FICHA TÉCNICA: MEDIDAS DE CALIDAD DE LA OPERACIÓN ESTADÍSTICA PRODUCCIÓN CIENTÍFICA BIBLIOMÉTRICA EN SCOPUS, WEB OF SCIENCE Y SCIELO CITATION INDEX

Esta ficha técnica presenta las medidas de calidad aplicadas para la interpretación de la información difundida en el marco de la operación estadística 'Producción Científica Bibliográfica en Scopus, Web of Science Core Collection y SciELO Citation Index'. Los datos han sido integrados, depurados y desduplicados para consolidar un único lago de datos confiable. El objetivo es fortalecer los atributos de calidad estadística exigidos por la NTC PE 1000:2020, especialmente en lo que respecta a la interpretabilidad, transparencia y continuidad de las series estadísticas. A continuación, se presentan las reglas de validación y consistencia realizadas durante las fases de acopio y procesamiento de datos, y las medidas de calidad para el análisis e interpretación de resultados.

1. Reglas de validación y consistencia

1.1. Acopio de datos

Los métodos de validación en el proceso de acopio 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.

En adición el OCyT ha desarrollado una API para verificar los nombres de los municipios y departamentos según lo emite el DANE en su última versión, junto a los códigos asignados por departamentos. La aplicación realiza un cruce de información y verifica el correcto etiquetado de las carpetas para proceder a realizar el ajuste según corresponda.

1.2. Procesamiento

Estas reglas fueron aplicadas en la fase de acopio, sin embargo, durante el procesamiento se realiza una doble verificación, 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 archivo base del sistema llamado "DEPARTAMENTOS.xlsx". Igualmente se realiza el nombramiento de las áreas de conocimiento4 para los archivos descargados por subject área para Scopus de acuerdo con lo consignado en el archivo base "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 base "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.

2. Comités de expertos

El Observatorio Colombiano de Ciencia y Tecnología ha implementado un proceso estructurado de análisis por comités de expertos en tres momentos clave de la operación estadística de producción científica bibliográfica en Colombia. El primero se dedica al análisis de necesidades insatisfechas; el segundo se centra en el análisis de los resultados, realizado tanto por expertos internos para una primera evaluación, como por expertos externos, cuyo propósito es validar y contrastar los resultados obtenidos, garantizando una visión integral y rigurosa.

3. Medidas de calidad para la interpretación de indicadores

3.1 Producción bibliográfica por tipo de producto

Medida de calidad	Descripción aplicada al indicador
Fuente de datos	Scopus, WoS y SciELO, integradas en un único lago de datos depurado.

Cobertura	Abarca artículos, revisiones, editoriales, cartas, proceedings, etc.
Definición uniforme	Se considera 'artículo' a: Article; Review; Research Article; Review Article. Y los demás documentos ingresan a la categoría de "Otros documentos"
De duplicación	Se eliminaron registros duplicados entre bases para evitar sobreestimaciones.
Nota sobre series temporales	Se especifica que los últimos años son provisionales.
Interpretabilidad	Se desagrega por año y tipo documental para facilitar el análisis longitudinal.

3.2Tipo de autoría de la producción bibliográfica

Medida de calidad	Descripción aplicada al indicador
Categorización clara	Autoría única, coautoría nacional, coautoría internacional, según afiliación institucional.
Homogeneidad metodológica	Criterios de clasificación se mantienen iguales en todo el periodo.
Fuente confiable	Información extraída de metadatos en Scopus, WoS y SciELO.
Interpretabilidad	Presentado como porcentaje anual sobre el total.
Provisionalidad	Datos recientes pueden variar por actualización de metadatos.

3.3Distribución de la producción bibliográfica por área de conocimiento

Medida de calidad	Descripción aplicada al indicador
Clasificación estandarizada	Se emplea la clasificación de áreas OCDE para garantizar la comparabilidad internacional y la consistencia temática. Los documentos indexados en dos o más áreas de conocimiento se asignan a la categoría "Ciencias Multidisciplinarias", permitiendo

	una representación adecuada sin duplicidad en el conteo por área.
Tratamiento multidisciplinario	Documentos con varias áreas se agrupan como 'Ciencias Multidisciplinarias'.
Comparabilidad	Metodología constante entre 2013 y 2023.
Desagregación temática	Permite análisis por campo disciplinar.
Interpretabilidad	Datos en porcentajes por año con leyenda clara.

3.4 Promedio anual de citas de la producción bibliográfica

Medida de calidad	Descripción aplicada al indicador
Cálculo estandarizado	Promedio = total citaciones del año / total publicaciones del año.
Fuente de citaciones	Obtenidas directamente desde Scopus, WoS y SciELO.
Sesgo temporal explicado	Se advierte menor acumulación de citas en años recientes.
Comparabilidad	Mismo criterio desde 2013.
Interpretabilidad	Presentado por origen (base de datos) con aclaraciones.
Transparencia	Promedio puede variar si se actualizan datos en fuentes externas.

Anexo: Cumplimiento del requisito 11.6 de la NTC PE 1000:2020

En atención al requisito 11.6 de la NTC PE 1000:2020, relativo a la incorporación de notas explicativas en las series históricas publicadas, se resalta que esta operación estadística incluye elementos clave para garantizar la correcta interpretación de la información difundida:

- Los datos de los últimos años se identifican como provisionales y sujetos a ajustes.
- Se emplean definiciones estandarizadas y consistentes para clasificar documentos y tipos de autoría.

- Se aplicó una metodología de integración y de duplicación para consolidar un lago de datos único, garantizando transparencia.
- Las gráficas y tablas se acompañan de desagregaciones temáticas y temporales que refuerzan la interpretabilidad.
- Las notas explicativas verifican cada vez que se publican resultados (anualmente).

1. Ficha de aprobación

Elaborado por:	Revisado por:	Aprobado por:
Nombre: Alicia Herrera **Cargo:** Asistente de Investigación **Fecha**: 18/02/2025	**Nombre:** Efrén Romero Riaño **Cargo:** Líder temática Bibliometría **Fecha**: 20/02/2025	**Nombre:** Andrea Guevara Rey **Cargo:** Líder temática Inversión **Fecha**: 20/02/2025

2. Control de Cambios

FECHA	DESCRIPCIÓN DEL CAMBIO	VERSIÓN
20/02/2025	Lanzamiento de documento	1