Operación estadística de producción científica bibliométrica.

INSTRUCTIVO DE PROCESAMIENTO

1. Objetivo

Establecer los procedimientos para realizar el procesamiento 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 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

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.

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 al procesamiento

- Actividades preparatorias

Las actividades preparatorias consisten en la adecuación de los archivos a procesar con los aplicativos[3] para las fuentes. Así, se realizan los cambios de ser necesario al nombramiento de las carpetas que contienen los datos en bruto verificando que queden tal cual se consigna en el archivo requerido por el aplicativo "DEPARTAMENTOS.xlsx". Igualmente se realiza el nombramiento de las áreas de conocimiento[4] para los archivos descargados por subject área para Scopus de acuerdo con lo consignado en el archivo "áreas_scopus.xlsx". Estos dos archivos se alojan en la carpeta base requerida por los aplicativos.

Por otra parte se verifica que en el archivo "periodos.xlsx" de la carpeta base se registren los años de interés correctos para el operativo en curso, colocando en la columna inicial el año inicial de interés al igual que en la final.

- Los aplicativos para el procesamiento son los siguientes:

wos-scielo: aplicativo que se encarga de procesar los datos obtenidos de las bases de datos WoS y Scielo (su detalle respecto a código fuente de ejecución de encuentra en el anexo para el aplicativo wos-scielo)

Scopus: aplicativo que se encarga de procesar los datos obtenidos de las bases de datos Scopus (su detalle respecto a código fuente de ejecución de encuentra en el anexo para el aplicativo Scopus)

Global-sco-wos: aplicativo que se encarga de la fusión de documentos generado por la aplicación Scopus y WoS-Scielo que están dentro de la capeta procesar_ resultados\producción\municipal (su detalle respecto a código fuente de ejecución de encuentra en el anexo para el aplicativo global-sco-wos)

[3] Este aplicativo se desarrolla con la versión de Python 3.10.9 y se utiliza la librería Pandas

[4] Para la homologación de las áreas de conocimiento en el caso de Web of Science se utiliza OECD Category Schema publicado por WoS, https://incites.help.clarivate.com/Content/Research-Areas/oecd-category-schema.htm y en el caso de Scopus se utiliza All Science Journal Classification Codes (ASJC) to FOS (Scopus), https://zenodo.org/records/8313183

- 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[5].

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. Procesamiento de los datos

A continuación, se describen las actividades correspondientes al procesamiento de los datos para obtener los indicadores de la operación estadística de bibliometría.

Verificación de requerimientos para las aplicaciones WoS - Scielo y Scopus.

Inicialmente se realiza la actividad de verificación respecto a los requerimientos para el funcionamiento de las aplicaciones o programas. En la ubicación donde se almacenan estas aplicaciones debe tener dos carpetas.

- Base: Hace referencia a los archivos esenciales para poder ejecutar las aplicaciones consta de 5 archivos o hojas de Excel de nombre:

[5] Revisar el plan de entrenamiento

- ✓ DEPARTAMENTO.xlsx
- ✓ areas-wos.xlsx
- ✓ periodos.xlsx
- ✓ areas_scopus.xlsx
- ✓ paises.xlsx

Asimismo, consta de la carpeta que contiene los datos de las áreas de Scopus que tiene el nombre de:

- ✓ areas_scopus

- Procesar: Hace referencia a la ubicación de archivos originales descargadas de las bases de datos científicas de Scopus, WOS y SciELO, se deben ubicar los archivos de forma ordenada y clasificada en orden DEPARTAMENTO/MUNICIPIO/ARCHIVO[6].

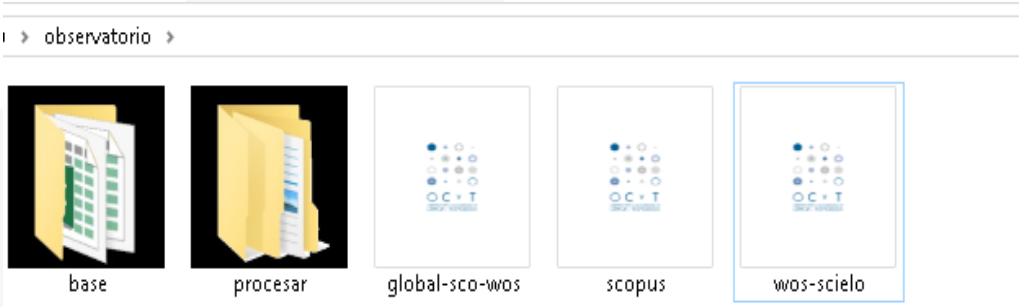
```
procesar
        |__Departamento
                    |__Ciudad/Municipio
                                |__archivo1.txt
                                |__archivo2.txt
                                |__archivo.csv
                                |__archivo.ris
```

Ejecución de aplicativo WoS-Scielo

1. Ingresar a la carpeta donde se alojan todos los archivos

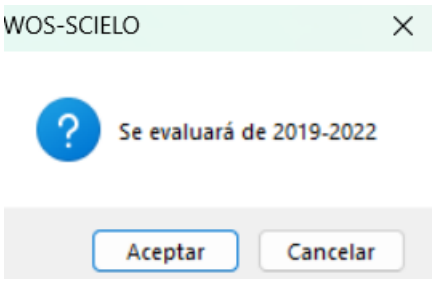
[6] Los archivos de extensión o formato .txt corresponden a WOS-SciELO y los archivos de extensión o formato .csv y .ris corresponden a Scopus

Gráfico 1. Ejemplo con carpeta donde se alojen las aplicaciones y los archivos requeridos



2. Dar doble clic en la aplicación wos-scielo. En este momento se presentará el cuadro de decisión que muestra los años de interés para realizar el procesamiento, le daremos clic en aceptar, como se muestra a continuación:

Gráfico 2. Ejemplo cuadro de decisión WOS-SCIELO



3. Al darle aceptar demora algunos segundos o minutos según la cantidad de información que se esté procesando
4. Al final de la operación en la carpeta resultados se obtienen las carpetas con los resultados que muestran los conteos para posteriormente trasladar esta información hacia los cuadros de salida y/o para que posteriormente con el aplicativo que une la información de las tres fuentes (WoS, Scopus y Scielo) se generen los indicadores respectivos.

Ejecución de aplicativo Scopus

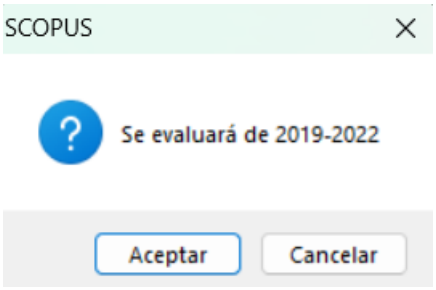
1. Ingresar a la carpeta donde se alojan todos los archivos

Gráfico 3. Ejemplo con carpeta donde se alojen las aplicaciones y los archivos requeridos



2. Dar doble clic en la aplicación Scopus. En este momento se presentará el cuadro de decisión que muestra los años de interés para realizar el procesamiento, le daremos clic en aceptar, como se muestra a continuación:

Gráfico 4. Ejemplo cuadro de decisión SCOPUS



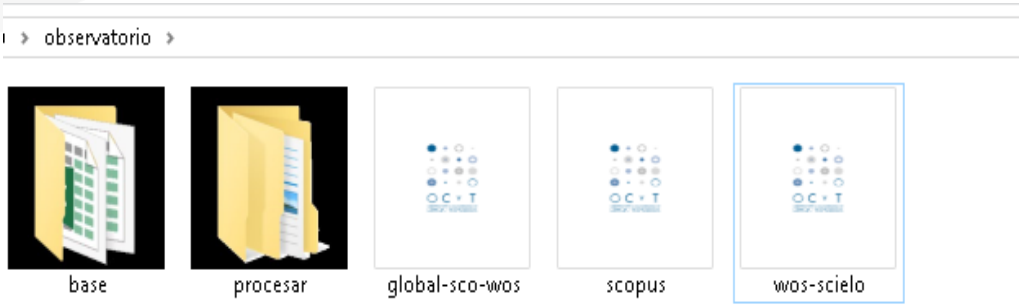
3. Al darle aceptar demora algunos segundos o minutos según la cantidad de información que se esté procesando
4. Al final de la operación en la carpeta resultados se obtienen las carpetas con los resultados que muestran los conteos para posteriormente trasladar esta información hacia los cuadros de salida y/o para que posteriormente

con el aplicativo que une la información de las tres fuentes (WoS, Scopus y Scielo) se generen los indicadores respectivos.

Ejecución de aplicativo global-sco-wos

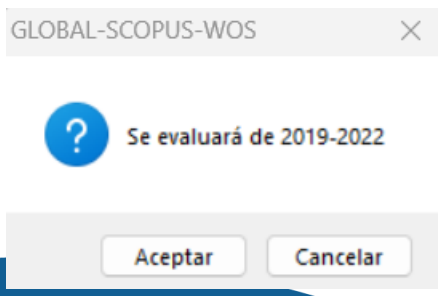
1. Ingresar a la carpeta donde se alojan todos los archivos

Gráfico 5. Ejemplo con carpeta donde se alojen las aplicaciones y los archivos requeridos



2. Verificar que anteriormente se hayan ejecutado las aplicaciones Scopus y wos-scielo dado que es obligatoria su ejecución previa para que este aplicativo funcione
3. Dar doble clic en la aplicación global-sco-wos. En este momento se presentará el cuadro de decisión que muestra los años de interés para realizar el procesamiento, le daremos clic en aceptar, como se muestra a continuación:

Gráfico 6. Ejemplo cuadro de decisión GLOBAL-SCOPUS-WOS



4. Al darle aceptar demora algunos segundos o minutos según la cantidad de información que se esté procesando
5. Al final de la operación en la carpeta resultados se obtienen las carpetas con los resultados que muestran los conteos para posteriormente trasladar esta información hacia los cuadros de salida y/o para que posteriormente con el aplicativo que une la información de las tres fuentes (WoS, Scopus y Scielo) se generen los indicadores respectivos
6. El último paso para el indicador de "*Distribución de la producción bibliográfica de autores vinculados a la institución por áreas de la ciencia*" se calcula agregando las categorías: Ciencias Agrícolas, Humanidades, Ciencias Naturales, Ciencias Sociales, Ingeniería y Tecnología, Ciencias médicas y Ciencias de la salud para "*Producción única por áreas OCDE*". Los conteos para "*Ciencias Multidisciplinarias*" y "*Desconocida*" ya se realizan automáticamente con el aplicativo.

Almacenamiento de los resultados

Una vez se cuenta con las estadísticas del operativo, se almacenan en un servidor en la nube tipo hosting, el cual se ubica en un datacenter del OCyT, en cual garantiza la protección de la información contra la perdida, alteración y confidencialidad, al contar con controles de acceso.

Igualmente, el OCyT ha establecido el respaldo de la información, de acuerdo el procedimiento de BackUP y respaldo de la información, los cuales son automáticos y manuales. Adicionalmente, por políticas de respaldo y políticas de seguridad y privacidad de la información de las estadísticas de Producción Científica Bibliométrica, se deben crear carpetas en OneDrive, de toda la información que se genera de cada proceso de producción de información.

6.4. Inconsistencias

Inconsistencias	Causal	Control
Error en la digitación del periodo de análisis	Digitación incorrecta en el archivo periodos para los periodos inicial y/o final de análisis	Como medida de calidad el aplicativo muestra una ventana emergente de confirmación para el periodo de procesamiento. Además si es incorrecto aparecerá error.
Errores en el nombramiento o ubicación de las carpetas con archivos a procesar	Error en digitación o ubicación de los datos a procesar	Verificación en la carpeta base de los archivos o carpetas para los cuales el aplicativo no arroja resultados
Error en ejecución aplicativo de unión global-sco-wos	Falta de ejecución de los aplicativos necesarios para la aplicación del global-sco-wos	Como medida de calidad el aplicativo muestra una ventana emergente Error: No hay información para procesar

6.5. Validaciones

Los métodos de validación en el proceso de procesamiento 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 previa del nombramiento y ubicación de archivos previo al procesamiento con los aplicativos. Igualmente el encargado de procesamiento en la ejecución interna del aplicativo verifica y valida con una muestra de datos que se esté ejecutando y generando los resultados adecuados respecto a la eliminación de duplicados[7] y la imputación de etiquetas de acuerdo con lo establecido para el procesamiento. Sin embargo y dado que puede ocurrir algún error humano se establecen medidas de calidad en los programas de tal forma que no se ejecuten o generen resultados sin la información correcta. Lo anterior se complementa con ventanas emergentes con mensajes de confirmación que indican el error en la ejecución.

[7] Los duplicados internamente el programa procede primero revisando título y luego DOI para la eliminación.

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 sean 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	Publicar los datos sobre producción científica

	datos sean relevantes para la toma de decisiones.	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 producción científica bibliométrica nacional.	Proporcionar datos desagregados para permitir un análisis detallado.
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