Operación estadística de producción científica bibliométrica.

INSTRUCTIVO DE RECOLECCIÓN

1. Objetivo

Establecer los procedimientos para realizar la recopilación de datos para la operación estadística de producción científica bibliométrica, que respondan a necesidades identificadas y a criterios de calidad técnica.

2. Alcance

Este instructivo aplica para la operación Estadísticas Producción Científica Bibliométrica que se encuentra cubierta por el alcance del SGC - ISO 9001:2015 y por la NTCPE 1000:2020.

3. Abreviaturas

OCyT: Observatorio Colombiano de Ciencia y Tecnología.

WoS: Web Of Science.

SEN: Sistema estadístico nacional

SGC: Sistema de Gestión de Calidad.

NTC PE: Norma Técnica de Calidad del proceso estadístico.

CTeI: Ciencia Tecnología e Innovación

4. Definiciones

Para los fines de este documento, se aplican los términos y las definiciones utilizadas en las bases de datos de consulta (WoS y Scopus), NTC-ISO 9000, el Decreto 2404 de 2019 y los especificados en la NTCPE 1000:2020.

5. Lineamientos Generales

De acuerdo con lo establecido en el Decreto 2404 de 2019 "por el cual se reglamenta el artículo 155 de la Ley 1955 de 2019 y se modifica el Título 3 de la Parte 2 del Libro 2 del Decreto 1170 de 2015 Único del Sector Administrativo de Información Estadística", el Sistema Estadístico Nacional (SEN) tiene como finalidad establecer e implementar un esquema de coordinación y articulación entre los componentes que lo conforman, que permita mejorar la información estadística producida para la toma de decisiones a nivel nacional y territorial con estándares de calidad, con lenguajes y procedimientos comunes, respetuosos de los estándares estadísticos internacionales y que contribuyan a la transparencia, pertinencia, interoperabilidad, acceso, oportunidad y coherencia de las estadísticas producidas en el país; de modo que la formulación de políticas públicas estén soportadas en evidencia verificables que propenda por una mejora en las condiciones de vida de la sociedad en general[1].

Por su parte el OCyT en cumplimiento de su objetivo: "*fortalecer la capacidad nacional para generar y usar indicadores que servirán para orientar y evaluar las políticas nacionales, regionales e internacionales, así como la acción de diversos actores en los polos científicos y tecnológicos*[2]", ha decidido implementar los

1 Decreto 2404 de 2019 (diciembre 27) Por el cual se reglamenta el artículo 155 de la Ley 1955 de 2019 y se modifica el Título 3 de la Parte 2 del Libro 2 del Decreto 1170 de 2015 Único del Sector Administrativo de Información Estadística.

2 Estatutos OCyT.

requisitos de la NTCPE 1000:2020 con el fin de facilitar que las operaciones del Observatorio hagan parte de las estadísticas oficiales nacionales y territoriales de ciencia y tecnología, así como también, fortalecer la relevancia de la operación para generar información que apoye un mejor seguimiento a las políticas públicas y la toma de decisiones sectoriales.

6. Desarrollo de Actividades

6.1. Perfiles y roles del equipo de trabajo

El personal necesario para realizar el operativo de bibliometría para 2022 se puede encontrar en la **matriz de roles y responsabilidades**.

6.2. Actividades previas a la recolección

- Actividades preparatorias

Por las características de la forma de recolección para las estadísticas de Producción Científica Bibliométrica, la cual se realiza a partir de la descarga de información de Web of Science (WoS), Scopus y Scielo, no se realiza actividades de sensibilización a fuentes. Sin embargo, se hace una revisión de los cambios que pueda presentar la estructura de las bases de datos antes de iniciar las descargas.

- Entrenamiento

De acuerdo con los lineamientos del OCyT, se realizarán sesiones de entrenamiento al equipo responsable del diseño y ejecución del operativo

estadístico en donde se deben impartir los conocimientos básicos teniendo en cuenta los requerimientos de la NTC PE 1000:2020 y las habilidades necesarias el desarrollo de la operación estadística y las actividades involucradas en la recolección de información[3].

Igualmente, se contempla la participación en talleres y encuentros que estén relacionados con temas estadísticos que permitan actualizar los conocimientos y obtener retroalimentación sobre los procesos de aprovechamiento de información que se realizan a nivel nacional, regional e internacional.

6.3. Recolección o captura de la información

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 fuentes WoS, Scopus y Scielo, insumo con el cual se producen las estadísticas de Ciencia y Tecnología.

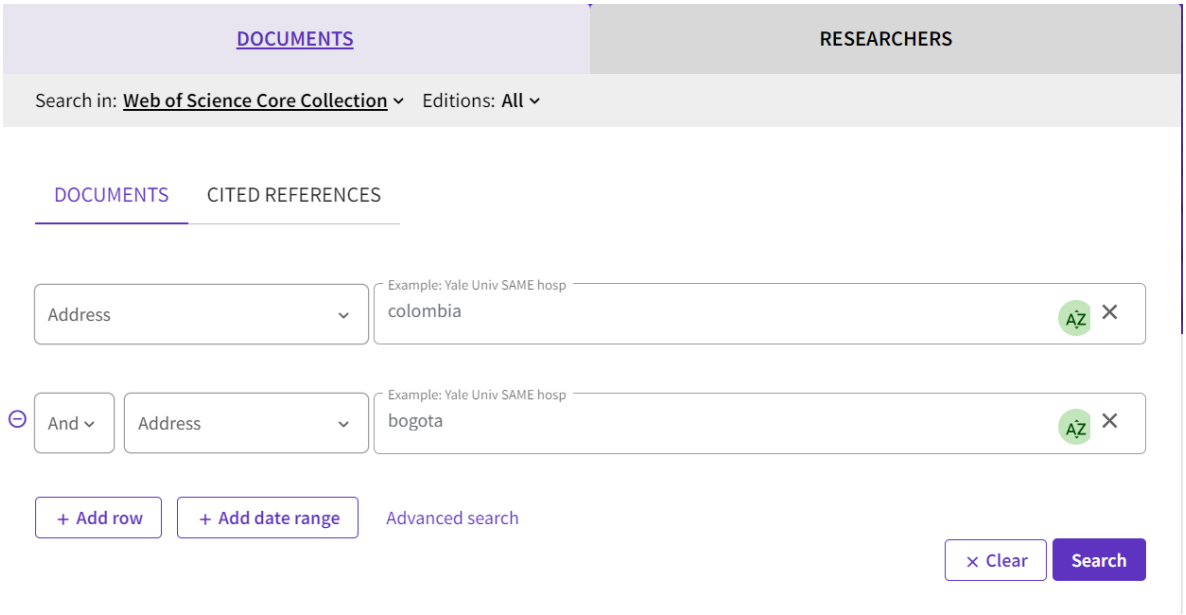
[3] Revisar el plan de entrenamiento

- **Descarga de WoS**

Como punto de partida se inicia con la descarga a partir de una consulta que se realiza en Web of Science, como se describe a continuación:

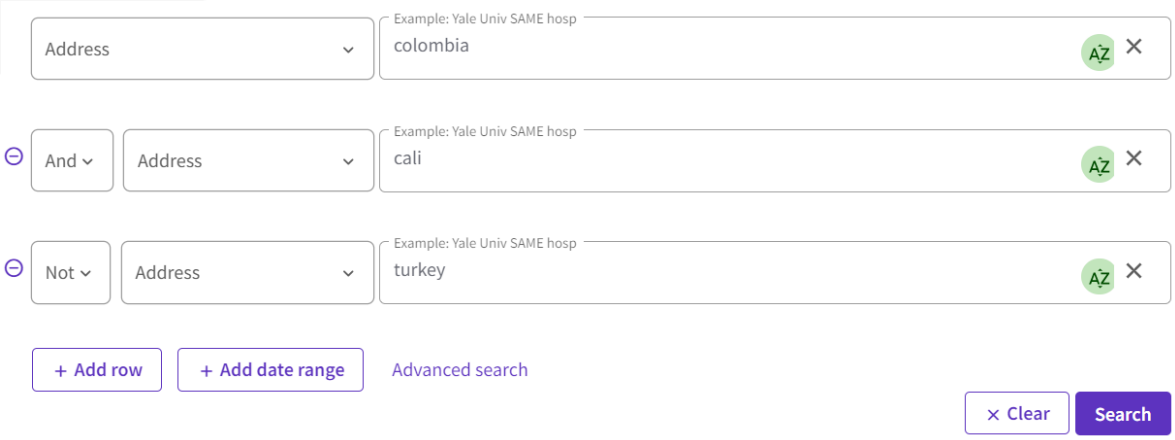
1. Ingresar a la página de Web of Science: https://www.webofscience.com/wos
2. Ingresar las credenciales de acceso, estas serán proporcionadas por el OCyT, en el caso que el responsable de la recolección no cuente con el acceso
3. Realizar la búsqueda básica seleccionando la colección principal de WoS (Web of Science Core Collection en gráfico 1), en esta se realiza la búsqueda del parámetro "Address" colocando Colombia y el municipio de interés a consultar, a continuación, se muestra un ejemplo:

Gráfico 1. Ejemplo de consulta/ecuación para descarga en WoS



4. Antes de continuar con la consulta, es necesario indagar por ciudades homónimas a nivel mundial para filtrar en la búsqueda y no recuperar documentos producidos por otros países, en este sentido con la ayuda de ChatGPT[4] y verificando con el directorio de ciudades homónimas[5] se realiza la búsqueda para filtrar con el parámetro "NOT". En caso de que no existan ciudades homónimas este paso es omitido.

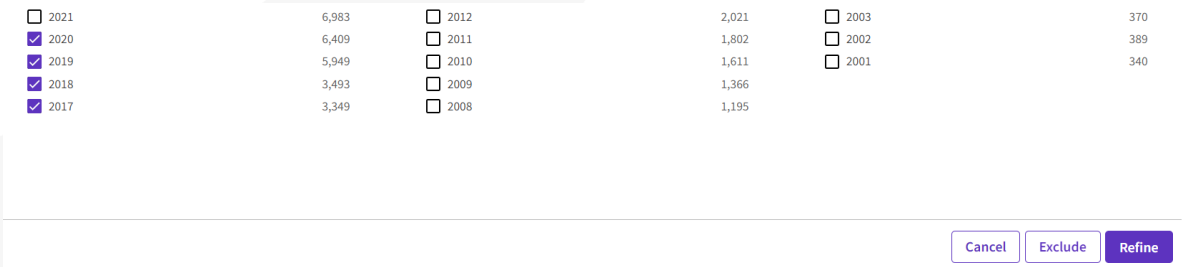
Gráfico 2. Ejemplo de consulta/ecuación para descarga en WoS con filtro por ciudad homónima



5. Posteriormente se procede con la búsqueda en la opción Search
6. Luego de obtener los resultados totales, se realiza el filtro para los años de interés, ejemplo:

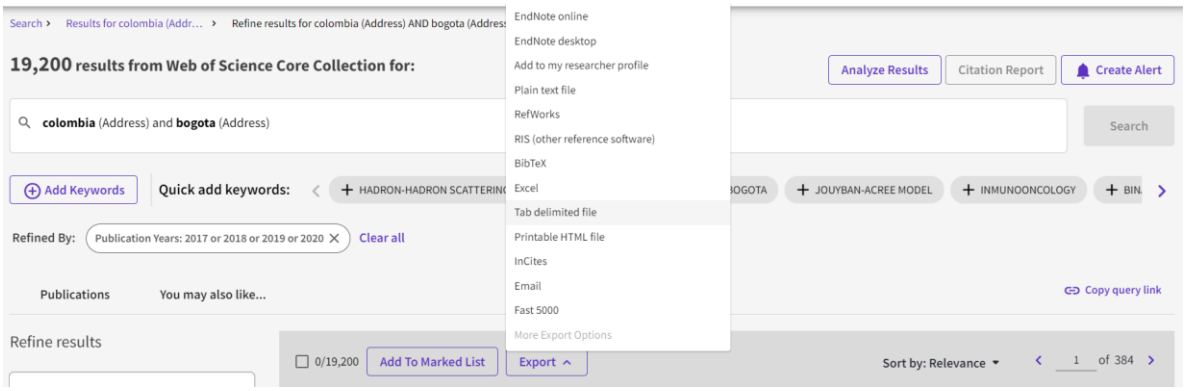
[4] Con el prompt "¿en qué ciudades se tienen ciudades con nombre de _____? Ejemplo: Cali".
[5] https://www.geodatos.net/en/homonymous-cities

Gráfico 3. Ejemplo para selección de años de interés para consulta



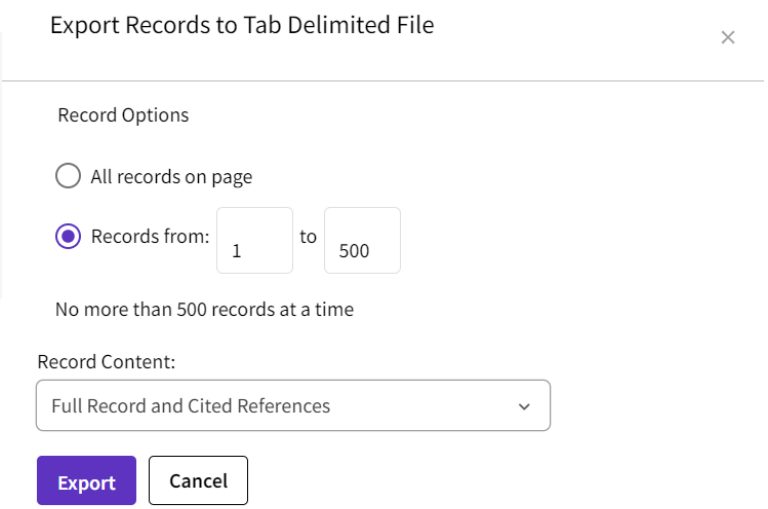
7. Teniendo los resultados de los cuatro años de interés para medición, se procede con las descargas. Es importante saber, que para WoS los lotes tienen un límite de 500 registros para exportación. En primer lugar se seleccionará el parámetro Tab delimited file como se muestra a continuación:

Gráfico 4. Ejemplo para parámetro inicial de descarga Tab delimited file



8. Luego se seleccionará el parámetro Full record and Cited references así:

Gráfico 5. Ejemplo para segundo parámetro de descarga Full record and Cited References



9. Se procederá con la descarga en la opción export obteniendo un archivo .txt
10. Se procederá con la repetición de la operación señalando los registros faltantes para la descarga de los lotes, por ejemplo, para la continuación del anterior será Records from: 501 to 1000

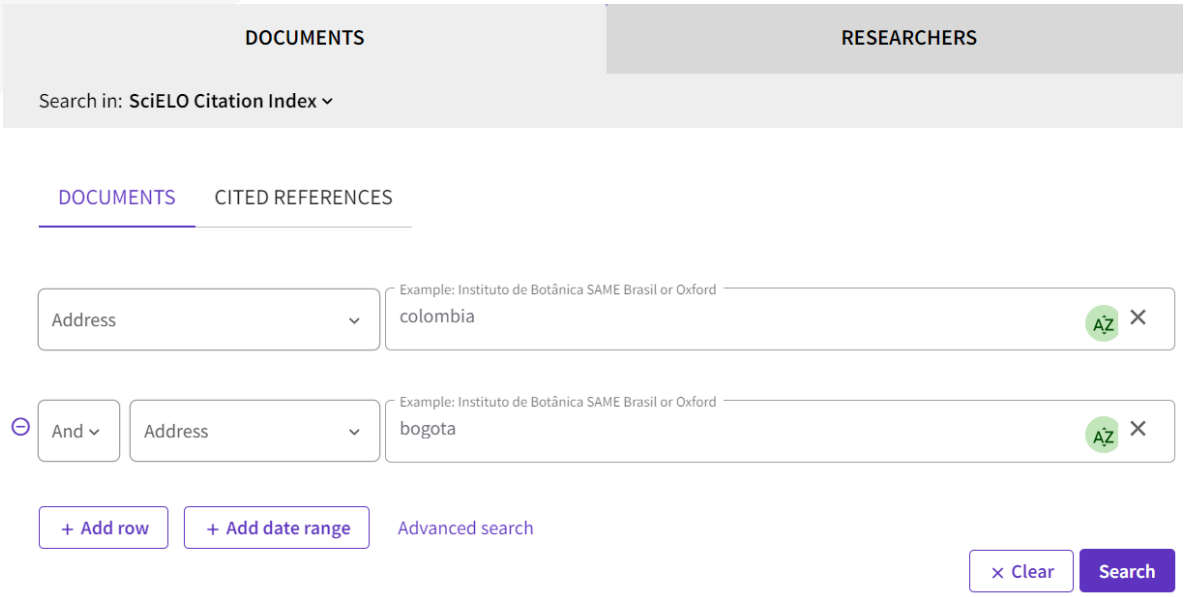
- **Descarga de Scielo**

Como punto de partida se inicia con la descarga a partir de una consulta que se realiza en Web of Science, como se describe a continuación:

1. Ingresar a la página de Web of Science: https://www.webofscience.com/wos
2. Ingresar las credenciales de acceso, estas serán proporcionadas por el OCyT, en el caso que el responsable de la recolección no cuente con el acceso

3. Realizar la búsqueda básica seleccionando Scielo (Scielo Citation Index en gráfico 6), colección principal del Web of Science (Web of Science Core Collection en gráfico 1), en esta, se realiza la búsqueda del parámetro "Address" colocando Colombia y el municipio de interés a consultar, a continuación, se muestra un ejemplo:

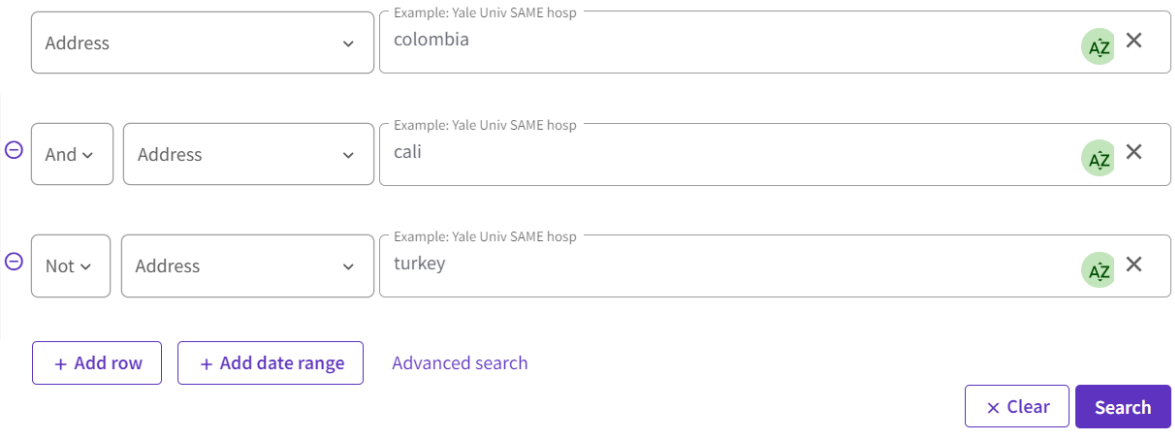
Gráfico 6. Ejemplo de consulta/ecuación para descarga en WoS



4. Antes de continuar con la consulta, es necesario indagar por ciudades homónimas a nivel mundial para filtrar en la búsqueda y no recuperar documentos producidos por otros países, en este sentido con la ayuda de ChatGPT[6] y verificando con el directorio de ciudades homónimas[7] se realiza la búsqueda para filtrar con el parámetro "NOT". En caso de que no existan ciudades homónimas este paso es omitido.

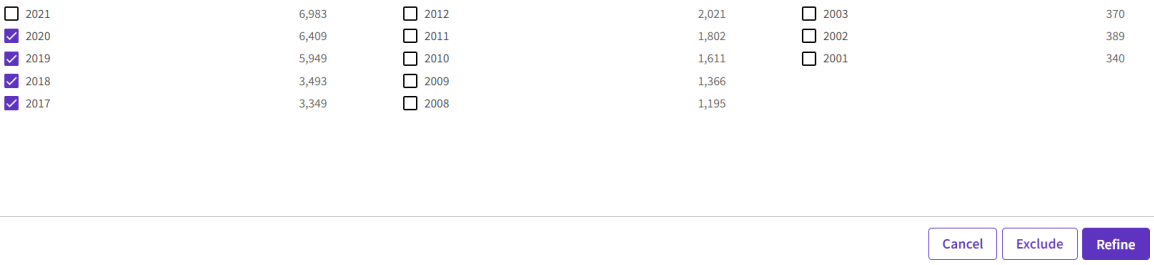
[6] Con el prompt "¿en qué ciudades se tienen ciudades con nombre de _____ Ejemplo?: Cali".
[7] https://www.geodatos.net/en/homonymous-cities

Gráfico 7. Ejemplo de consulta/ecuación para descarga en WoS con filtro por ciudad homónima



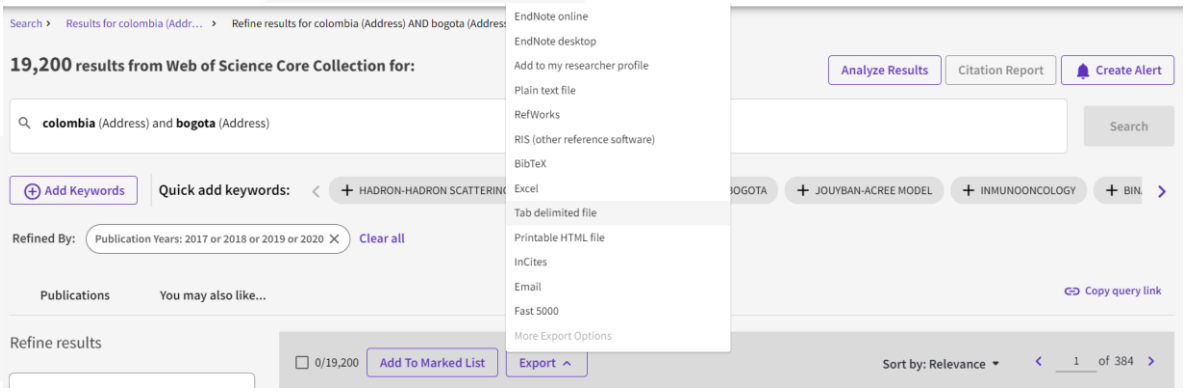
5. Posteriormente se procede con la búsqueda en la opción Search
6. Luego de obtener los resultados totales, se realiza el filtro para los años de interés, ejemplo:

Gráfico 8. Ejemplo para selección de años de interés para consulta



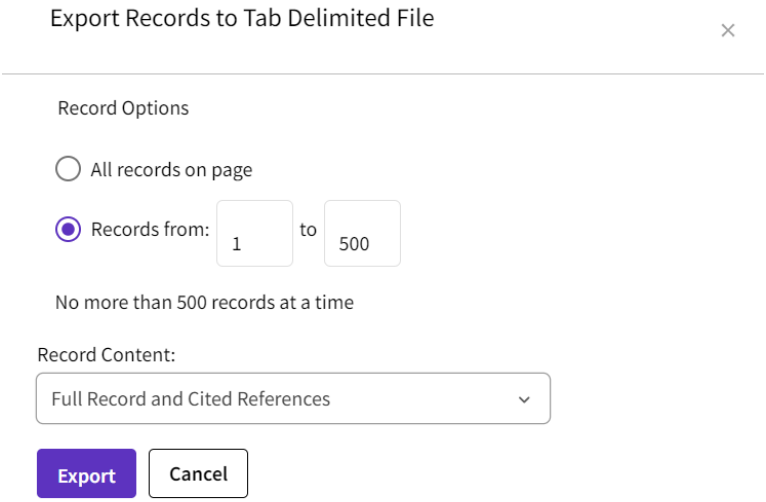
7. Teniendo los resultados de los cuatro años de interés para medición, se procede con las descargas. Es importante saber que para WoS los lotes tienen un límite de registros para exportación, el cual es 500. En primer lugar se seleccionará el parámetro Tab delimited file como se muestra a continuación:

Gráfico 9. Ejemplo para parámetro inicial de descarga Tab delimited file



8. Luego se seleccionará el parámetro Full record and Cited references así:

Gráfico 10. Ejemplo para segundo parámetro de descarga Full record and Cited References



9. Se procederá con la descarga en la opción export obteniendo un archivo .txt

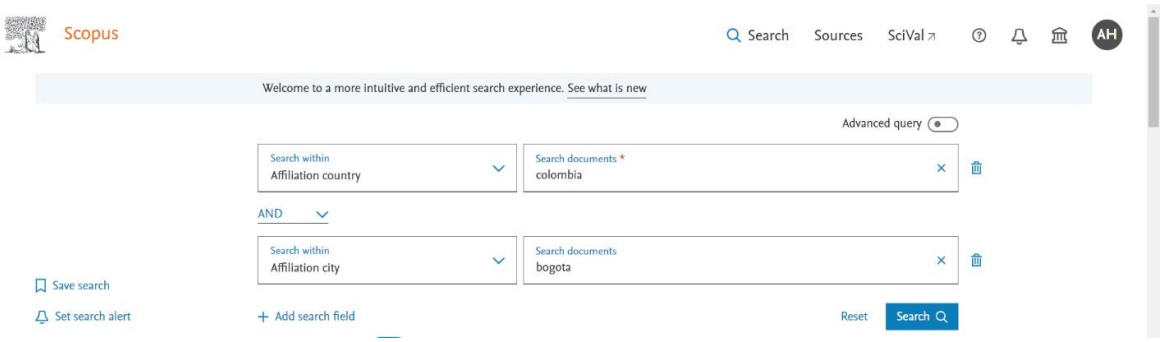
10. Se procederá con la repetición de la operación señalando los registros faltantes para la descarga de los lotes, por ejemplo para la continuación del anterior será Records from: 501 to 1000

Descarga de Scopus

Como punto de partida se inicia con la descarga a partir de una consulta que se realiza en Scopus, como se describe a continuación:

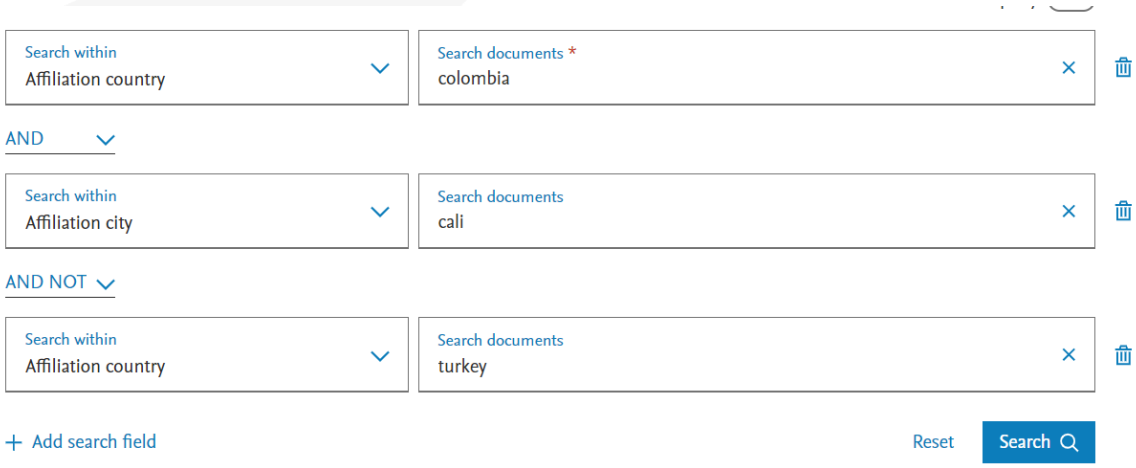
1. Ingresar a la página de Scopus: https://www.scopus.com/home.uri
2. Ingresar las credenciales de acceso, estas serán proporcionadas por el OCyT, en el caso que el responsable de la recolección no cuente con el acceso
3. Realizar la búsqueda básica, en esta se realiza la búsqueda del parámetro "Affiliation Country" colocando Colombia y con el parámetro "Affiliation City" el municipio de interés a consultar, a continuación, se muestra un ejemplo:

Gráfico 11. Ejemplo de consulta/ecuación para Scopus



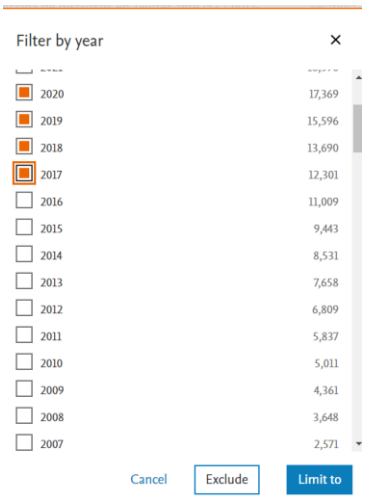
4. Antes de continuar con la consulta, es necesario indagar por ciudades homónimas a nivel mundial para filtrar en la búsqueda y no recuperar documentos producidos por otros países, en este sentido con la ayuda de ChatGPT y verificando con el directorio de ciudades homónimas se realiza la búsqueda para filtrar con el parámetro "NOT". En caso de que no existan ciudades homónimas este paso es omitido.

Gráfico 12. Ejemplo de consulta/ecuación para descarga en Scopus con filtro por ciudad homónima



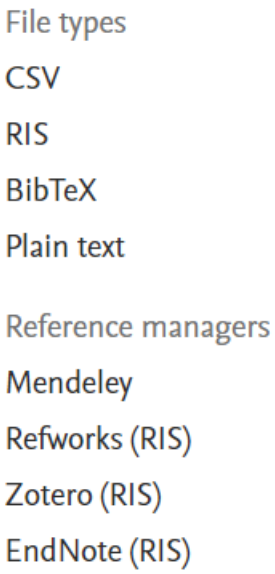
5. Posteriormente se procede con la búsqueda en la opción Search
6. Luego de obtener los resultados totales, se realiza el filtro para los años de interés, ejemplo:

Gráfico 13. Ejemplo para filtro de años de interés en el operativo Scopus



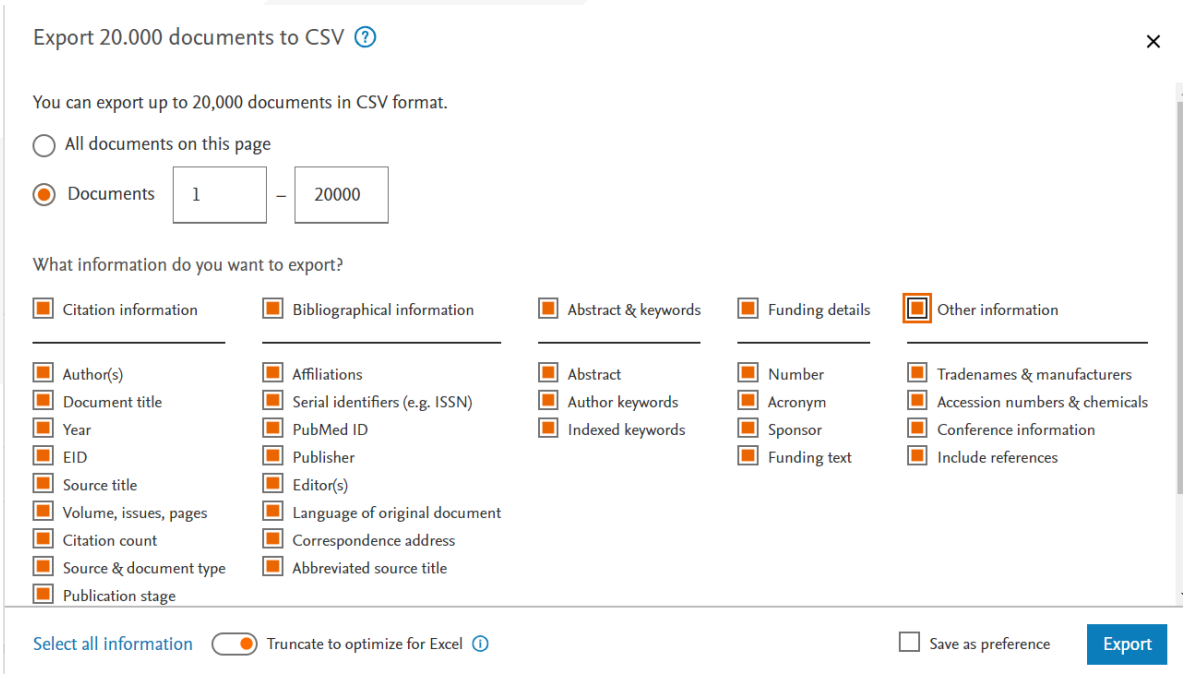
7. Teniendo los resultados de los cuatro años de interés para medición, se procede con las descargas. Es importante saber que para Scopus los lotes tienen un límite de registros para exportación, el cual es 20000. Las descargas se deberán realizar en archivos .csv y .ris seleccionando la opción del listado, por ejemplo para CSV es así:

Gráfico 14. Ejemplo para selección de descarga en formato .csv



8. Para la descarga se seleccionarán todos los parámetros para descarga, como se muestra a continuación en formato CSV y RIS:

Gráfico 15. Ejemplo para la selección de los parámetros de descarga en Scopus



9. Con estos se procederá con la descarga en la opción export obteniendo un archivo .csv o .ris según corresponda
10. Se procederá con la repetición de la operación señalando los registros faltantes para la descarga de los lotes, por ejemplo, para la continuación del anterior será Documents: 20001 to 29000

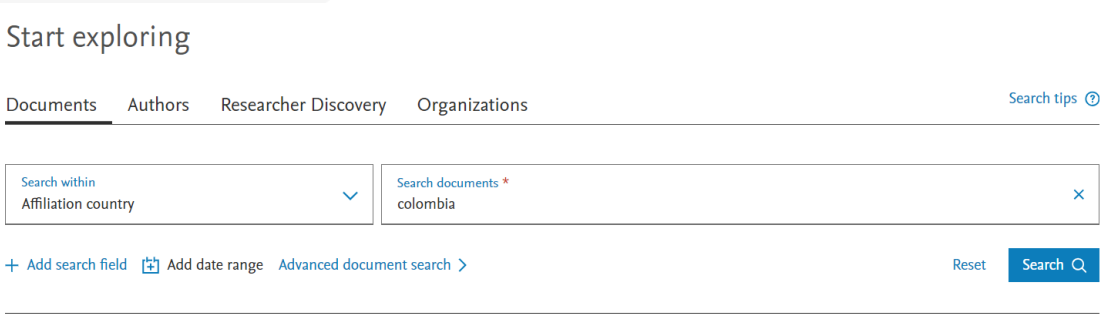
Descarga de Scopus para áreas de conocimiento

Como punto de partida se inicia con la descarga a partir de una consulta que se realiza en Scopus, como se describe a continuación:

1. Ingresar a la página de Scopus: https://www.scopus.com/home.uri
2. Ingresar las credenciales de acceso, estas serán proporcionadas por el OCyT, en el caso que el responsable de la recolección no cuente con el acceso

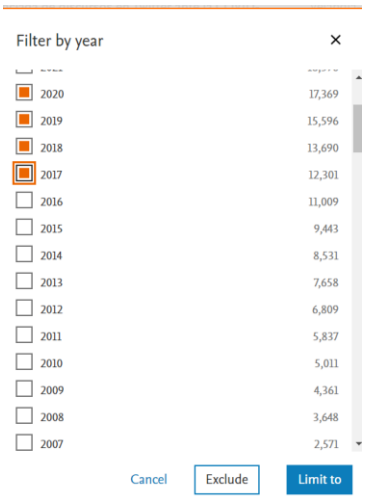
3. Realizar la búsqueda básica, en esta se realiza la búsqueda del parámetro "Affiliation Country" colocando, a continuación, se muestra un ejemplo:

Gráfico 16. Ejemplo de consulta/ecuación para Scopus para áreas de conocimiento



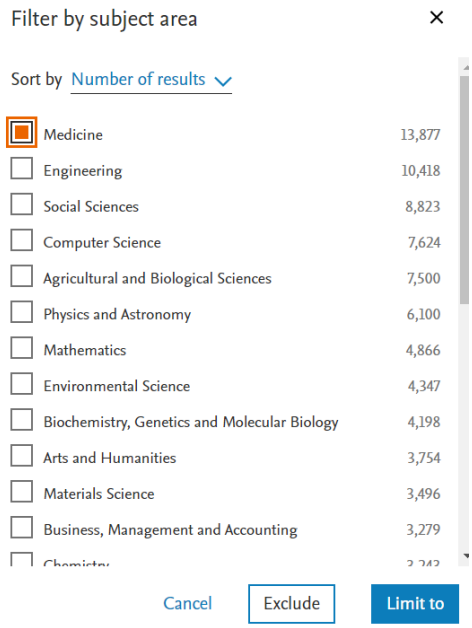
4. Posteriormente se procede con la búsqueda en la opción Search
5. Luego de obtener los resultados totales, se realiza el filtro para los años de interés, ejemplo:

Gráfico 17. Ejemplo para filtro de años de interés en el operativo Scopus



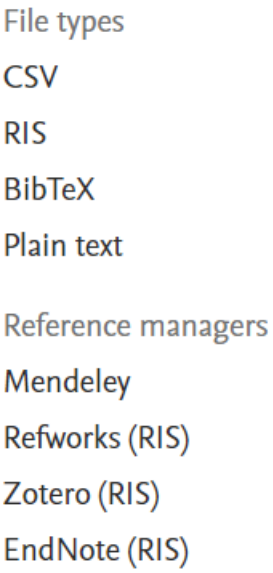
6. Teniendo los resultados de los cuatro años de interés para medición, se procede con el filtrado por áreas para las descargas. Para cada descarga se selecciona un área, en este caso el etiquetado de Scopus para el filtro es subject área y la selección para el filtro se realiza como se presenta a continuación:

Gráfico 18. Ejemplo para filtro por subject area en Scopus



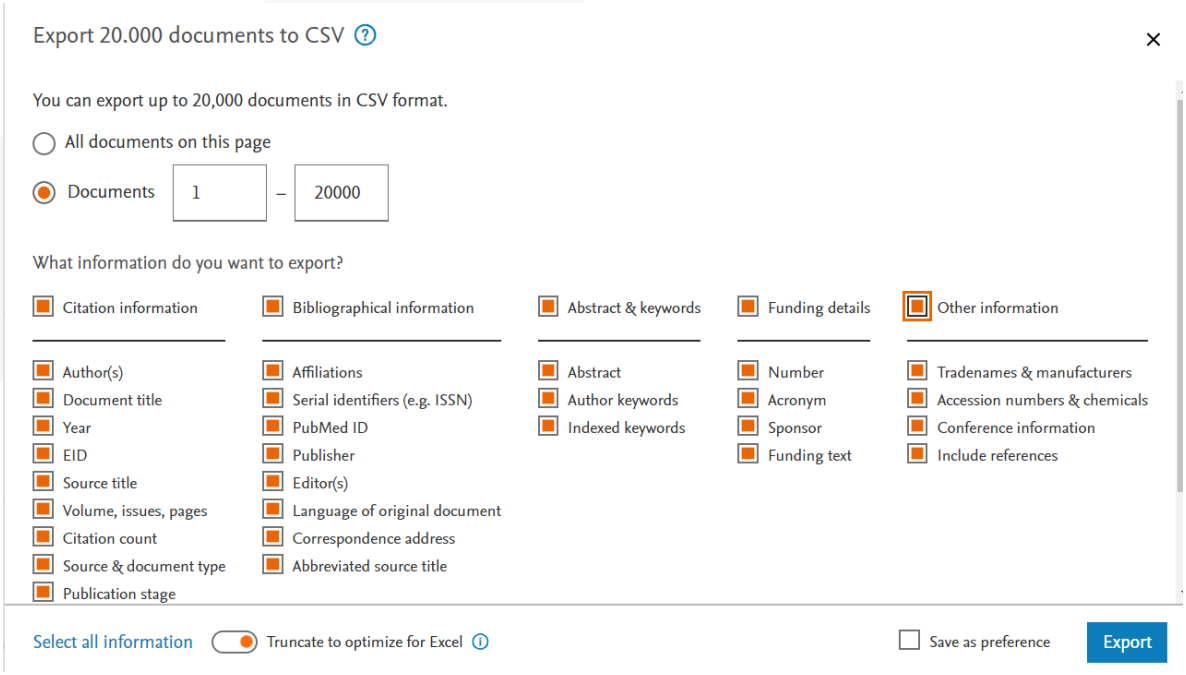
7. Es importante saber que para Scopus los lotes tienen un límite de registros para exportación, el cual es 20000. Las descargas se deberán realizar en archivos .csv seleccionando la opción del listado, así:

Gráfico 19. Ejemplo para selección de descarga en formato .csv



8. Para la descarga se seleccionarán todos los parámetros para descarga, como se muestra a continuación en formato CSV:

Gráfico 20. Ejemplo para la selección de los parámetros de descarga en Scopus



9. Con estos se procederá con la descarga en la opción export obteniendo un archivo .csv según corresponda
10. Se procederá con la repetición de la operación señalando los registros faltantes para la descarga de los lotes, por ejemplo para la continuación del anterior será Documents: 20001 to 29000

6.4. Inconsistencias

Inconsistencias	Causal	Control
Errores en la descarga del tipo de archivo	Incorrecta exportación del archivo .txt, .csv o .ris	Verificar los tipos de archivo en las descargas en bruto
Errores en la descarga por lotes con mensaje en la plataforma	Error propio en la fuente	Consulta con la fuente directamente por medio de PQRS que contenga ecuación y error arrojado

Error en la descarga por periodo	Ingreso incorrecto de la consulta, ecuación o filtro para los años de interés	Como medida de calidad el programa en la etapa de procesamiento solo procesará los años de interés

6.5. Validaciones

Los métodos de validación en el proceso de recolección de datos incluyen diversas técnicas cruciales para garantizar la calidad y precisión de la información. Para la operación estadística de bibliometría se realiza la revisión de variables descargadas y el diligenciamiento del formato de verificación Primero, se realiza una revisión exhaustiva para detectar y corregir cualquier carácter no numérico en los campos de la base de datos, asegurando así la integridad de los datos. Además, se verifica que las variables utilizadas en los cálculos de los indicadores sean las requeridas y adecuadas. La información territorial debe contar con una cobertura nacional completa, garantizando que todos los territorios estén representados adecuadamente sin ninguna representación atípica. Las tipologías por autoría deben ser las descritas en las fichas de indicadores, así como con las áreas establecidas por la OCDE. Otro importante método de verificación es el diligenciamiento del formulario de verificación de descargas, el cual contempla variables de seguimiento por fuente de descarga, asegurando un control y monitoreo de todas las fuentes de datos utilizadas para el proceso estadístico.

6.6 Control de sesgos

Es importante aplicar controles de sesgo para garantizar que los datos sean precisos y representativos. A continuación, se presentan algunos controles de sesgo que se pueden aplicar en la operación estadística de bibliometría.

6.6.1 Muestreo representativo

Asegurarse de que la muestra de las instituciones que hacen parte del universo de la medición sea representativas de la población por tipo de institución. En la actualidad este control no se aplica, ya que al ser una operación estadística a partir

de registros administrativos se recolecta todos los registros que las bases de datos contengan.

6.6.4 Control de sesgo de retroceso

Asegurarse de que los datos recopilados reflejen el período de tiempo deseado y no se vean afectados. La información que se recoge en cada operativo corresponde a los cuatro años anteriores al operativo. Esto permite tener cifras cerradas contablemente lo que facilita el reporte y tener claridad del año de referencia.

6.6.5 Control de sesgo geográfico y sectorial

Asegurarse de que la medición incluya una representación adecuada de diferentes regiones geográficas y sectores de la economía para evitar un sesgo regional o sectorial. Para lograr una correcta representatividad se realiza las consultas en las bases de datos de los registros para todos los municipios de todos los departamentos del país.

6.6.6 Revisión y auditoría de datos

Realizar auditorías y revisiones de datos para identificar y corregir posibles errores, inconsistencias o sesgos en la información recopilada. Esto proceso se hace por parte del investigador encargado del procesamiento quien revisa formatos o tipos de archivos, etiquetas de datos de las bases descargadas, características propias de las exportaciones en bruto de las bases.

6.7 Medidas de calidad

La calidad de las estadísticas de producción científica bibliométrica es esencial para garantizar la fiabilidad y utilidad de los datos. A continuación, se presentan algunas medidas de calidad que se pueden aplicar a la medición:

Medidas de calidad	**Descripción**	**Aplicación**
Validez	Asegurarse de que los datos reflejen de manera precisa y completa la producción científica bibliométrica nacional. Esto implica que los indicadores utilizados deben medir con precisión los conceptos que se pretenden evaluar.	Utilizar indicadores específicos, para medir la producción científica bibliométrica nacional en lugar de indicadores menos relacionados.
Fiabilidad	Garantizar que los datos sean consistentes y reproducibles con el tiempo y en diferentes situaciones. Esto implica utilizar métodos y técnicas de recopilación de datos que minimicen la variabilidad y el error.	Asegurarse de que los datos de producción científica bibliométrica nacional sean recopilados de manera consistente cada año, utilizando los mismos métodos y fuentes confiables.
Consistencia	Asegurar que los datos se recopilen y presenten de manera coherente a lo largo del tiempo y en diferentes fuentes. La consistencia es clave para permitir la comparación de datos a lo largo de diferentes períodos y entre diferentes fuentes.	Garantizar que las definiciones y clasificaciones utilizadas para la producción científica bibliométrica nacional sean consistentes con las normas internacionales y se mantengan constantes a lo largo del tiempo.
Precisión	Minimizar los errores y sesgos en la recopilación de datos. Esto implica el uso de definiciones claras y consistentes, así como la validación y verificación de datos.	Realizar auditorías periódicas de las fuentes de datos y utilizar técnicas de validación para reducir los errores en la recopilación de datos, como errores de entrada o duplicación.
Actualización y puntualidad	Mantener los datos actualizados y disponibles en un tiempo oportuno. La puntualidad es importante para que los datos sean relevantes para la toma de decisiones.	Publicar los datos sobre producción científica bibliométrica nacional anualmente.
Desagregación	Proporcionar datos desagregados que permitan un análisis más detallado y específicos con factores relevantes, esto puede ser importante para una comprensión más completa de la	Proporcionar datos desagregados para permitir un análisis detallado.

	producción científica bibliométrica nacional.	
Transparencia	Documentar claramente la metodología utilizada para recopilar, procesar y analizar los datos. Esto permite que los usuarios comprendan cómo se obtuvieron los datos y evalúen su calidad.	Documentar detalladamente la metodología utilizada para recopilar y procesar los datos.
Comparabilidad	Asegurarse de que los datos sean comparables con estándares internacionales y con datos de otras fuentes. La comparabilidad facilita la evaluación de la posición de un país o región en relación con otros.	Alinear los indicadores y clasificaciones utilizados con los estándares internacionales, para facilitar la comparación entre países.
Auditoría y revisión	Realizar auditorías y revisiones periódicas de los datos para identificar y corregir posibles errores y problemas de calidad.	Realizar auditorías de los datos y permitir que expertos independientes revisen y evalúen la calidad de los datos de producción científica bibliométrica nacional.
Consulta a expertos	Buscar la opinión y asesoramiento de expertos en el campo de CTI para garantizar que los indicadores y metodologías utilizados sean apropiados y se ajusten a las mejores prácticas.	Colaborar con expertos en estadísticas y en producción científica bibliométrica nacional para asesoramiento en la selección de indicadores, la metodología de recopilación de datos y la interpretación de resultados.

7. Ficha de aprobación

Elaborado por:	**Revisado por:**	**Aprobado por:**
Nombre: Alicia Herrera Sánchez **Cargo:** Asistente de Investigación **Fecha**: 10/10/2023	**Nombre:** Efrén Romero Riaño **Cargo:** Líder de Bibliometría **Fecha**: 10/10/2023	**Nombre:** Andrea Guevara **Cargo:** Líder de inversión **Fecha**: 2 13/10/2023

8. Control de Cambios

Identificar los criterios mencionados a continuación.

FECHA	DESCRIPCIÓN DEL CAMBIO	VERSIÓN
10/10/2023	Lanzamiento de documento	1