la innovación.

Política Nacional de Ciencia, Tecnología e Innovación (CTI)
Esta política define los objetivos y prioridades para el desarrollo de la CTI en Colombia, promoviendo la generación de conocimiento científico y tecnológico que responda a las necesidades del país. Los lineamientos incluyen la promoción de la investigación en áreas estratégicas, el fortalecimiento de las capacidades científicas, el fomento a la internacionalización de la ciencia colombiana, y la articulación entre la academia, el sector productivo y el Estado.

Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI)
El Minciencias lidera este sistema, el cual es el conjunto de actores, instituciones y procesos que interactúan para promover la ciencia, tecnología e innovación en Colombia. Dentro de este sistema, se establecen las directrices para la financiación de proyectos científicos, la evaluación de la calidad de la investigación, y la promoción de la transferencia de conocimiento al sector productivo y a la sociedad en general.

Fomento a la Investigación y Desarrollo (I+D)
Minciencias impulsa la investigación mediante programas de financiamiento que buscan fortalecer las capacidades investigativas de las universidades, centros de investigación y empresas. Las políticas incluyen incentivos fiscales para la I+D, fondos concursables para proyectos de investigación y programas de becas para la formación de investigadores.

Ética y Buenas Prácticas en Investigación

El Ministerio también establece lineamientos éticos para la realización de investigaciones científicas en Colombia, promoviendo prácticas responsables y transparentes en la producción de conocimiento. Estos lineamientos buscan garantizar la integridad científica, la protección de los derechos de los participantes en investigaciones, y el respeto por el medio ambiente y la biodiversidad.

DANE. Normas y Estándares para la Producción y Difusión de Estadísticas Oficiales. El Departamento Administrativo Nacional de Estadística (DANE) es la entidad responsable de la producción y difusión de las estadísticas oficiales en Colombia. Sus normas y estándares son esenciales para asegurar que la información estadística sea precisa, confiable y comparable a nivel nacional e internacional.

Normas Técnicas y Metodológicas

El DANE establece normas técnicas que regulan la recolección, procesamiento, análisis y difusión de datos estadísticos. Estas normas buscan asegurar la calidad y consistencia de las estadísticas producidas, y son aplicables tanto a las operaciones estadísticas que realiza el DANE como a aquellas realizadas por otras entidades públicas y privadas bajo su supervisión.

Política de Calidad Estadística

El DANE implementa una política de calidad que abarca desde el diseño de las operaciones estadísticas hasta la publicación de los resultados. Esta política incluye procedimientos de control de calidad, auditorías estadísticas, y la certificación de estadísticas oficiales, con el objetivo de garantizar la precisión y la transparencia de la información difundida.

Difusión y Acceso a la Información Estadística

El DANE tiene la responsabilidad de hacer que la información estadística esté disponible y sea accesible para todos los ciudadanos. Para ello, implementa políticas de difusión que incluyen la publicación de bases de datos, informes, boletines y herramientas interactivas en su portal web. Además, promueve la transparencia y el acceso a los datos mediante plataformas de datos abiertos, asegurando que la información estadística pueda ser utilizada por investigadores, tomadores de decisiones y el público en general.

Confidencialidad Estadística

Un principio fundamental del DANE es la confidencialidad de la información recolectada. Las normas del DANE establecen que los datos individuales recopilados en las encuestas y censos no pueden ser divulgados, y sólo se publican de manera agregada. Esto garantiza la protección de la privacidad de los individuos y entidades que proporcionan la información estadística.

2.1.5 Definición de variables y construcción de indicadores estadísticos

La operación estadística implementada está basada en registros administrativos, calculando a través de conteos la producción científica de los investigadores vinculados a instituciones colombianas. Todos los indicadores se obtienen de una única base, con producto únicos que se ha construido mediante la aplicación de un algoritmo producido por el OCyT.

En este sentido, se han definido cuatro indicadores de producción científica de agregado nacional, los cuales permiten conocer la dinámica de los productos científicos generando en una ventana de observación mínima de 4 años. Este indicador puede ser presentado de manera analítica en diversas desagregaciones, tales como; tipo de repositorio, tipo de coautorías, tipo de productos y número de citaciones por productos.

Tabla 1. Lista de indicadores

Nombre del indicador	Producción bibliográfica de autores vinculados a la institución por tipo de producto
Código del indicador en la base de datos	prod_cientifica_1

Descripción	Número de documentos publicados, que dan cuenta de la generación de conocimientos a través de documentos (artículos científicos, artículos cortos, notas, editoriales, reseñas, memorias de conferencia, proceedings) en revistas de alta visibilidad internacional y que por lo tanto se encuentran indexadas en las principales bases de datos de revistas científicas (Web of Science, SciELO yScopus). Este es uno de los indicadores de resultado de las actividades de I+D más utilizados, el cual permite conocer las dinámicas en generación de conocimientos científicos en el país. Una mirada más profunda de estos documentos permite observar tenencias predominantes, las instituciones más productivas, y el impacto sobre la comunidad científica.				
Método de cálculo o fórmula	A partir de la base de datos de productos nacionales, se realiza un conteo de los documentos y artículos reportados en los productos.				
Desagregación	- Territorial				
Periodicidad	Bianual	Anual x	Semestral	Trimestral	Mensual
Cobertura geográfica	Nacional X	Departamental	Municipal	Otra Cual:	
Último año con datos disponibles	2022				
Periodo que cubre el indicador	2019 – 2022				
Observaciones:	Desde el año 2014, SciELO CI (Scielo CitationIndex) comienza a ser operativo como parte del Web of Science. Esto obedece a una intención de hacer el Web of Science competitivo en términos de cobertura regional. Hay que tener en cuenta que algunas de las revistas de SciELO CI, se encontraban incluidas dentro de Web of Science antes del 2014. Incluye además de artículos científicos, capítulos de libros artículos o abstracts en conferencia (proceedings), correcciones y fe de erratas, material editorial, cartas, reseñas de libros y capítulos de libro.				

	Las descargas se realizan desde la unidad de estudio territorial por municipio, aun así, estos datos no se presentan por entidad territorial en los productos de la operación.
Fuente	Web of Science: Core Collection, Web of Science: SciELO Citation Index y Scopus

Nombre del indicador	Producción bibliográfica de autores vinculados a la institución por tipo de coautoría				
Código del indicador en la base de datos	prod_cientifica_5				
Descripción	Este indicador mide el número de documentos construidos en cooperación en tres modalidades: i) Coautoría Internacional: Se considera un documento en esta categoría si el documento cuenta con más de un autor y por lo menos uno de los autores cuenta con una afiliación institucional del extranjero. ii) Coautoría nacional: Se considera un documento en esta categoría, si el documento cuenta con más de un autor y todos los autores están vinculados a instituciones colombianas. iii) Autoría única (Solo autor): Se considera un documento en esta categoría, si el documento cuenta con un único autor vinculado a una institución colombiana.				
Método de cálculo o fórmula	A partir de la base de datos de productos nacionales, se realiza un conteo de los documentos y artículos reportados en los productos.				
Desagregación	- Territorial				
Periodicidad	Bianual	Anual x	Semestral	Trimestral	Mensual
Cobertura geográfica	Nacional X	Departamental	Municipal	Otra Cual:	
Último año con datos disponibles	2022				
Periodo que cubre el indicador	2019 – 2022				
Observaciones:	Las descargas se realizan desde la unidad de estudio territorial por municipio, aun así, estos datos no se presentan por entidad territorial en los productos de la operación.				

Fuente	Web of Science: Core Collection, Web ofScience: SciELO CitationIndex y Scopus

Nombre del indicador	Distribución de la producción bibliográfica de autores vinculados a la institución por áreas de la ciencia
Código del indicador en la base de datos	prod_cientifica_9
Descripción	Este indicador muestra la distribución porcentual de la producción científica por áreas de la ciencia (Ciencias Multidisciplinarias y Desconocida) en tres modalidades: - Ciencias Multidisciplinarias: Se considera un documento que se clasifica en más de dos áreas de conocimiento - Producción única por áreas OCDE: Se considera un documento que se clasifica en una única área OCDE - Desconocida: Se considera un documento que se reporta en las bases con área desconocida Para su construcción el primer paso consistió en la homologación de las subáreas de Scopus y categorías de WOS a las áreas según clasificación OCDE más un área "multidisciplinaria" lo que se considera cómo áreas de la ciencia. Lo anterior permite el conteo de productos científicos únicos en cada las áreas. Se debe aclarar que los productos reportados en varias áreas se clasifican bajo el área ciencias multidisciplinarias incluyendo el área multidisciplinarias propia de Scopus.
Método de cálculo o fórmula	Paso 1: a partir de la base de datos de productos nacionales, se realiza un conteo de los productos en un área específica. Paso 2: Se contabilizan los productos reportados en las todas las áreas de conocimiento OECD únicas por producto*. **Nota: Dado que existen productos con múltiples áreas de este contará en la categoría de ciencias multidisciplinarias.* Paso 3: División entre el paso 1 y el paso 2, multiplicado por 100.
Desagregación	- Territorial - Producción única por áreas OCDE: Ciencias Agrícolas,

	Humanidades, Ciencias Naturales, Ciencias Sociales, Ingeniería y Tecnología, Ciencias médicas, Ciencias de la salud				
Periodicidad	Bianual	Anual x	Semestral	Trimestral	Mensual
Cobertura geográfica	Nacional X	Departamental	Municipal	Otra Cual:	
Último año con datos disponibles	2022				
Periodo que cubre el indicador	2019 – 2022				
Observaciones:	Las descargas se realizan desde la unidad de estudio territorial por municipio, aun así, estos datos no se presentan por entidad territorial en los productos de la operación.				
Fuente	Web ofScience: Core Collection, Web ofScience: SciELO CitationIndex y Scopus				

Nombre del indicador	Promedio de citas de documentos publicados por autores vinculados a la institución
Código del indicador en la base de datos	prod_cientifica_13
Descripción	Este indicador refleja el promedio de veces que los documentos han sido citados por los otros documentos dentro de la misma base de datos, desde el inicio de su publicación. Para los documentos más antiguos, al estar disponibles por un lapso de tiempo mayor, aumentan las posibilidades de ser citados, pero también se incrementa el riesgo de obsolescencia.
Método de cálculo o fórmula	Paso 0: Promedio aritmético de las citas alcanzas productos reportados. Paso 2: Asignación del valor máximo de citas logrado por el producto en las tres bases de datos. Paso 3: Promedio aritmético de las citas alcanzas productos reportados.

Desagregación	- Territorial				
Periodicidad	Bianual	Anual x	Semestral	Trimestral	Mensual
Cobertura geográfica	Nacional X	Departamental x	Municipal	Otra Cual:	
Último año con datos disponibles	2022				
Periodo que cubre el indicador	2019 – 2022				
Observaciones:	Las descargas se realizan desde la unidad de estudio territorial por municipio, aun así, estos datos no se presentan por entidad territorial en los productos de la operación.				
Fuente	Web of Science: Core Collection, Web of Science: SciELO Citation Index y Scopus				

Fuente: Observatorio Colombiano de Ciencia y Tecnología OCyT

La producción Científica departamental/municipal se realiza sobre Entidades gubernamentales, IES privadas, IES Publicas, Centros de investigación y desarrollo tecnológico, ONG, Asociaciones y agremiaciones profesionales, Hospitales y clínicas.

- **Detección y análisis de requerimientos**

  El líder de bibliometría reúne a su equipo de trabajo con el propósito de retroalimentar el proceso de producción científica departamental/municipal, y de realizar ajustes o cambios necesarios para optimizar la operación.

- **Diseño y Pruebas**

  Se revisa la documentación actual sobre el proceso de producción científica departamental y municipal, comenzando con la descarga de bases científicas seleccionadas. El objetivo es evaluar y optimizar los

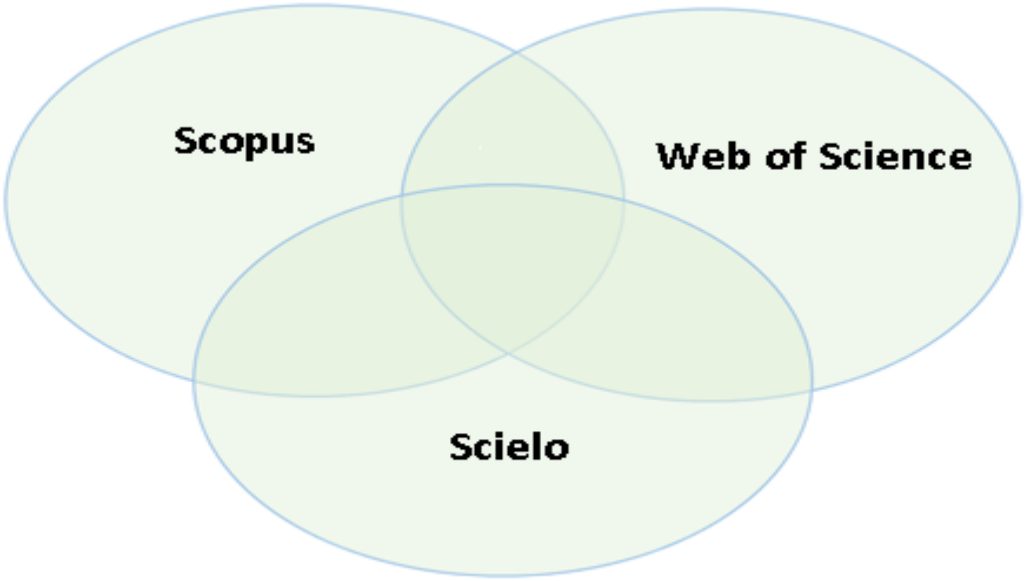
algoritmos existentes. Se realizan ajustes necesarios tanto en la documentación como en los algoritmos para garantizar el éxito del proceso.

- **Ejecución**

  Para esta operación estadística se utiliza la información sobre los documentos de autores afiliados a instituciones colombianas publicados en revistas indexadas en el Web of Science de Clarivate, en Scopus de Elsevier, y en Scielo. Su selección se basa en que estas se consideran los servicios de indexación de mayor prestigio.

  Cabe resaltar que para el ingreso a estas bases de datos requiere tener autenticación de estas.

Figura 2. Bases de datos fuente



Fuente: Observatorio Colombiano de Ciencia y Tecnología OCyT

Es importante anotar que esta operación estadística se fundamenta en algoritmos que permiten homologar los productos (documentos de

autores) presentes en las tres bases de datos para identificar productos únicos; en una ventana de tiempo de 4 años.
La búsqueda de información en estas bases científicas se realiza con delimitación de país departamento y municipio, lo cual genera archivo por cada uno de los municipios de cada departamento

El formato descarga de WoS y Scielo son documentos de texto delimitados por tabuladores cuya extensión es .txt

El formato de las descargas de Scopus incluye conjuntos de registros de documentos de texto en formato .csv.

Bases de datos de WoS y Scielo, se consolida todos los archivos a nivel país (Colombia) en una sola base de datos, pues estas bases comparten las mismas etiquetas y no se realizan procesos de homologación entre ellas, las etiquetas utilizadas son:

- TI: Título del documento
- AU: Autores del documento
- TC: Citaciones del documento
- WC: Campo de conocimiento del documento.

Base de Scopus, se consolida todos los archivos previamente de este a origen nivel país (Colombia)

Las Etiquetas Usadas son:
- Title: Titulo documento formato .csv y que se homologa a TI
- Authors: Autores de documento formato .csv y se homologa a AU
- Cite by: Citaciones de documento formato .csv y se homologa a TC

Consolidación global entre las dos bases de datos que se han generado WOS-SCIELO y SCOPUS quedando una sola base de datos integrada con productos únicos

- **Análisis**

Realización análisis no paramétricos son herramientas estadísticas que se emplean cuando los datos no cumplen con los supuestos de normalidad o cuando la escala de medición departamental presenta anomalías de ser así se aborda hasta encontrar posibles fallas o errores en el proceso.

Además, se hace revisión de información histórica lo que implica examinar detenidamente los antecedentes y el contexto histórico relacionado con la producción científica departamental. Chequeo de la revisión de estudios previos, datos históricos relevantes, documentos y literatura pertinente que ayuden a contextualizar los hallazgos y a comprender mejor la evolución de la producción científica departamental a lo largo del tiempo.

Una vez completados los análisis y revisión histórica, se procede a la generación de Unos resultados finales. Esto implica la presentación y síntesis de los hallazgos obtenidos a partir de los análisis realizados, proporcionando conclusiones significativas y recomendaciones basadas en la evidencia recopilada. Los resultados finales son fundamentales para validar las hipótesis de investigación, identificar tendencias, establecer relaciones entre variables y contribuir al conocimiento en el área de la producción científica departamental.

2.1.6 Plan de resultados

La publicación de los resultados de las estadísticas de producción científica bibliométrica sigue un cronograma de difusión establecido anualmente. En los primeros 10 días de abril del año siguiente al año de los resultados se actualiza la información de la operación estadística en la página web https://www.ocyt.org.co/

Por la naturaleza de las operaciones de la producción bibliográfica, los datos obtenidos provienen de bases dinámicas y en el momento en que se realiza las consultas, estas pueden variar respectos a años anteriores, es por esto por lo que la serie se actualiza en una ventana de observación de 4 años.

A continuación, se presenta el diseño del plan de resultados,

a. Cuadros de salida o resultados: Son producidos con la base de datos única, validada y depurada. Estos se ponen a disposición del público en el sitio web del OCyT.

b. Herramienta Interactiva: Con el fin de centralizar la información relacionada con la operación estadística, se cuenta con un micrositio de bibliometría que reúne:

- Indicadores temáticos
- Documentos metodológicos
- Mecanismos de soporte
- Cronograma

Enlace de acceso: https://bibliometria.ocyt.org.co/

c. Documentos de interpretación de resultados: Puede ser un capítulo, informe o boletín, la presentación está sujeta a decisión del líder de bibliometría, de la dirección del OCyT y del comité editorial.

2.1.7 Diseño de cuadros de salida o de resultados

Los cuadros de salida fueron diseñados a partir del análisis de las necesidades de los usuarios con formato del Observatorio y de acuerdo con las tendencias internacionales de presentación del conjunto de variables sobre producción científica bibliométrica[3]

2.2 Diseño del formulario o cuestionario

La operación estadística está basada en registros administrativos, que, para el caso de las estadísticas de Producción Científica Bibliométrica, la información se obtiene a partir de las bases de datos que se descargan de las páginas WEB of Science, Scopus, Scielo, insumo con el cual se producen las estadísticas de Ciencia y Tecnología.

[3] Revisar la ficha metodológica

2.3 Normas, especificaciones o reglas de validación, consistencia e imputación

Como primera medida, se establece que las descargas se realizan por municipio, siendo un enfoque territorial más local que garantiza la cobertura nacional para el cálculo de producción científica de bibliometría.

Una vez se realiza la descarga se debe diligenciar el formato de verificación de descargas, este formato contiene los aspectos para tener en cuenta como: medidas de validación y la completitud de las descargas. Este formato de verificación está conformado por dos tablas de seguimiento.

La primera tabla reúne las variables que se deben validar para las descargas de la fuente de Scopus, la segunda tabla reúne los aspectos de verificación de descargas de la fuente Web Of Science y Scielo. Esto se debe a que WoS contiene colecciones de citas de Scielo e integran las mismas revistas indexadas, por lo tanto, se evalúan bajo los mismos criterios establecidos.

Por otra parte, y en el sentido de compromiso de la operación se ha desarrollado un sistema de limpieza por medio de scripts en Python que permiten, depurar los datos descargados, realizar validaciones de los datos y encontrar casos atípicos en los datos suministrados por las fuentes.

Por el carácter de la información de esta operación estadística, no se lleva a cabo imputación de datos, ni estimaciones, dado que los cálculos se realizan con base a registros administrativos, para dar a conocer el avance nacional en la producción científica[4]. **Anexo 2**

2.4 Nomenclaturas y clasificaciones utilizadas

DIVIPOLA: La operación estadística utiliza la División Político-Administrativa de Colombia – DIVIPOLA de fuente del Departamento Administrativo Nacional de Estadísticas-DANE, que está definido como "un estándar de codificación que

[4] Revisar el formato de verificación de descargas, el instructivo de recolección y de procesamiento.

permite contar con un listado organizado y actualizado de la totalidad de unidades en que está dividido el territorio nacional"[5]

Áreas de conocimiento -OCDE: Los Campos de Investigación y Desarrollo (FORD) constituyen un esquema de distribución del conocimiento, propuesto por la Organización para la Cooperación y el Desarrollo Económicos (OCDE), para clasificar las unidades de ejecución de investigación y desarrollo experimental (I+D) y distribuir sus recursos en función del ámbito de conocimiento en el que se lleva a cabo.[6]https://concytec-pe.github.io/vocabularios/ocde_ford.html

3 DISEÑO ESTADÍSTICO

3.1 Universo de estudio

Producción científica bibliográfica de los investigadores adscritos a instituciones colombianas

3.2 Población Objetivo

Producción científica bibliográfica de los investigadores adscritos a instituciones colombianas que registran sus productos en las bases de datos de Scopus, WoS, Scielo.

3.3 Cobertura geográfica: Nacional

3.4 **Desagregación geográfica:** Se presenta a nivel **nacional**

3.5 **Desagregación temática:** Tipo de producto, tipo de autoría, área OCDE, promedio de citas

[5]Definición tomada del Geoportal DANE

[6](Fuente: OCDE (2015), Frascati Manual 2015: GuidelinesforCollecting and Reporting Data onResearch and Experimental Development, TheMeasurement of Scientific, Technological and InnovationActivities. Publicado por acuerdo con la OCDE, París (Francia). DOI: http://dx.doi.org/10.1787/9789264239012-en)

3.6 Fuente de datos (tipo de operación)

Bases de datos de producción Scopus, WoS, Scielo, con las cuales se consolida la operación estadística a partir de registros administrativos

3.7 Comprobación de datos fuentes secundarias

Para un adecuado uso de los datos provenientes de fuentes secundarias se hace un proceso de revisión de la información que proviene de estas fuentes para la consolidación.

Tabla 2. Comprobación de datos fuentes secundarias

Aspectos evaluados	Scopus	WoS	Scielo
La utilidad de las variables de interés y los conceptos	Se utilizan variables seleccionadas por ser de interés	Se utilizan variables seleccionadas por ser de interés	Se utilizan variables seleccionadas por ser de interés
El nivel de desagregación geográfica del archivo de datos	Se desagrega por entidad territorial	Se desagrega por entidad territorial	Se desagrega por entidad territorial
La relevancia de los datos contenidos en los archivos de datos	Muy relevantes	Muy relevantes	Relevantes
Que no existe duplicidad de datos	Muy relevante	Muy relevante	Muy relevante
La transmisión oportuna de las variables de interés para la producción estadística, de acuerdo con la	Oportuna	Oportuna	Oportuna

periodicidad de envío			
Las capacidades tecnológicas para transmitir, cargar, almacenar y procesar el archivo de datos	Se cuenta con dichas capacidades	Se cuenta con dichas capacidades	Se cuenta con dichas capacidades
La confidencialidad, la disponibilidad y la protección de los datos	Disponibles	Disponibles	Disponibles
El acceso continuo a los archivos de datos para la producción estadística	Datos disponibles, pero se necesita credenciales de acceso	Datos disponibles, pero se necesita credenciales de acceso	Datos disponibles

Fuente: Observatorio Colombiano de Ciencia y Tecnología OCyT

3.8 Unidades estadísticas

- **Unidad de observación:** Productos científicos bibliográficos nacionales y producción bibliográfica por autores.
- **Unidad de muestreo:** Productos científicos bibliográficos nacionales y la producción bibliográfica por autores.
- **Unidad de análisis:** Documentos construidos en cooperación por tipo de coautoría, número de documentos publicados y la distribución porcentual de la producción científica por área de conocimiento.

3.9 Período de referencia: Ventana de observación de 4 años, partiendo desde el año respectivamente anterior, es decir, si la operación estadística se ejecuta en el año 2023, la ventana de observación será 2019 a 2022.

3.10 Período de recolección/ acopio: dos meses finales del año de referencia.

4 DISEÑO DE LA RECOLECCIÓN/ACOPIO

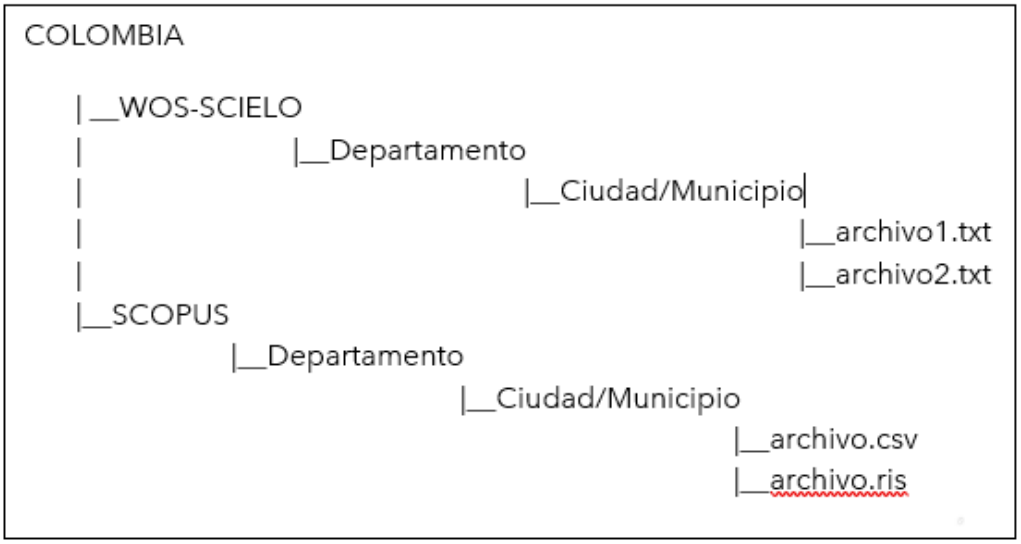
4.8 Métodos y estrategias de recolección o acopio de datos

Para la operación estadística denominada "Estadística de la Producción Científica Bibliométrica" elaborado por el Observatorio Colombiano de Ciencia y Tecnología – OCyT, se utiliza información sobre los documentos de autores afiliados a instituciones colombianas publicados en revistas indexadas en Web of Science de Clarivate, en Scopus de Elsevier y SciELO.

La búsqueda de información en estas bases científicas se realiza con delimitación de país y municipio, lo cual genera un archivo por cada uno de los municipios de dicho departamento. Esta se fundamenta en un algoritmo que permite homologar los productos (documentos de autores) presentes en las tres primeras bases de datos para identificar productos únicos, en una ventana de tiempo de 4 años. *Véase en el instructivo de recolección.*

La recolección de información se organiza en la carpeta procesar la cual utilizará el aplicativo para ejecutarse. Internamente dicha carpeta posee subcarpetas con el nombre estándar para los departamentos que contienen los resultados de las descargas, es decir los archivos .csv, .ris y .txt la parametrización se presenta en el siguiente esquema.

Figura 3. Parametrización de búsqueda en WoS



Fuente: Observatorio Colombiano de Ciencia y Tecnología OCyT

Hay que tener en cuenta que se debe tener credenciales para poder acceder estas bases de datos científicas.

4.9 Estructura organizacional del operativo y definición del equipo requerido

El personal involucrado en la operación estadística es seleccionado de acuerdo con los perfiles técnicos definidos por el OCyT que son verificados por la coordinación administrativa y financiera al momento de su incorporación, acorde con las necesidades del cargo y con las competencias que se requieren para las actividades involucradas en el desarrollo de la operación estadística. A continuación, se muestra las actividades y funciones del equipo operativo.

Tabla 3. Roles y funciones del equipo operativo

Rol en el operativo	Funciones
Líder de bibliometría, coordinador operacional o analítico	Identificar las necesidades de información nuevas para la operación estadística. Identificar las necesidades de fuentes de información y bases de datos que apoyen la operación estadística. Generar el cronograma de la operación estadística Elaborar informes de evaluación por fase. Supervisar el proceso de acompañamiento en la recolección de los datos de las fuentes definidas Generar los cuadros de salida, de acuerdo con lo diseñado, para los diferentes espacios de difusión. Generar reportes de información por solicitud de los usuarios. Realizar seguimiento al equipo de la operación y coordinar sus actividades con el equipo de apoyo transversal.
Investigador	Apoyo en el procesamiento de los datos de las fuentes definidas para la generación de los cuadros de salida e indicadores de la operación estadística Desarrollo de software para el soporte de los procedimientos de recolección y/o procesamientos señalados por el Coordinador operativo

Asistente de investigación	Recolectar y acopiar datos de las fuentes definidas Realizar informes periódicos de avance de la operación. Acompañamiento al operativo de recolección
Líder de inversión, apoyo transversal	Dar lineamientos, directrices y políticas para implementar la gestión de la calidad estadística en el OCyT y su mejora continua. Liderar la centralización y difusión oficial de la información estadística que produce y administra el OCyT. Revisar el diseño de las operaciones estadísticas. Tomar decisiones estratégicas, de lineamientos para la gestión de la información estadística del OCyT. Aprobar la información estadística, antes de su difusión.
Gestor de la información, apoyo transversal	Especificar, diseñar, implementar, disponer y gestionar la infraestructura tecnológica requerida que permita el cargue y procesamiento de los datos producidos, para satisfacer las necesidades de información estadística. Liderar la estrategia tecnológica para facilitar y optimizar la gestión de la información. Liderar la automatización del procedimiento de gestión de la información estadística del OCyT en todas sus fases y la disposición para su difusión Velar por la seguridad informática y la protección de datos acorde con la normatividad y políticas internas vigentes. Implementar nuevas tecnologías para la construcción y publicación de tableros de indicadores amigables, usables y de fácil visualización. Gestionar el trámite a las peticiones, quejas y reclamos que se reciban por los diferentes canales definidos por el OCyT. Evaluar la satisfacción de usuario frente a la calidad del servicio prestado por el Observatorio
Estadístico, apoyo transversal	Recopilar e interpretar los datos del OCyT para descubrir patrones y tendencias que permitan el análisis e interpretación de la información de CteI.
Asistente de estadística, apoyo transversal	Apoyar transversalmente el análisis estadístico en diferentes áreas del OCyT.

Asistente de comunicaciones, apoyo transversal	Apoyar en la implementación de la política de comunicación interna y externa para una adecuada difusión de la información estadística a los públicos identificados.
Líder de gestión humana, apoyo administrativo	Brindar el soporte y apoyo administrativo y proveer el recurso humano competente, requerido para la gestión estadística. Liderar las políticas para la gestión documental del OCyT, incluyendo la conservación, custodia y transferencia de los archivos documentales generados de la operación estadística.
Subdirectora administrativa y financiera, apoyo administrativo	Proceso convocatorio, selección y contratación de personal, programación pagos, reportes financieros, tributarios, apoyo al SGC

Fuente: Observatorio Colombiano de Ciencia y Tecnología OCyT

4.10 Esquema de entrenamiento del personal

De acuerdo con los lineamientos del OCyT, se realiza un entrenamiento al equipo de Bibliometría en donde se imparten los conocimientos y las habilidades necesarias para el desarrollo de la operación estadística y las actividades involucradas en el procesamiento de la información. *Véase en el plan de entrenamiento.*

4.11 Proceso de sensibilización y acuerdos de intercambio

Por las características de la forma de recolección para las estadísticas de producción científica, la cual se realiza a partir de la descarga de información de WEB of Science, Scopus y Scielo, no se realiza actividades de sensibilización a las fuentes. Sin embargo, se hace una revisión de los cambios que pueda presentar la estructura de las bases de datos.

4.12 Elaboración de instructivos

Para el desarrollo del proceso de recolección de información se cuenta con guías e instructivos que permiten desarrollar el proceso de descarga de la información, homologación y depuración de las bases de datos de WoS, Scielo y Scopus.

A continuación, vamos a mencionar los instrumentos, con los cuales cuenta el proceso de la recolección de la información:

- Instructivo de recolección de la información
- Instructivo de procesamiento

4.13 Diseño de la estrategia de comunicación y plan de contingencias

Para asegurar que se cumpla la recolección de información de manera correcta, todos los inconvenientes, obstáculos o contratiempos durante dicha actividad se comunicarán al cargo inmediatamente anterior o superior directo para ser resuelto. *Véase en el manual de funciones y organigrama.*

4.14 Diseño de la estrategia de seguimiento y control

Con el objetivo de supervisar y controlar la operación para la recolección y acopio de la información desde las bases de datos se establece la entrega de informes de avance e informe final de recolección a los asistentes de investigación durante el operativo de recolección. Lo anterior para crear y alimentar los indicadores de gestión definidos en el área.

4.15 Diseño de sistemas para la obtención de datos

En la siguiente gráfica se presenta el proceso de captura y consolidación de la base única a partir de las bases de datos de revistas científicas (Web of Science Core Collection, SciELO y Scopus).

Para el proceso de recolección de la información, el OCyT con base en la revisión temática que ha realizado sobre la obtención de los datos, ha identificado que las fuentes idóneas de información donde se puede establecer cómo se encuentra la producción de información científica bibliométrica de diferentes autores por país e institución, son las bases de datos de WoS, Scielo y Scopus.

En este sentido se ha establecido la descarga de información de las páginas de producción bibliométrica antes mencionadas en los dos meses finales del año de referencia. Para llevar este proceso, se ha definido la homologación, estandarización entre las diferentes fuentes y codificación de la información para

garantizar un producto consistente frente al usuario de los resultados. Lo anterior, se realiza de acuerdo con el detalle que se encuentra en la guía de recolección de la información.

Se implementan mecanismos de seguimiento continuo para asegurar el cumplimiento de las metas establecidas en el cronograma de recolección. Estos incluyen:

Monitoreo semanal: Seguimiento de la cantidad de registros extraídos y su cobertura geográfica, se documentará en el informe de avance y final de la fase de recolección.

Indicadores de Rendimiento: Evaluación del rendimiento del equipo en términos de cantidad y calidad de los datos recolectados.

Almacenamiento Seguro: Bases de datos locales que almacenan los registros bibliográficos de manera segura, asegurando la protección de la información.

4.16 Transmisión de datos

Se realiza una preparación de los archivos, la cual consiste en obtener de las bases de datos Web of Science, Scielo y Scopus archivos en .txt, .csv y .ris respectivamente para el posterior procesamiento. Esto permite realizar la consolidación de los archivos y generar una base consolidada.

Teniendo en cuenta los archivos esenciales contenidos en la carpeta *base* del aplicativo para el procesamiento (para mayor detalle *ver Instructivo de procesamiento*), se estandariza la información de acuerdo con los criterios que se han definido por el OCyT.

5 DISEÑO DEL PROCESAMIENTO

5.1 Consolidación de archivos de datos

A continuación, se describen de forma general los procedimientos para el procesamiento de la información necesaria para el cálculo de los indicadores de la operación estadística de bibliometría. Estos procedimientos determinan la construcción de los cuadros de salida del operativo. Para mayor detalle del

procedimiento para el procesamiento se puede consultar el *Instructivo de procesamiento*.
La metodología actual para la unificación de la información exportada de las fuentes se fundamenta en una estructura jerárquica integrada en los aplicativos scopus y wos-scielo que a partir de los archivos .ris, .csv y .txt obtenidos de las bases de datos generan los resultados para el aplicativo global-sco-wos, así:

```
procesar
    |__Departamento
            |__Ciudad/Municipio
                    |__archivo1.txt
                    |__archivo2.txt
                    |__archivo.csv
                    |__archivo.ris
```

En este sentido, con la anterior estructura se ejecutan los dos aplicativos (Scopus y wos-scielo) que generan internamente una base de acuerdo con los años de interés de la data a procesar la cual alimenta el aplicativo que produce los resultados globales.
Este aplicativo se encarga de la fusión de documentos generado por la aplicación scopus y wos-scielo que están dentro de la capeta procesar_ resultados\produccion\municipal.

5.2 Revisión y validación

Una vez recolectada la información a partir de las descargas de las bases de datos de las páginas Web of Science, Scopus y Scielo, se realiza el análisis donde se contrasta la información preliminar obtenida en el año de referencia con las cifras históricas. Esta actividad permite identificar rápidamente cambios atípicos entre uno y otro proceso y posibles inconsistencias, a partir de esta actividad se obtienen una base de datos consistente y depurada.

En el proceso de recolección de la información se realizan una serie de controles los cuales se desarrollan de acuerdo con lo establecido para la operación estadística. Es así, que se realiza monitoreo a situaciones en las cuales se puedan

generar algunas afectaciones al dato y se establecen las acciones a realizar para poder contar con información de calidad.

Posteriormente una vez se cuenta con la base de datos, se realiza una depuración validando los datos descartados, en esta depuración se descartan los espacios vacíos y con base a la DIVIPOLA se realiza la revisión de los códigos territoriales para asegurar la calidad de la información y que en el cálculo sean procesados los campos validados[7]

5.3 Diseño para la generación de cuadros de resultados

Una vez se cuenta con la información validada y la base de datos normalizada de cada uno de los diferentes repositorios, se implementa el algoritmo se da inicio al proceso de generación de los indicadores y cuadros de salida producto de las estadísticas de Producción Científica Bibliométrica, los cuales se realizan por medio de los aplicativos y su ejecución detallados en el instructivo de procesamiento.

5.4 Seguridad y almacenamiento de la información

Las bases de datos en ambiente productivo del Observatorio Colombiano de Ciencia y Tecnología se almacenan en un servidor en la nube tipo hosting, el cual se ubica en un datacenter con certificación TIER III / SOC 2 tipo II, lo que garantiza un 99.982% de disponibilidad al año, seguridad basada en controles de acceso y protección contra ataques DDoS, la arquitectura del servidor cuenta con una plataforma bajo Linux a 64bits, motores de bases de datos MySQL, con un canal de acceso de 1GB y un disco de almacenamiento 100GB + 50GB MYSQL, administración gráfica vía CPanel y finalmente cuenta sistemas de BackUP automático y manual.

A nivel LAN en el DataCenter del OCyT se cuenta con un servidor dedicado de archivos donde se centraliza el sistema de BackUP de la información y bases de datos para procesar, el cual, utilizando herramientas de BackUP automatizadas se realizan copias de tipo completas, diferenciales e incrementales dependiendo de

[7] Revisar el instructivo de procesamiento.

la criticidad de la información, estas copias se almacenan de manera local y una segunda copia se carga a la nube en los periodos establecidos, esto para garantizar el proceso de restauración y minimizando el riesgo de pérdida de información.

Finalmente, el servidor de archivos y los demás servidores ubicados en el DataCenter y en la red LAN del OCyT cuentan con una protección adicional perimetral de tipo Firewall UTM para la gestión unificada de amenazas, este dispositivo incluye una amplia variedad de funciones, como firewall de inspección con estado, antivirus HTTPs / FTP, filtro de contenido, antivirus POP3 / SMTP, antiphishing y herramientas de contra el spam, SSL / TLS VPN, IDS y otras características de seguridad.

5.5 Diseño del plan de pruebas

Hasta el momento el desarrollo de pruebas piloto asociadas a cambios significativos en la metodología, la toma de información, los instrumentos o la herramienta informática de esta operación estadística no han sido necesarios, puesto que ésta no ha sufrido cambios en los aspectos antes señalados que justifiquen la aplicación de dichas pruebas o revisiones. No obstante, lo anterior se cuenta con una guía de revisiones y pruebas a metodología, instrumentos y herramientas de las estadísticas de producción científica bibliométrica, para que sea tenida en cuenta por el equipo técnico y la aplique en los casos que sean necesarios.[8]

6. DISEÑO DEL ANÁLISIS

5.6 Métodos de análisis de resultados

En este diseño de análisis se detalla una metodología paso a paso para llevar a cabo el análisis de las estadísticas de producción científica en Colombia. El objetivo es proporcionar una guía clara y estructurada que permita a los investigadores y analistas realizar evaluaciones exhaustivas y significativas.

[8] Revisar el plan de pruebas e informe de pruebas

Definición de Variables

Identificar las variables clave que serán analizadas para evaluar la coherencia de la producción científica, como la cantidad de publicaciones por año, áreas temáticas, colaboraciones, citaciones, entre otras.

Análisis de Coherencia

Análisis de tendencias temporales. Aplicar técnicas estadísticas para identificar tendencias a lo largo del tiempo en variables clave como:

a. El número de publicaciones por año
b. Número de colaboraciones
c. Número de citaciones

Detección de valores atípicos. Utilizar métodos estadísticos para identificar y analizar valores atípicos que puedan indicar anomalías o errores en los datos recopilados.

Los métodos contemplados y aplicables son:

Distribuciones estadísticas específicas. Dependiendo de la naturaleza de los datos (por ejemplo, si siguen una distribución normal o no), se pueden utilizar métodos específicos como la prueba de Grubb para datos normales o métodos basados en distribuciones t de Student, estos métodos son sugerencias, los métodos de distribución están sujetos al criterio del analista o el responsable de la interpretación de datos.

Consistencia entre variables. Realizar análisis bivariados para verificar la consistencia entre diferentes variables, por ejemplo, la relación entre el número de publicaciones y las citaciones recibidas.

Análisis Descriptivo

Se estudiarán diversas variables, como el número de artículos publicados, tipos de documentos (artículos, conferencias, editoriales), citas recibidas, y colaboraciones nacionales e internacionales.

Se aplicarán técnicas invariadas (análisis de frecuencias, medidas de tendencia central y dispersión), para describir la distribución y características de las variables.

Análisis Bivariado y Multivariado

Se explorarán las relaciones entre variables utilizando análisis bivariados (correlaciones entre el número de citas y la institución de afiliación, por ejemplo) y multivariados (análisis de regresión para determinar factores predictivos del impacto científico).

Se emplearán técnicas estadísticas como regresiones lineales y análisis de varianza (ANOVA) para examinar las relaciones entre variables múltiples.

Análisis Espacial (Cienciometría Espacial):

Se georreferenciarán las publicaciones científicas para evaluar la distribución territorial de la producción científica en Colombia.

5.7 Anonimización

Identificación de variables sensibles. Identificar las variables que podrían revelar la identidad de las instituciones educativas u otras entidades involucradas en la producción científica, como nombres específicos, ubicaciones detalladas, o datos directamente identificables.

Selección de técnicas de anonimización. Elegir técnicas adecuadas basadas en el tipo de variables que deben ser anonimizadas:

- Supresión: Eliminación de variables directamente identificables o sensibles.

- Pseudonimización: Reemplazo de datos identificables por pseudónimos o códigos que no permitan la identificación directa.

- Generalización y/o Agregación: Reducción de la precisión de los datos (por ejemplo, agrupando datos geográficos en niveles más generales).

5.8 Verificación de anonimización

Implementación en el proceso de preparación de datos: Integrar las técnicas de anonimización en el flujo de trabajo de preparación de datos antes del análisis estadístico.

Examinar los cuadros de resultados y tablas estadísticas para verificar que los valores presentados no puedan ser atribuidos a una fuente específica.

Confirmar, que las estadísticas descriptivas, medidas de tendencia central, dispersión y otros indicadores no revelen detalles que comprometan la privacidad de las fuentes.

7. DISEÑO DE LA DIFUSIÓN Y COMUNICACIÓN

El OCyT anualmente prepara el plan de divulgación en el cual se presenta los medios sobre los cuales se van a difundir los resultados de las Estadísticas de Producción Científica Bibliométrica en cuanto a las tipologías de entrega de información, los productos estadísticos para la difusión de resultados y las actividades de difusión de productos informativos**.**

Los resultados de la operación estadística se difunden mediante el desarrollo de cuadros de salida y gráficas de una forma clara, precisa y de fácil interpretación para el lector, donde dichos resultados evidencien por si solos el panorama de la producción científica en Colombia. El principal medio de difusión de los microdatos y metadatos utilizado por el OCyT es la página web https://www.ocyt.org.co/

7.1 Diseño de productos de difusión y comunicación

La estructura del plan de difusión y comunicación está bajo el racional estratégico CONCIENCIA DE DATOS, donde pretenden de manera integral atender los

propósitos de las premisas estratégicas identificadas, recogen las mejores prácticas en comunicación y definen acciones específicas orientadas a aumentar la percepción positiva del OCyT en las audiencias clave identificadas. Este plan de acción en 360° mantiene la coherencia que se define en el racional estratégico, en cada una de sus líneas de alistamiento, percepciones, posiciones, acciones y relaciones.

7.2 Entrega de productos

El Observatorio de Ciencia y Tecnología informa la fecha de publicación de resultados en la página web.

La información publicada incluye diferentes tipos de documentos: metadatos (Ficha metodológica, glosario y metodología) y las series históricas desde los últimos 10 años.

El OCyT ha dispuesto de un tablero de control online denominado portal de datos en el cual el usuario pueda hacer consultas.

7.3 Estrategia de servicio

Según los lineamientos establecidos por el OCyT, el área de comunicaciones asigna las PQRS (Peticiones, Quejas, Reclamos y Sugerencias) emitidas por los usuarios a través de la sección "Contáctenos" en la página web, al área de bibliometría para su correspondiente gestión y tratamiento. *Véase en el plan de difusión*

8. DISEÑO DE LA EVALUACIÓN DE LAS FASES DEL PROCESO

El equipo técnico del OCyT cuenta con auto evaluaciones en periodos donde se observa que es susceptible de mejorar algún aspecto de las fases del proceso estadístico.

Así mismo, se ha establecido la implementación de un sistema de gestión en el cual se ha definido un programa de auditorías internas que permitirán verificar la eficacia de los procesos.

Criterios de Evaluación por Fase

1. Fase de recolección de datos

1.1 Compleción de la recolección

- La recolección de datos se debe completar dentro del plazo establecido.
- Todos los datos deben ser recolectados desde las fuentes especificadas (Web of Science, Scopus, SciELO) y deben cubrir todas las unidades geográficas relevantes.

1.2 Validación de datos

- Se deben aplicar reglas de validación y consistencia a los datos recolectados para garantizar su calidad.
- Las inconsistencias o anomalías identificadas en los datos deben ser corregidas de manera oportuna.

1.3 Cobertura geográfica:

- Los datos recolectados deben cubrir todas las áreas geográficas especificadas en el diseño de la operación.

2. Fase de Procesamiento de Datos

2.1 Consolidación de archivos

- La consolidación de archivos debe integrar de manera adecuada todas las fuentes de datos en una única base de datos.
- Los procedimientos de consolidación deben seguir las directrices establecidas en el diseño metodológico.

2.2 Codificación y homologación:

- La codificación de la información debe realizarse de acuerdo con las nomenclaturas y clasificaciones (DIVIPOLA, áreas de conocimiento OCDE) establecidas.
- El algoritmo de homologación de productos únicos debe aplicarse correctamente para evitar duplicaciones.

2.3 Depuración y limpieza:

- El sistema de limpieza de datos debe eliminar registros duplicados o inconsistentes.
- Los casos atípicos o errores identificados deben ser gestionados adecuadamente.

3. Fase de análisis de resultados

3.1 Métodos de análisis

- Los métodos de análisis aplicados deben ser apropiados para los datos recolectados y alinearse con los objetivos del estudio.
- Se deben realizar revisiones históricas y análisis comparativos, según corresponda.

3.2 Interpretación de resultados

- La interpretación de los resultados debe ser coherente, clara y proporcionar conclusiones y recomendaciones útiles.
- La coherencia interna de los resultados debe ser verificada en relación con estudios anteriores.

4. Fase de difusión y comunicación

4.1 Difusión de productos

- Los productos deben publicarse según el cronograma de difusión establecido.

- Los productos (informes, cuadros, gráficos) deben cumplir con los estándares de presentación establecidos.

4.2 Accesibilidad de la Información

- La información debe estar disponible y ser accesible para todos los usuarios interesados.
- Se deben utilizar canales de comunicación efectivos para asegurar la difusión de los resultados.

5. Fase de evaluaciones

5.1 Autoevaluaciones y auditorías

- Se deben realizar las autoevaluaciones y auditorías internas planificadas para evaluar la eficacia de los procesos.
- Las recomendaciones derivadas de las evaluaciones deben implementarse correctamente y documentarse.

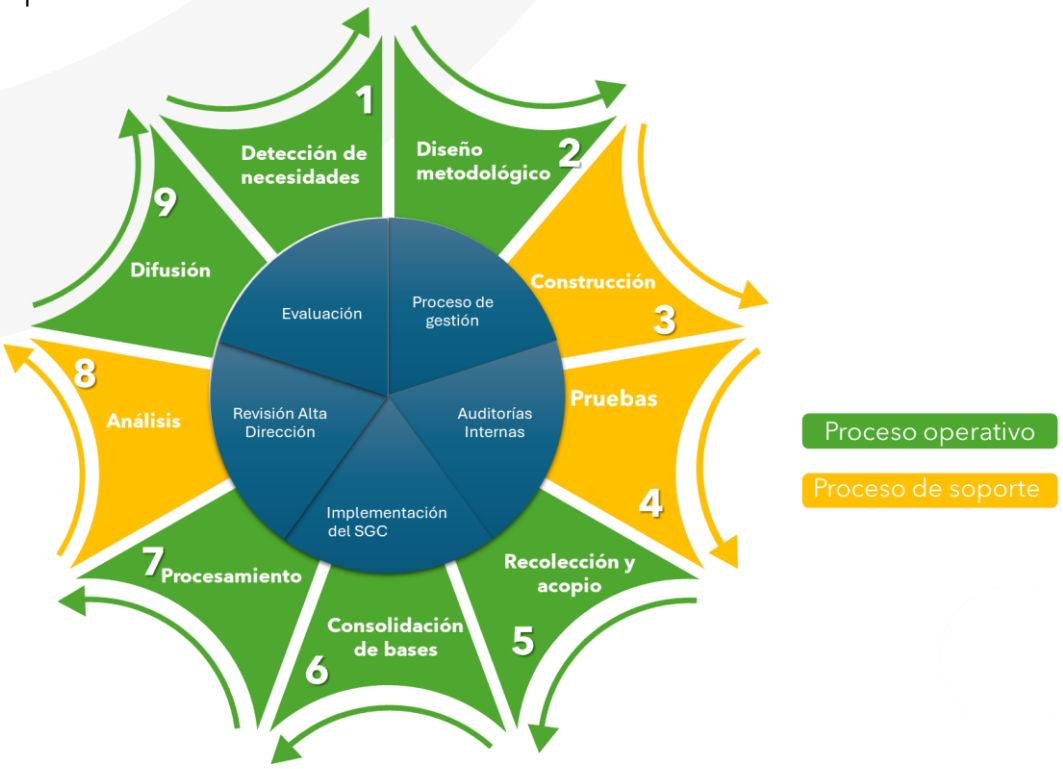
5.2 Plan de acción

- Se debe desarrollar un plan de acción basado en los hallazgos de la evaluación para mejorar los procesos futuros.
- El progreso de las acciones de mejora identificadas debe ser seguido y documentado.

8.1 Mapa de procesos

El mapa de procesos proporciona una representación visual de las principales etapas involucradas en la operación estadística, desde la detección de necesidades hasta la difusión de resultados. Este mapa organiza los procesos en categorías clave: operativos, de soporte y de gestión. Los procesos operativos incluyen el diseño metodológico, recolección de datos, procesamiento y análisis, mientras que los procesos de soporte y gestión aseguran la calidad y supervisión

continua a través de auditorías internas y revisiones por la alta dirección. En conjunto, este mapa ofrece una visión integral del flujo de trabajo, destacando cómo se interrelacionan las diferentes fases para cumplir con los objetivos de la operación estadística.



Fuente. OCyT

9. DOCUMENTACIÓN RELACIONADA

Para el desarrollo de esta operación estadística se diseñaron los siguientes documentos:

Ficha Metodológica: presenta un resumen de los aspectos metodológicos de la operación estadística.

Librerías de Phyton: paquetes con funcionalidades de cálculo y graficación que facilitan el procesamiento de información.

Instructivo de recolección de la información: características de cómo se lleva a cabo la recolección o captura de información.

Instructivo de procesamiento: características de cómo se desarrolla el procesamiento de la información y la generación de resultados.

Bases de homologación: Web of Science y Scielo usan la misma base de homologación. Es decir que se relacionan dos hojas en el documento, área para Scopus y áreas para WoS y Scielo.

Informe de pruebas: Documento que presenta los resultados de las pruebas realizadas.

Plan de pruebas: Documento que define el enfoque, los recursos, y el cronograma para las actividades de prueba de un sistema o proceso.

10. BIBLIOGRAFÍA

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA, DANE http://www.dane.gov.co/files/sen/normatividad/Lineamientos_Proceso_Estadistico.pdf, Lineamientos para el proceso estadístico en el Sistema Estadístico Nacional.

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA, DANE https://www.dane.gov.co/files/sen/lineamientos/DSO_020_LIN_04.pdf, Lineamientos para el proceso estadístico en el Sistema Estadístico Nacional.

Arango-Botero, D., Sánchez, L. E., Salazar, M. V, Valencia-Arias, A., & Salazar, C. V. (2019). Research approaches to technology transfer in the triple helix context: A bibliometric analysis. *Knowledge Management*, *19*(1), 1–20. https://doi.org/10.18848/2327-7998/CGP/v19i01/1-20

Chubin, D., & Garfield, E. (1980). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *2*(1), 91–94. https://doi.org/10.1007/BF02016602

Ferreira, M. L. A., & Ramos, R. R. (2015). Making university-industry technological partnerships work: A case study in the Brazilian oil innovation system. *Journal of Technology Management and Innovation*, *10*(1), 173–187. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84928228578&partnerID=40&md5=af6d83aa97771eb157b2cd1cd9d77919

Fishman, G., Alperin, J. P., Cetto, A. M., Alonso-Gamboa, J. O., Packer, A., Aguado-López, E., Bongiovani, P., Gómez, N., de Oliveira Amorim, K. M., Degani-Carneiro, F., & others. (2015). *Hecho en Latinoamérica: acceso abierto, revistas académicas e innovaciones regionales*. CLACSO.

Garfield, E. (1972). Bibliometric Analysis as a Tool in Journal Evaluation. *Science*, 471–479. https://doi.org/DOI: 10.1126/science.178.4060.471

Larivière, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. *Springer Handbooks*, *2018*, 3–24.

https://doi.org/10.1007/978-3-030-02511-3_1

Millar, C. C. J. M., Lockett, M., & Mahon, J. F. (2016). Knowledge intensive organisations: on the frontiers of knowledge management. *JOURNAL OF KNOWLEDGE MANAGEMENT*, *20*(5), 845–857. https://doi.org/10.1108/JKM-07-2016-0296

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(SUPPL. 1), 5200–5205. https://doi.org/10.1073/pnas.0307545100

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2017). Estudios Económicos de la OCDE: México. *Enero*, *Visión general*, 53. https://doi.org/10.1787/9789264227682-es

Pritchard, A. (1969). Statistical bibliogrphy or bibliometrics? *Journal of Documentation*, *25*((4)), 348–349.

Salazar, M., Lozano-Borda, M., & Lucio-Arias, D. (2014). Chapter 5: Science, technology and innovation for inclusive development in Colombia: pilot programmes developed by Colciencias. In *National Innovation Systems, Social Inclusion and Development* (pp. 133–168). https://doi.org/10.4337/9781782548683.00010

Uribe-Tirado, A., & Alhuay-Quispe, J. (2017). Metric study of information literacy in Latin America: From bibliometrics to altmetrics . *Revista Espanola de Documentacion Cientifica*, *40*(3). https://doi.org/10.3989/redc.2017.3.1414

Vargas-Quesada, B., & de Moya Aragón, F. (2007). *Visualizing the structure of science* (Springer (ed.)).

Zárate Rueda, R., Becerra Ardila, L. E., & Fuentes Suárez, F. L. (2017). State of scientific production of Social Responsibility in Colombia | Estado de la producción científica de la Responsabilidad Social en Colombia. *Espacios*, *38*(53).

11. CONTROL DE CAMBIOS

FECHA	DESCRIPCIÓN DEL CAMBIO	VERSIÓN
13/10/2018	Lanzamiento del documento	1
08/01/2020	Se ajusta el documento con los lineamientos DANE para la NTCPE 1000:2017, se añaden medidas de calidad, pruebas no paramétricas con el rango de muestra, formalización de listas de chequeo	2
01/11/2023	Se ajusta el documento a las especificaciones de la NTC PE 1000:2020 y guía actualizada para la elaboración del documento metodológico	3

12. APROBACIÓN

Elaborado por:	Revisado por:	Aprobado por:
Nombre: Alicia Herrera Sánchez y Efrén Romero Riaño **Cargo:** Asistente de Investigación y Líder de bibliometría **Fecha**: 01/11/2023	**Nombre:** Efrén Romero Riaño **Cargo:** Líder de Bibliometría **Fecha**: 04/11/2023	**Nombre:** Andrea Guevara **Cargo:** Líder de inversión **Fecha**: 07/11/2023

13. ANEXOS

Anexo 1. Glosario

Para los efectos del presente documento es importante tener claridad sobre los siguientes conceptos:

- **Análisis de citas**: técnica bibliométrica que estudia las citas como indicadores de la calidad de una publicación científica, atendiendo tanto a sus aspectos cuantitativos como cualitativos, como reflejo del reconocimiento de una publicación anterior. Permite examinar hábitos de comportamiento en la investigación e identificar relaciones entre temáticas, autores e instituciones, así como su evolución en el tiempo.

- **Base de datos bibliográficos**: conjunto de registros con información bibliográfica (autor, título de la contribución, de la publicación, fecha de la publicación, editorial, entre otros), almacenado y gestionado mediante sistemas informáticos. Muchas de estas bases de datos contienen descriptores, palabras claves y resúmenes. Y unas pocas contienen citas (índices de citas).

- **Bases de datos bibliométricas**: Recursos de información que suministran no solo información bibliográfica sino indicadores relativos a distintas facetas (impacto, citación, obsolescencia, etc.) de las publicaciones científicas recogidas, tanto referidos a autores, como a instituciones, publicaciones, países, materias u otros agentes intervinientes. Entre las principales podemos señalar Web of Science, Scopus, Scielo.

- **Bibliometría**: Disciplina que emplea métodos y técnicas cuantitativas para medir y evaluar la publicación de la investigación científica y las relaciones entre los agentes que la producen.

- **Factor de impacto (FI):**es una medida de la frecuencia con la que el artículo medio de una revista ha sido citado en un período. Básicamente mide la relación entre las citas recibidas y los artículos publicados en una revista.

- **Filiación institucional:** por la cual un autor indica expresamente en su publicación a qué institución pertenece.

- **Indicadores bibliométricos:** miden distintos aspectos cuantitativos sobre la generación, uso e impacto de la producción académica dentro de un área temática o un espacio geográfico determinado.

- **Indicadores de colaboración:** miden las relaciones que han existido entre los productores o agentes científicos y que han finalizado con la publicación conjunta de resultados científicos. Se basan en los datos de autoría.

- **Indicadores de producción:** se basan en el recuento de publicaciones científicas. Son los más fáciles de utilizar ya que nos proporcionan información sobre las características de las unidades analizadas. El número de publicaciones es una medida razonada de la actividad científica, pero no del progreso científico. Estos indicadores están condicionados por factores políticos y sociales. Miden la productividad científica, que es la cantidad de publicaciones producida por un autor, país o institución durante un período determinado.

- **Indicadores de visibilidad o impacto:** Se basan en los recuentos del número de citaciones que reciben los documentos durante un período determinado de tiempo o a partir de su publicación.

- **Índice h:** Indicador bibliométrico por Hirsch en 2005 que mide la cantidad de trabajos publicados por un investigador con igual o mayor número de citas. Por ejemplo, un investigador tiene un Índice H = 5 si cuenta con 5 trabajos con al menos 5 citas cada uno. Por su propia esencia, es especialmente idóneo para evaluar trayectorias científicas consolidadas.

- **Journal Citation Reports(JCR):**Base de datos que contiene información sobre parte de las revistas incluidas en las bases de datos de la Web of Science. Aporta tanto información descriptiva de las publicaciones como datos sobre el impacto de las mismas. Publica dos ediciones, una para las revistas de Ciencias Sociales indexadas en la base de datos Social Science Citation y otra para las revistas de

Ciencias Experimentales y Tecnológicas indexadas en la base de datos Science Citation Index (JCR-ScienceEdition).

- **Peer-review:** La revisión por pares, aspecto clave en las publicaciones científicas, representa un filtro de calidad mediante el cual especialistas de la misma materia establecen si una propuesta de publicación reúne las características y requisitos exigibles para ser publicada.

- **Publicación académica:** Publicación resultado de actividades de investigación, cuyo objetivo es aportar al desarrollo de un área científica, técnica o humanística determinada. Este tipo de publicación se suele difundir a través de artículos, libros, actas de congresos, etc.

- **Publicación indizada**: es la que ha sido sometida a un proceso de selección y análisis previo por parte de los servicios de indización para asegurar que sus contenidos cumplan con criterios de calidad como: cumplimiento de periodicidad, presencia de Comité Editorial y Científico, proceso de revisión por pares externos, entre otros.

- **Revista científica**: medio especializado donde la investigación científica puede publicarse.

- **Scielo**: Biblioteca científica en formato electrónico (Sigla en inglés)

- **Scopus:** Base de datos multidisciplinar, contiene resúmenes y citas de artículos publicados en revistas científicas.

- **WoS**: Web of Science

Anexo 2. Especificaciones de validación

ESPECIFICACIONES DE VALIDACIÓN Y CONSISTENCIA

ESTADÍSTICAS DE PRODUCCIÓN CIENTÍFICA BIBLIOMÉTRICA

CODIGO:
VERSIÓN:
FECHA:

Tabla	Denominación	Campo		Descripción	Dominio	Regla de validación	Sentencia SQL
		Tipo	Long				
prod_principal	cod_ti	Carácter	10	codigo del titulo del documento	0-9999999	1. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE cod_ti IS NULL
prod_principal	py	Numerico	4	Año de publicación	[2007, 2017]	1. Valores de dominio entre 2007 y 2017	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE py NOT BETWEEN 2007 AND 2017
prod_principal	bd	Numerico	1	Origen de base de datos (Wos, Scielo, Scopus)	1 : Scopus 2 : Scielo 3 : Wos 4 : Scopus, Scielo 5 : Scielo, Scopus 6 : Scopus, Wos 7 : Scopus, Scielo, Wos	1. Valores de dominio entre 0 y 7 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE bd NOT BETWEEN 0 AND 7 SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE bd IS NULL
prod_principal	tipo	Carácter	9	Si el documento es articulo o pertenece a otro tipo de documento	Articulo, Otros	1. Valores de dominio (Articulo, Otros) 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE tipo NOT IN ('Articulo','Otros') SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE tipo IS NULL"
prod_principal	coautorias	Carácter	20	Tipo de coautoria del documento	Coautoria Nacional, Internacional, Solo autor	1. Valores de dominio (Coautoria|Nacional, Internacional, Solo|autor) 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE coautorias NOT IN ('Coautoria|Nacional', 'Internacional', 'Solo|autor') SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE coautorias IS NULL
prod_principal	citas	Numerico	3	Numero de citas del documento	>=0	1. Valores de dominio >=0 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas <0 SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas IS NULL
prod_principal	citas_scopus	Numerico	3	Numero de citas en scopus	>=0	1. Valores de dominio >=0 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas_scopus <0 SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas_scopus IS NULL
prod_principal	citas_wos	Numerico	3	Numero de citas en wos	>=0	1. Valores de dominio >=0 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas_wos <0 SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas_wos IS NULL
prod_principal	citas_scielo	Numerico	3	Numero de citas en scielo	>=0	1. Valores de dominio >=0 2. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas_scielo <0 SELECT * FROM OCYTRG_PROD_PRINCIPAL WHERE citas_scielo IS NULL
prod_area	cod_ti	Carácter	10	Código del titulo del documento	0-9999999	1. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_AREA WHERE cod_ti IS NULL SELECT * FROM OCYTRG_PROD_AREA WHERE cod_gran_area IS NULL
prod_area	cod_gran_area	Numerico	1	Código de gran área de conocimiento al que pertenece el documento	1 : Natural Sciences 2 : Agricultural Sciences 3 : Medical and Health Sciences 4 : Engineering and Technology 5 : Social Sciences 6 : Humanities 7 : Multidisciplinary	1. Campo Obligatorio no puede tener valores nulos 2. Valores de dominio 1 - 7	SELECT * FROM OCYTRG_PROD_AREA WHERE cod_gran_area IS NULL SELECT * FROM OCYTRG_PROD_AREA WHERE cod_gran_area NOT BETWEEN 1 AND 7
prod_area	nombre_gran_area	Carácter	30	Nombre de gran área de conocimiento al que pertenece el documento	Natural Sciences;Agricultural Sciences; Medical and Health ; Sciences ; Engineering and Technology ;Social Sciences; Humanities ; Multidisciplinary	1. Campo Obligatorio no puede tener valores nulos 2. Valores de dominio ('Natural Sciences','Agricultural Sciences','Medical and Health','Sciences', 'Engineering and Technology','Social Sciences','Humanities','Multidisciplinary')	SELECT * FROM OCYTRG_PROD_AREA WHERE nombre_gran_area IS NULL SELECT * FROM OCYTRG_PROD_AREA WHERE nombre_gran_area NOT IN ('Natural Sciences','Agricultural Sciences','Medical and Health Sciences', 'Engineering and Technology','Social Sciences','Humanities','Multidisciplinary')
prod_inst	cod_ti	Carácter	10	Código del titulo del documento	0-9999999	1. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_AREA WHERE cod_ti IS NULL SELECT * FROM OCYTRG_PROD_AREA WHERE cod_gran_area IS NULL
prod_inst	institucion_normalizada	Carácter	300	Nombre de la Institucion		1. Campo Obligatorio no puede tener valores nulos	SELECT * FROM OCYTRG_PROD_INS WHERE institucion_normalizada IS NULL
prod_inst	nme_departamento	Carácter	30	Nombre del Departamento	Ver Tabla de referencia DIVIPOLA	1. Campo Obligatorio no puede tener valores nulos 2. El valor del campo debe existir en la tabla Divipola, para el caso de no identificado no se encuentra en Divipola	SELECT * FROM OCYTRG_PROD_INS WHERE nme_departamento IS NULL SELECT * FROM OCYTRG_PROD_INS WHERE nme_departamento NOT IN (SELECT NOM_DEPTO FROM REF_DIVIPOLA_CENP_CABE) AND nme_departamento != 'No identificado'
prod_inst	cod_departamento	Carácter	3	Código de divipola para los Departamentos	Ver Tabla de referencia DIVIPOLA	1. Campo Obligatorio no puede tener valores nulos 2. El valor del campo debe existir en la tabla Divipola, para el caso del código 999 no se encuentra en Divipola	SELECT * FROM OCYTRG_PROD_INS WHERE cod_departamento IS NULL SELECT * FROM OCYTRG_PROD_INS WHERE cod_departamento NOT IN (SELECT COD_DEPTO FROM REF_DIVIPOLA_CENP_CABE) AND cod_departamento != 999