

Análisis de la Producción científica sobre Big Data en la Base de Datos Scopus.(2014-2022).



Natalia Zambrano Venegas

Brayan Sneider Cruz Colmenares

Universidad Distrital Francisco José de Caldas

Facultad de Ciencia y Educación

Archivística y Gestión de la Información Digital

Bogotá D.C. 2023

Análisis de la Producción Científica sobre Big Data en la Base de Datos Scopus. (2014-2022).



Natalia Zambrano Venegas

Brayan Sneider Cruz Colmenares

Johann Pirela Morillo.

Director.

Modalidad de grado: Investigación-Innovación (Participación en el semillero YUARDI)

Universidad Distrital Francisco José de Caldas

Facultad de Ciencia y Educación

Archivística y Gestión de la Información Digital

Bogotá D.C. 2023

## Agradecimientos

*A mi familia que me apoyaron en todo este proceso, con ayuda de ellos estoy logrando mis sueños de ser archivista. Gracias por estar pendientes de mis estudios por orientarme para tomar mejores decisiones y por eso este logro es de todos, resaltando el apoyo de mi madre Andrea Zambrano que sacando adelante a sus hijos puede decir que la mayor se va a graduar como Archivista. Al profesor Johan Pirela que nos asesoró en este trabajo y nos enseñó la importancia de la investigación para nuestra área de trabajo. A la Universidad Francisco José de Caldas en donde me dieron la oportunidad de estudiar una carrera universitaria.*

*A mis padres Carlos Cruz y Elisabeth Colmenares por todo el apoyo que me dieron a lo largo de la carrera, por los consejos que me ayudaron a lograr mi título profesional, gracias por enseñarme los valores que me ayudaron a ser la persona que soy ahora, a los profesores del proyecto de archivística por todas las enseñanzas a lo largo de la carrera sobre todo al profesor Johan Pirela quien nos apoyó en todo momento para realizar el trabajo de grado, gracias por su asesoría y el tiempo que nos brindó. Por último, a la universidad Distrital por brindarme la oportunidad de estudiar y obtener mi título profesional.*

## Dedicatoria

*A mi familia por ser mi apoyo, principalmente a mi madre, hermano y a mi gato Orión que me acompañó hasta altas horas de la noche para lograr esta meta.*

*Natalia Zambrano Venegas*

*Este trabajo de grado lo dedico a mis padres, a mi hermana y mi novia por el apoyo que me dieron a lo largo de mi carrera, ellos son mi motivación para obtener mi título profesional.*

*Brayan Sneider Cruz Colmenares*

## Tabla de Contenido

<b>Agradecimientos</b>	<b>3</b>
<b>Dedicatoria</b>	<b>4</b>
<b>Resumen</b>	<b>6</b>
<b>Abstract</b>	<b>7</b>
<b>Introducción</b>	<b>8</b>
<b>Capítulo I. Problema de Investigación</b>	<b>10</b>
1.1 Planteamiento del problema.	10
1.2 Justificación	14
1.3. Objetivos	16
1.3.1. Objetivo General	16
1.3.2. Objetivos Específicos	16
<b>Capítulo II. Marco Teórico y Estado del Arte</b>	<b>17</b>
2.1. Antecedentes.	17
2.2 Estado del Arte	23
2.3 Categorías Conceptuales.	28
2.3.1. Producción Científica	28
2.3.2. Estudios Métricos	29
2.3.3. Bibliometría.	30
2.3.5. Big Data.	31
<b>Capítulo III. Metodología de Investigación.</b>	<b>33</b>
3.1 Enfoque	33
3.2 Tipo de Investigación	34
3.3 Método de Investigación	35
3.4 Técnicas	36
3.5 Instrumentos	37
3.6 Fases	39
<b>Capítulo IV. Resultados.</b>	<b>41</b>
4.1 Documentos por Autores	41
4.2 Documentos producidos por país	43
4.3 Producción científica a través de fuentes	44
4.4 Documentos por año.	45
4.5 Documentos por Afiliación.	46
4.6 Documentos por Tipo.	47
4.7 Documentos por Área Temática.	48
4.8 Documentos por patrocinador de financiación.	49
<b>CONCLUSIONES.</b>	<b>51</b>
<b>RECOMENDACIONES.</b>	<b>53</b>
<b>REFERENCIAS</b>	<b>55</b>

## Resumen

En la actualidad el Big data ha tomado más relevancia para el análisis de información en grandes volúmenes, es por esto que mediante la base de datos Scopus, se analizó la producción científica sobre este campo que constituye un conjunto de herramientas y técnicas para la identificación de patrones de regularidad de los datos que se producen en una organización. En el semillero YUARDI en cual está orientado a la reflexión y el análisis de la información relacionado con diferentes tecnologías empleadas en la archivística como en caso el Big Data En este sentido, el objetivo general de este trabajo es analizar la producción científica sobre Big Data en la base de datos Scopus (2018-2022). Para el desarrollo del trabajo se estableció que los autores Ernesto Spinak, Tomas Gorriz, Jairo Marquez Diaz, trabajaron en sus artículos y trabajos de investigación todo el tema de la bibliometría como instrumento para medir el desarrollo de la información científica sobre Big data. La metodología consistió en un estudio basado en indicadores bibliométricos, tales como: autor con mayor producción, países, producción a través de fuentes, por año, por afiliación por tipo por área temática y por patrocinador financiero entre términos esenciales relacionados con el big data. Dichos indicadores fueron considerados en la revisión de las bases de datos, que no solo permiten definir la cuantificación de las publicaciones, a partir de las mismas se obtendrán gráficas que permitan realizar un análisis más a fondo de los aspectos que intervienen en la publicación de información, ya sea económico, político o social.

Palabras claves: Bibliometría, producción Científica Big Data, análisis masivo de datos

## Abstract

Currently, Big data is becoming more relevant for the analysis of information in large volumes, which is why through the Scopus database. The scientific production on this field will be analyzed, which constitutes a set of tools and techniques for the identification of patterns of regularity of the data produced in an organization. In the YUARDI seedbed, which is oriented towards reflection and analysis of information related to different technologies used in archiving, such as Big Data.. In this sense, the general objective of this work is to analyze the scientific production on Big Data in the Scopus database (2018-2022). In this degree work we rely on authors such as Ernesto Spinak, Tomas Gorriz, Jairo Marquez Diaz, who in their articles and research papers cover the entire subject of bibliometrics as an instrument to measure the development of scientific information on Big data. The methodology consisted of a study based on bibliometric indicators, such as: author with the highest production, countries, production through sources, per year, by affiliation by type by subject area and by financial sponsor among essential terms related to big data. These indicators were considered in the review of the databases, which not only allow defining the quantification of the publications, but from them graphs will be obtained that allow a more in-depth analysis of the aspects involved in the publication of information. be it economic, political or social.

Key words: Bibliometrics, Big data, Scientific production, information

## Introducción

El Big Data es el análisis de grandes volúmenes de datos lo cual contribuye a lograr los objetivos estratégicos de un proyecto. Por medio del Big Data y su adecuada implementación se garantiza el éxito de proyectos de investigación aplicable a cualquier campo de investigación. Los estudios métricos son herramientas por las cuales se determina la cantidad de producción científica referente a un tema en particular teniendo en cuenta diferentes características de la información publicada. En la actualidad, no se encuentra un estudio métrico relacionado al Big Data, por ende, este trabajo de grado se centra en la base de datos Scopus.

La importancia del Big Data mundial crea la necesidad de la implementación de los estudios métricos para tener claridad relacionada con la producción académica del tema. Por medio de la caracterización de la base de datos, resaltando su campo social y herramientas que facilitan la apropiación de los estudios métricos, en este caso relacionados al Big Data. Se establecen indicadores bibliométricos como los autores, países, a través de fuentes, por año, por afiliación, por tipo, por área temática y por patrocinador, en donde se busca identificar las cualidades de la producción científica del análisis masivo de datos.

En este sentido el presente trabajo se estructura de la siguiente manera en el Capítulo I. Problema de Investigación, se encuentra el planteamiento del problema en donde se visualiza el contexto nacional e internacional del Big Data, en donde se encuentra la necesidad de mirar la producción académica de este tema y desarrollar la importación de este para la archivística. En la justificación del problema se establece la relación del Big Data con la Archivística y donde se registra que se pueden utilizar las ciencias métricas por medio de indicadores para realizar un



análisis de los mismos. El objetivo general es Analizar la producción científica sobre Big Data en la base de datos Scopus (2014-2022).

En el Capítulo II. Marco teórico y estado del arte, se desarrollan las referencias en donde se toma como base la investigación en los repositorios de las universidades de la Salle, Javeriana y Antioquia. En el estado del arte se desarrollan los principales autores relacionados con el Big Data, la bibliometría y la archivística. En las categorías conceptuales se desarrollan la producción académica, los estudios métricos, bibliometría, indicadores bibliométricos y Big Data.

En el Capítulo III. Metodología de investigación, se desarrolla el enfoque cualitativo, por medio del tipo de investigación de la descripción analítica, desde donde se desarrolla un análisis de la información. Utilizado como método de investigación la investigación documental teniendo en cuenta que se despliega en bases de datos en donde se recopiló información por medio de las técnicas de bibliometría y utilizando formatos de registro de información.

En el Capítulo IV. Resultados se encuentran ilustraciones en donde se puede visualizar los indicadores bibliométricos y cómo se adaptan al Big Data. El análisis de datos y de los indicadores bibliométricos nos ayuda a determinar las características propias del tema y conclusiones teniendo como base factores y criterios claros para este estudio de la producción académica de la Big Data.

## Capítulo I. Problema de Investigación

En este capítulo se muestra la caracterización inicial del objeto de estudio que se contempla en el presente ejercicio investigativo, el cual se basa en la realización de un estudio métrico para analizar la producción académica que se encuentra en la base de datos: Scopus, referente al tema del Big Data. Se resalta la importancia del Big Data, de los estudios métricos y la aplicación de indicadores bibliométricos a las bases de datos ya mencionadas. Las características propias de cada base de datos crean un ambiente óptimo para realizar este proyecto de investigación.

### **1.1 Planteamiento del problema.**

La bibliometría ayuda a determinar las características y la cantidad de producción académica que se tiene en un tema en especial en este caso referente a Big Data, que se concibe como un conjunto de técnicas que basadas en el análisis masivo de datos que permiten proyectar el comportamiento de escenarios, realizando análisis predictivo y de diferentes alcances y niveles. Con los estudios métricos de información se ofrece información para asegurar contextos claros acerca de la producción científica sobre diversos campos y áreas. En el caso particular de este trabajo, el foco será el reconocimiento de indicadores bibliométricos específicos, desde los cuales será posible obtener las principales tendencias temáticas relacionadas con el Big Data.

Los estudios métricos se apoyan en el uso de gráficas para determinar una mayor visualización y comprensión de la información. Se debe tener en cuenta que este proceso se puede realizar para cualquier tema de investigación, lo que contribuye al manejo de información y da un contexto claro en donde se pueden basar los investigadores para realizar sus producciones académicas. En este caso se va a desarrollar el análisis bibliométrico basado en Big Data, en

donde los resultados van a brindar una orientación sobre los volúmenes de producción académica sobre un tema con suma importancia a nivel internacional, Latinoamérica y nacional.

En la actualidad, se presenta un gran creciente de la producción de información lo que genera la necesidad de la implementación del Big Data para la gestión de la información.

Galiana (2019), plantea que el Big Data es el conjunto de datos, en su mayoría extensos, según sus características técnicas de variedad, velocidad, variabilidad, veracidad y volumen . Los procesos de gestión documental en donde se evidencia con más cabalidad la integración del Big Data y la gestión de información es por medio de la minería de datos.

Se han publicado diferentes normas ISO orientadas hacia la arquitectura de referencia de Big Data, relacionado también con la tecnología de la información y la estadística. Las principales son:

- ISO/IEC DIS 20546 Tecnología de la información- Big Data- Resumen y vocabulario.
- ISO/IEC AWI TR 20547-1 Tecnología de la información – Arquitectura de referencia de Big Data – Parte 1: Marco y proceso de solicitud.
- ISO/IEC TR 20547-3 Tecnología de la información – Arquitectura de referencia de Big Data -Parte 2: Casos de uso y requisitos derivados.
- ISO/IEC DIS 20547-3 Tecnología de la información -Arquitectura de referencia de Big Data – Parte 3: Arquitectura de referencia.
- ISO/IEC AWI 20547-4 Tecnología de la información – Arquitectura de referencia de Big Data – Parte 4: Estructura de seguridad y privacidad.
- ISO/IEC TR 20547-5: 2018 Tecnología de la información – Arquitectura de referencia de Big Data – Parte 5: Hoja de ruta de estándares.

La importancia del Big Data para las organizaciones es tan amplia que se llevo a cabo el 9 y 10 de diciembre del 2021, “el encuentro Mundial Big Data ha contado con los principales expertos mundiales de Nuevas Tecnologías o tecnologías exponenciales que ofrecen casos de uso, mejores prácticas, desafíos a enfrentar, éxito y lecciones que aprender” ( “6 TO EMBDATA”, 2021)

En Latinoamérica se presentan problemáticas como falta de interés en el desarrollo del Big Data y esto ocasiona que no hallan expertos en el tema, a diferencia que a nivel mundial es una herramienta eficaz para los negocios y se han comenzado a invertir recursos en esta área. Uno de los ejemplos más claros es la falta de interés por medio de las entidades públicas para invertir en procesos relacionados con este tema, dejando al lado esta oportunidad de negocio que puede beneficiar a las entidades del estado. La necesidad de implementar el Big Data en las entidades públicas también está orientada a la atención al ciudadano y para estudios estadísticos, con los cuales se puede establecer políticas públicas.

En Colombia, universidades como la Universidad Externado de Colombia (2021),

señalan que la normativa actual es insuficiente para garantizar las operaciones de Big Data con respecto a la información pública el cual es un gran reto teniendo en cuenta que: Solo el 3,3% de las entidades [oficiales de Colombia] cuenta con las condiciones para avanzar a la implementación de Big Data y el aprovechamiento de los datos.

El 96,7% restante requieren de intervención específicas para afrontar los retos actuales de generación de valor social y económico con la explotación de datos digitales”(p. 1)

La importancia que tiene el Big Data para las organizaciones, los investigadores y los interesados radica que a través de sus técnicas es posible proyectar patrones de ocurrencia de hechos, procesos, fenómenos, actividades y estudios. En este sentido, desarrollar un análisis de la producción científica sobre este tema constituye un ejercicio que permite reconstruir tendencias, autores, países de mayor de mayor producción y las co-ocurrencias con términos, alrededor de los cuales gira la información científica que se produce, para lo cual se recurrió a la realización de este ejercicio utilizando cuatro bases de datos Scopus, las cuales tienen características particulares.

Scopus es una base de datos de referencias bibliográficas la cual tiene herramientas que ayudan al análisis y visualización de la investigación.. Esta base de datos “proporciona una visión general completa de la producción de investigación en todo el mundo en los campos de la ciencia, la tecnología, la medicina, las ciencias sociales, las artes y las humanidades” (Scopus, 2022)

Estas bases de datos además de ser una gran alternativa de búsqueda también brindan herramientas para el desarrollo de estudio métricos, a partir del análisis de datos y porque tienen un amplio nivel de información. Se enfocan en áreas de conocimiento similares y en la que se apoya qué es la ciencia social en donde se encuentran temas relacionados con Big Data, adicionalmente que ofrecen herramientas para realizar el proceso bibliométrico y de colaboración para los resultados de esta investigación.

Se busca por medio de la bibliometría hacer un análisis bibliométrico que por medio de gráficas o tablas se puede representar las características que tiene la producción académica sobre Big Data, estableciendo opciones claras para su mejor comprensión. Los campos de investigación que se encuentra en la base de datos Scopus, teniendo en cuenta su alto nivel de producción en áreas sociales y las herramientas que ofrecen para el análisis de la producción académica.

Reconociendo la importancia del Big Data para las organizaciones actuales, las producciones científicas y el análisis bibliométrico de la literatura del Big Data se centra hacia sectores específicos como el financiero, turístico, empresarial, entre otros. No se encuentra un análisis bibliométrico de la producción académica a nivel general. La bibliometría permite visualizar desde el punto de vista cuantitativo la producción académica según país, autor, idioma y la cantidad de producción en bases de datos confiables. Por medio de los resultados que se obtienen mediante esta clasificación de producción científica podemos evidenciar cuales son los países con mayor producción sus enfoques y años de producción. Lo que lleva a preguntarnos **¿Qué producción científica se encuentra disponible sobre Big Data en la base de datos Scopus (2014-2022)**

## **1.2 Justificación**

El Big Data es el manejo de grandes volúmenes de información para la creación de conocimiento para una organización, es un tema muy extenso; si bien, estos conocimientos y prácticas son de suma importancia para el campo de la Archivística y Gestión de la Información Digital, son escasos los estudios métricos que aporte datos sobre las tendencias en producción intelectual. Por esta razón se ve la necesidad de analizar la producción científica, por medio de la bibliometría, para determinar la producción por país, autores, años, entre otros se determinan las características más comunes sobre la producción científica.

A partir de los resultados producto del análisis bibliométricos, fue posible determinar la situación de los países referente a la producción científica y la Big Data, esto evidencia su compromiso con los grupos de investigación y a su vez la relevancia que se le da a la educación universitaria y científica ya que en muchos países especialmente en Latinoamérica no se hace la

inversión adecuada de los presupuestos públicos para generar una mayor producción científica (León, et al. 2020). A partir de la Big data y los grandes volúmenes de información que se manejan en base de datos como Scopus, las cuales se tomaron como herramienta que brindan la producción científica en diferentes aspectos como: países y autores, entre otros, lo que permite definir esta información como referencia para medir la distribución de la producción científica en diferentes áreas o disciplinas en las que se aplican.

Con este ejercicio de investigación, que se deriva del Semillero YUARDI, del Proyecto Curricular en Archivística y Gestión de la Información Digital, de la Universidad Distrital “Francisco José de Caldas”, se analizó la producción científica de Colombia comparada con las de otros países, así se concluye o se da a entender la importancia del estudio científico en las universidades del país, esto con el fin de demostrar si hay un déficit, así como la caracterización de la producción científica. También se pretende demostrar, por medio de estos autores, cuáles disciplinas y áreas de conocimiento muestra mayores niveles de producción están investigando de formas más amplia la Big Data y cuales son en las que menos se están invirtiendo conocimientos para este tema.

Esto permite dar a conocer como la Big Data es un eje primordial para el análisis de la información en grandes volúmenes ya que demuestra a través de ella las tendencias globales, teniendo en cuenta que vivimos en la era de la información digital y esto permitirá conocer a qué temas se están aplicando la Big Data, teniendo en cuenta que en Colombia no se conoce mucho sobre este aspecto como es el caso de las entidades de carácter público que no lo tienen en cuenta para reconocer problemáticas a nivel social, económico, político entre otros aspectos que afectan a la sociedad.

Este estudio corresponde a un inicio para plantear y estudiar la bibliometría más a fondo, relacionándolo con otras metodologías de medición de datos en torno a la generación de conocimiento. Y corresponde a archivistas y gestores de la información digital tener en cuenta estos métodos de medición de información para así mismo aplicarlo a entidades públicas o privadas que trabajen con grandes cantidades de información y evidenciar las necesidades en las que se deben mejorar.

### **1.3. Objetivos**

#### 1.3.1. Objetivo General

Analizar la producción científica sobre Big Data en la base de datos Scopus (2014-2022)

#### 1.3.2. Objetivos Específicos

- Identificar los autores, países, años y tipos de mayor producción científica sobre BIG DATA.
- Determinar las producciones a través de fuentes, por afiliación, por área temática y por patrocinador financiero que mayormente se publica sobre BIG DATA.
- Establecer el balance de la producción científica sobre BIG DATA en la fuente de datos seleccionada.



## Capítulo II. Marco Teórico y Estado del Arte

El análisis bibliométrico se puede aplicar en diferentes áreas del conocimiento, lo que genera una gran producción académica. En Colombia, distintas universidades tienen en sus repositorios trabajos de grados con aplicación de los estudios métricas entre estos la bibliometría, aunque son orientados a distintos temas se puede encontrar algunos asociados. Se desarrollan temáticas como los datos abiertos, ciencias de la información, la participación de las universidades para la visualización de las producciones científicas y las disciplinas asociadas a las soluciones de problemáticas en las ciencias de la información.

### **2.1. Antecedentes.**

Esta investigación se desarrolla a partir de la exploración en los repositorios de la Universidad de Antioquia, Universidad Salle y Pontificia Universidad Javeriana, con variables como la implementación de estudios métricos en la producción científicas orientada en áreas de la información, los cuales se desarrollan en centros de investigación, ciencia abierta, ciencias de la información y revistas.

En el repositorio de la Universidad de Antioquia se encuentra la investigación elaborada por Juan Carlos Cardenas Isaza titulada: “Visibilidad de la Producción Científica y Académica de los Grupos de Investigación Adscritos al Centro de Investigaciones Sociales y Humada (CISH), desde la Ruta Verde del Repositorio Institucional de la Universidad de Antioquia”, elaborada en 2019. El objetivo es visualizar la producción científica y académica de los Grupos de Investigación adscritos al Centro de Investigaciones Sociales y Humanas desde la ruta verde del repositorio institucional de la Universidad de Antioquia. En la metodología de investigación se desarrolla un enfoque mixto, utiliza como técnicas la recolección de información y los productos

científicos y académicos y tiene tres fases las cuales son diagnóstico, organización y tratamiento de los productos académicos y científicos del CISH y la caracterización.

Los resultados se desarrollaron por cada una de las fases en donde el diagnóstico se presenta la información que se tienen en el CISH, se divide por los grupos que conforman el centro, se reconocieron tipos como: productor de conocimiento los artículos de nuevo conocimiento y divulgación que se encuentran publicados. Se visualiza su disponibilidad en el repositorio y el acceso a los mismos. Se puede observar en la conclusión de Cárdenas (2019), “Se recomienda continuar trabajando en el proceso de organización y tratamiento para contribuir a la renovación y también al ascenso de la categoría que tiene los grupos de investigación en el sistema de clasificación de Colciencias” (p.31)

El trabajo anterior se orienta en los indicadores bibliométricos, una de las conclusiones señalan la importancia de la aplicación de las buenas prácticas archivísticas para garantizar el acceso del conocimiento en este caso en el CISH. Las ilustraciones brindan una base para armonizar las ilustraciones con el texto.

En nuestra búsqueda de antecedentes una de la fuente fue el repositorio de la Universidad Salle, donde se encontró un trabajo de Yeimmy Rocio Cárdenas Reyes (2020) titulado: “Análisis bibliométrico aplicado a la Ciencia abierta 2010-2019” . El cual tiene como objetivo principal identificar el comportamiento de la producción señalado aplicando conceptos obtenidos a través de la academia. La metodología utilizada en este trabajo consistió en un enfoque cuantitativo, utilizando medidas estadísticas por medio del análisis de la información recolectada, con un método de investigación documental, desarrollando la técnica de la bibliometría estableciendo indicadores de actividad e indicadores de impacto, adicional las fases van relacionadas con los objetivos específicos del trabajo.

Las fases se centran en el análisis del comportamiento temático de la ciencia abierta en las bases seleccionadas y el determinar los indicadores bibliométricos, estos son fueron, para medir la producción científica, autores más relevantes, producción por país, temáticas de los documentos, recursos más importantes, afiliaciones, factor de impacto patrones de colaboración y altmetrics. Generando resultados claros representados por tablas e ilustraciones para mejor comprensión.

Uno de los resultados más interesantes corresponde al crecimiento en los últimos años sobre la investigación referente al tema de los datos abiertos, adicionalmente uno de los autores más representativos es Bockelman B, el cual tiene gran cantidad de artículos publicados en la base de datos Scopus. En la producción por país se presentan los tres países con mayor producción los cuales son Estados Unidos, Alemania y el Reino Unido, en cuanto al área del conocimiento se visualiza las ciencias computacionales y las ciencias sociales. Por cada uno de los indicadores se encuentra su ilustración y el análisis de cada una, lo que contribuye a la creación de conclusiones.

La conclusión a la que llega la autora desarrolla la importancia de la bibliometría para la toma de decisiones, “de acuerdo a dicho comportamiento se espera siga creciendo la producción científica en este campo de estudio y además adicionado nuevos recursos y herramientas de libre acceso, más allá de las publicaciones científicas indexadas, también como alienación a la filosofía propia de la Ciencia Abierta” (Cárdenas Reyes, 2020, p. 74). También señala los resultados más importantes de este trabajo. Este trabajo tiene una amplia contribución con la presente investigación porque se desarrolla un tema relacionado con la archivística el cual es la ciencia abierta y el uso de los indicadores bibliométricos tales como: producción científica, autores más

relevantes y producción por país, los cuales también se van desarrollar el primer objetivo específico.

En el repositorio de la Pontificia Universidad Javeriana se encontraron dos trabajos el primero por Johana Marcela Ladino Lemus, titulado “Evaluación Bibliométrica y Temática de la Revista Colombiana Internacional 2011-2020” elaborado en el 2021, el objetivo correspondía a realizar un análisis sobre la producción científica a la Revista Colombiana Internacional 2011-2019, para identificar aspectos significativos que contribuyan a la toma de decisiones y el apoyo a los investigadores. En la metodología se encuentra un estudio que señala la aplicación de indicadores bibliométricos los cuales fueron productividad, citación, colaboración y redes. En las áreas temáticas se desarrollaron categorías historia, ciencias políticas y ciencias sociología y ciencias políticas. Por medio de esta metodología, se visualizaron resultados que se ilustran por medio de imágenes y tablas, con su respectivo análisis.

En los resultados se puede evidenciar que el autor con mayor producción científica en la Revista es Rettberg A, también se visualiza que el idioma con mayor publicación es español, le sigue el inglés y por último el portugués. Se presenta un análisis de frecuencia de citación, en las temáticas que se presentó “mayor incidencia a la citación están enfocadas en temáticas como políticas de carácter regional y latinoamericano, seguido de conflicto, posconflicto y conflictos socioambientales” (Ladino, 2021, p. 82). Estos resultados contribuyen a la creación de conclusiones, orientadas hacia la importancia de la bibliometría, idioma, países (resaltado la colaboración entre países) y la frecuencia de citación con las temáticas correspondientes.

Este estudio orienta sobre la aplicación del indicador de citación teniendo en cuenta que se pueden evidenciar países con mayor frecuencia de citación y las temáticas de citación. La

colaboración en la producción científica en donde se contempla la participación de países como: Colombia, Argentina, Brasil, España, México y Estados Unidos.

En este trabajo se puede evidenciar una adecuada aplicación de indicadores como el idioma y países. La articulación de los indicadores con la temática en este caso la gestión de la información en el ámbito periodístico. Es un claro ejemplo de adopción de las ciencias bibliométricas, exactamente con los indicadores de idioma y países, en la producción científica relacionada con el periodismo.

El otro trabajo que se encontró en el repositorio de Pontificia Universidad Javeriana, corresponde al desarrollado por Julián Felipe Ardila Pérez, titulado “Análisis Bibliométrico de la Visibilidad de la Producción Científica de América Latina en Ciencia de la Información Propuesta para revisar la Interdisciplinariedad”, elaborado en el 2021, el objetivo hacía énfasis en el Desarrollar un análisis bibliométrico de la producción de América Latina en revistas científicas indexadas en el área temática (subject category) de ciencias de la información y bibliotecología por medio de la library and information Science que se encuentra en la base de datos Web of Science se busca valorar la interdisciplinariedad de esta producción.

La metodología utilizada en este trabajo se realiza con el enfoque mixto teniendo en cuenta que se utiliza un método bibliométrico desde un punto de vista cuantitativo. Se desarrollan tres fases las cuales son identificación, análisis interdisciplinariedad por difusión y análisis interdisciplinariedad por los documentos citados. Los indicadores bibliométricos son los países, años y citas, los cuales están representados por medio de ilustraciones y tablas dinámicas con su respectivo análisis.

Los resultados están orientados a la producción científica de América Latina en Ciencias de la Información por idioma de publicación, en donde se evidencia que hay mayor producción en el idioma portugués, inglés y español. En cuanto a los documentos más citados se evidencian temáticas de cadenas de suministros, logística, comunicación, medios sociales, política, tecnología de la información y salud pública. En cuanto a los autores, en donde se resalta en la cantidad de citas por autor se concluyó que el autor con mayor citación es Maciel M. Queiroz, lo que refuerza el resultado de concentración de contribuciones ya que es Brasil es país con mayor participación en la producción de documentos. Estos resultados contribuyeron a la creación de conclusiones orientadas hacia la práctica de citación.

Este proyecto señala el panorama de la ciencia de la información relacionada con la producción científica. Al tener en cuenta que diferentes disciplinas se emplean para solucionar problemáticas que se pueden generar en las ciencias de la información, se comienza a visualizar la importancia de este análisis “Específicamente en la Ciencias de la Información, se ha encontrado que los aportes de investigadores pertenecientes a diferentes disciplinas y su colaboración permite un avance significativo en el campo” (Ardila Pérez, 2021).

En los antecedentes encontrados no se evidencia uno que trate sobre el Big Data, pero sí asociados a la biometría en el área de la información, su importancia para los interesados, las interdisciplinas asociadas y la difusión de la producción científica.

Se puede evidenciar que no se encuentran trabajos de grados que relacionen los estudios métricos sobre la producción científica relacionada con el Big Data, pero sí estudios métricos orientados a temas de la gestión de la información como los datos abiertos. Por medio de autores

reconocidos en el campo de los estudios métricos y del Big Data comenzamos a visualizar la importancia, en los últimos años, en las redes sociales y con temas relacionados al Covid- 19.

## **2.2 Estado del Arte**

En diferentes universidades se han llevado a cabo investigaciones relacionadas a la Big data el cual se le da un enfoque cuantitativo en los que se analizan las bases de datos como Scopus y Google Scholar, estos estudios miden los países en los que más se consulta la tecnología de big data colocando a China, Estados Unidos y Japón en los países donde más se consulta este tipo de información, donde lo relacionan más que todo a las redes sociales y los usos que se le da a la información masiva que recogen actualmente. En estos estudios se analizan las condiciones tecnológicas con las que cuenta cada país debido a que es un factor importante a la hora de consultar estos temas, esto demuestra que muchos países en vía de desarrollo no cuentan con las condiciones necesarias para dar a conocer la Big Data, lo que causa que este tema no tenga la relevancia que se le debe otorgar..

Entre 2010 y 2020 como menciona Hernández et al. (2021) la big data ha tomado un papel importante en empresas debido al aumento de usuarios en las redes sociales y el acceso a los celulares los cuales son canales fundamentales de información en masa y en tiempo real. Muchas personas al ver el potencial de la nube suben sus datos a este gran conjunto de datos, aunque este tema también tiene una preocupación en cuanto a la seguridad o respaldo que tenga la información que se sube ya que el riesgo de que sea eliminada siempre va a estar presente.

La bibliometría ha tenido un carácter multidisciplinar ya que se ha usado como metodología de evaluación y medición de esta producción científica y académica para diversas disciplinas. En el caso de la producción científica se plantea el problema de si esta información

publicada tiene un nivel de relevancia realmente bueno o es de alto impacto en el conocimiento o por el contrario pasa desapercibido y es considerado un esfuerzo en vano que no aporta al progreso ya sea académico, social, económico o político entre otras.

La importancia del big data en la medicina como lo menciona Wu (2021), en China se recurrió a la big data como forma de parar el contagio del Covid a través de la información que se recopila de cada ciudadano para así a través de los celulares saber quién está contagiado o ha tenido contacto con el virus, este sistema realmente ayudó a disminuir los contagios y tener un mejor control sobre el virus.

A través de la big data se recopilaban datos de información como la ubicación de la persona, y las personas con las que se frecuentaba, aunque también se dio el debate si realmente la big data es una violación a la privacidad es por eso que muchas empresas piden autorización de uso de datos para evitar este tipo de problemas como lo afirma Marquez (2020)

Debido a estos aportes tan importantes para la sociedad se han analizado otros métodos de recolección de información. Las investigaciones bibliométricas constituyen una de las tendencias más relevantes de la Ciencia en su concepción más empírica, la Cienciometría; puesto que las publicaciones científicas ofrecen gran cantidad de datos que son susceptibles de cuantificación y análisis para el estudio de la ciencia (Romera, 1992).

Por lo tanto “La bibliometría como conocimiento teórico es la caracterización cuantitativa de las propiedades del discurso impreso, entendiendo por caracterización cuantitativa, la exposición de ideas probabilísticamente verdaderas acerca de los fenómenos seleccionado” (Schrader, 1981, p. 2). Es decir que a través de los datos podemos plantear distintos escenarios



frente a tendencias y patrones en el cual se analizan las soluciones o se trazan los objetivos frente al tema.

Razón por la cual la Spinak (1996), menciona que la bibliometría es un método de investigación documental en el cual se aplican un grupo de técnicas donde no solo se toman en cuenta citas y publicaciones, también se toman en cuenta diversos modelos matemáticos a los cuales se les aporta diferentes técnicas de análisis.

La bibliometría enfocada a ciencia ha surgido como un producto del contacto interdisciplinar, como mencionó López Piñeros (como citó Sanz, 2022), el desarrollo ha dependido de dos fuentes distintas, aunque con numerosas relaciones. La primera es la llamada 'ciencia de la Ciencia' y tiene un carácter fundamentalmente teórico. La segunda, de significación casi puramente técnica o aplicada, corresponde al gigantesco desarrollo que últimamente ha adquirido la documentación científica. Debido a estos conceptos los indicadores bibliométricos han tomado relevancia en distintos campos de la investigación.

Las investigaciones bibliométricas se encuentran divididas en dos partes, la investigación descriptiva y las investigaciones que utilizan el análisis de citas los cuales se generan a través de datos básicos que se utilizan como referencias bibliográficas. En cuanto a la investigación descriptiva se ha centrado principalmente en dos parámetros el personal científico y las publicaciones científicas en los cuales se mide la producción científica referente a un área en específico como en este caso se va a realizar con el big data. Es por lo que diferentes autores han realizado un análisis del crecimiento exponencial de la información científica.

La bibliometría como lo menciona Gorriz y Castera (2018) es fundamental para el análisis y evaluación de la información científica, es imprescindible para programas académicos ya que

constituyen herramientas clave para la gestión de procesos tecnológicos y así mismo poder tomar decisiones estratégicas, como se evidencia la bibliometría como ciencia se aplica principalmente a dos ámbitos que son el desarrollo y su aplicación, su implementación puede traer grandes aportes a la ciencia.

Teniendo en cuenta que como mencionó Price (como cito Romera, 1992) “ La ciencia crece a interés compuesto, multiplicándose por una cantidad determinada en iguales períodos de tiempo. Matemáticamente, la ley del crecimiento exponencial resulta de la sencilla condición de que en cualquier momento la tasa de crecimiento sea proporcional al tamaño de la población o magnitud total adquirida” (pag. 464)

En cuanto a marketing, la Big data en los últimos años ha generado un gran interés en diferentes campos de la sociedad, no solo en la academia, también las empresas que utilizan la información en masa para identificar las necesidades del usuario o sus preferencias. Con base en los estudios de la Big data surge la necesidad de contar con profesionales que entiendan sobre el mismo, y puedan procesar toda esta información que se genera ya sea a través de bases de datos o plataformas que recolectan información.

Es interesante como presenta Kotler & Keller (como cito Goyzueta, 2015) el término *sistema de inteligencia de marketing* atribuyéndole como el conjunto de procedimientos y fuentes que utilizan los gerentes para obtener información diaria sobre las novedades que se dan en el entorno del marketing. Ambos autores postulan que el sistema de registros internos proporciona datos de resultados, pero el sistema de inteligencia de marketing proporciona datos de acontecimientos. Este último concepto será útil a la hora de explicar el proceso que puede alcanzarse con el Big Data-Marketing. En el área empresarial han surgido diferentes conceptos como el Business Analytics en el que se aplican técnicas con las cuales se almacena y analiza

información, estos procesos se orientan a los usuarios para la toma de decisiones en futuros negocios o acuerdos. Este análisis de la información les da una ventaja competitiva sobre otras organizaciones ya que el potencial de la big data trae consigo conclusiones para un mejor entendimiento del consumidor.

Con el constante avance en la gestión de la información junto con las nuevas tecnologías, se evidencia la necesidad de generar de manera eficiente patrones, tendencias y conocimientos que permitan concluir respuestas de manera inmediata. Pero la Big data no solo abarca estos conceptos, también permite conocer técnicas, métodos y paradigmas que se orientan a la solución de problemas relacionados a la datos, esto da entender que la Big data brinda una perspectiva diferente sobre el procesamiento de la información esto es una gran ventaja para las necesidades que tienen las empresas respecto a sus usuarios, lo importantes en este caso en el valor que se le da a los datos recolectados, con los cuales se pueden generar estrategias que permitan una mayor productividad con lo que se está ofreciendo a los consumidores. Gracias a las tendencias de las redes sociales se facilita en gran medida la aplicación de la Big data ya que en los últimos años estas han tenido un crecimiento exponencial. Teniendo en cuenta “la cantidad de datos digitales a nivel mundial creció de 150 exabytes (billones de gigabytes), en 2005, a 1.200 exabytes, en 2010. En 2007 se preveía que dicha cantidad aumentará en un 40% anual en los años siguientes (unas 40 veces el crecimiento de la población del mundo), por lo que se esperaba que la cantidad de datos digitales aumentará 44 veces entre 2007 y 2020 (replicándose cada 20 meses)” (Tabarez, 2014, p. 3).

Tabares (2014) relaciona tres aspectos que son de gran ayuda para el desarrollo de la Big data por lo cual se debe tener en cuenta características como: El “Volumen”, lo cual hace referencia a la cantidad de datos generados, la “Velocidad” que está relacionado con la entrada y

salida de datos a través de diferentes canales y por último la “Variedad” en los que se tiene en cuenta los formatos y fuentes en los que se encuentran los datos. Esta última es importante tenerla en cuenta debido a que la veracidad de la información nos permite tener más garantías en los resultados, con respecto a las bases de datos o plataformas es necesario evidenciar si la información recolectada se puede utilizar para los procesos analíticos.

## **2.3 Categorías Conceptuales.**

En las categorías conceptuales se contextualizan los términos que se deben tener en cuenta para el óptimo desarrollo de los objetivos planteados. Se explica su definición, aplicación, relación con la archivística y la importancia para este trabajo. Las categorías conceptuales que se van a analizar son: la producción científica, los estudios métricos, la bibliometría, los indicadores bibliométricos y el Big Data.

### **2.3.1. Producción Científica**

Las bases de datos han tenido un papel importante en la producción científica en diferentes campos de aplicación, por esta razón cada vez se crean más plataformas digitales que permiten difundir las investigaciones científicas que se producen a través de una gran cantidad de artículos. La producción científica permite ver las tendencias que van surgiendo a través de los diferentes ámbitos de la ciencia, debido a esto es importante analizar esta producción a través de las revistas científicas las cuales se han transformado en garantes del conocimiento, también se encargan de analizar las investigaciones especializadas ya que son un apoyo para garantizar la calidad de la información.

Como lo menciona Ríos y Herrero (2005) se plantea que se ha visto un crecimiento en la producción científica entre 1989 y 2003 en contexto latinoamericano como principales productores de conocimiento científica se encuentra países como Brasil, Argentina, Chile y Venezuela, debido a que los países más desarrollados han encabezado la producción científica a lo largo de la historia, esto como consecuencia de que en estos países subdesarrollados no se cuentan con las oportunidades necesarias para apoyar las investigaciones científicas.

Para países como Colombia la producción científica es de las más bajas en Latinoamérica ya que aporta un porcentaje muy bajo en lo cual en su mayoría es gracias a los artículos investigativos de la universidad nacional y la universidad de los andes, A raíz de esto han surgido diversas propuestas y proyectos para incrementar las publicaciones investigativas, porque se ha identificado la relación del factor socioeconómico con este tipo de indicadores de producción científica.

### 2.3.2. Estudios Métricos

En el campo de la archivística se llevan a cabo actividades a través de las cuales han surgido técnicas y procesos que van orientadas a objetivos específicos uno de ellos son los estudios métricos, para estos procesos se aplican métodos, modelos matemáticos y estadísticos para el análisis del desarrollo científico, estos métodos permiten recopilar todo tipo de información científica a través de los índices bibliográficos. Los métodos cuantitativos se han ido desarrollando y aplicando a las ciencias de la información, esto debido a que cada vez se genera más interés en estos indicadores cuantitativos. Es por lo que se proponen esfuerzos que se orienten hacia la integración de diferentes disciplinas que se relacionen con el estudio de la información, metodologías, métodos de estudio para la generación de conocimiento.

Esto permite conocer de modo fiable, cuáles son las materias o áreas temáticas que se han abordado en un periodo de tiempo determinado y su evolución, así como conocer las tendencias en investigación científica, otro de los objetivos de estos estudios también es analizar las características y la cantidad de publicaciones científicas, para evidenciar las necesidades en el uso de la información de un grupo determinado que puede ser aplicado por países o idiomas y así generar soluciones, por medio de la accesibilidad al quehacer científico.

De acuerdo con Vargas y madrigal (2007) se ha visto la necesidad de aplicar las matemáticas a las ciencias de la información y gracias a las nuevas tecnologías y sistemas que gestionan las bases de datos se ha hecho más fácil la aplicación de estas técnicas a los índices bibliográficos.

### 2.3.3. Bibliometría.

La bibliometría se creó debido a la necesidad de aplicar métodos cuantitativos y estadísticos en la producción científica y promover en la comunidad académica el análisis bibliométrico como herramienta de evaluación desarrollo y seguimiento de la producción científica, el principal objetivo de estos métodos es la planeación, la búsqueda y análisis de información científica para la investigación. Por eso es importante indagar cuáles son las leyes e indicadores bibliométricos y cómo aplicarlos en las prácticas académicas y científicas, debido a esto se debe definir y comprender los conceptos básicos y los usos de la bibliometría.

Este método juega un papel muy importante para comparar ya sean diferentes grupos investigativos dedicados a un campo en específico, así como a instituciones educativas que generan conocimiento científico, esto con el objetivo de tomar decisiones y direccionar de la mejor manera la producción científica, también sirve como referencia para evidenciar la calidad

de la información, en caso de las revistas científicas donde se analiza la calidad de sus artículos investigativos.

A través de los años se han identificado y analizado diferentes tendencias de investigación en las cuales se describe un tema que va desarrollando un crecimiento exponencial en la producción científica, por lo cual se van indagando diferentes indicadores bibliométricos para destacar las conclusiones.

Recientemente con el tema de la pandemia Chaviano (2020) identificó el aumento en la cantidad de publicaciones relacionadas con este tema, por lo cual se realizó un estudio bibliométrico en el que se utilizaron bases de datos como web of science, scopus y pubmed para recuperar la producción científica sobre el Covid 19 en Latinoamérica. Para la búsqueda de datos se crea una ecuación de búsqueda con el objetivo de abarcar toda la información posible, esta ecuación se compone de las palabras claves y las tipologías que se van a usar.

A partir de esta información se mencionan las variables como lo es el autor, país e institución, para así generar gráficas con las que se pueda realizar un análisis general que identifique los patrones o las tendencias como el país donde más se contribuyó a la investigación sobre covid 19, para una mejor comprensión de los datos se exponen los principales indicadores con una respectiva descripción de las conclusiones.

### 2.3.5. Big Data.

Big Data se refiere al conjunto de técnicas, tecnologías y aplicaciones utilizadas para gestionar y analizar grandes volúmenes de datos, tanto estructurados como no estructurados. Estos datos pueden provenir de diversas fuentes, como redes sociales, sensores, transacciones comerciales, entre otros. El análisis de Big Data permite obtener información valiosa y tomar decisiones basadas en datos en diversos

campos, como la economía, la meteorología, la biología computacional, entre otros como lo plantea Hernandez (2017).

La Big Data contiene características específicas como el volumen, en el cual se miden la capacidad o el tamaño de los datos, los cuales se miden desde bytes hasta terabytes, también está la variedad ya que los datos pueden estar estructurados de formas diferentes, otra característica a tener en cuenta es la velocidad que se refiere a la rapidez con la que se generan los datos y está relacionado con el desarrollo de software, cada una de estas características se interrelacionan para comprender y gestionar la Big Data según Camargo y Aguilar (2014).

Según marx (2013) los principales retos de la big data incluyen la privacidad y seguridad de los datos, la interoperabilidad de las herramientas de análisis, la capacidad de almacenamiento y procesamiento de grandes volúmenes de datos, la falta de recursos y el reconocimiento para el desarrollo de herramientas de software.

Estas necesidades se deben tener en cuenta por qué los datos deben ser más accesibles para la comunidad científica e investigativa, puesto que el conocimiento es una herramienta que se debe ser compartido, pero también se deben mejorar diversos aspectos en el manejo de los datos en grandes volúmenes.



### Capítulo III. Metodología de Investigación.

En la metodología de investigación se va a sustentar que tipo de enfoque, tipo de investigación, método de investigación, las técnicas y las fases, el enfoque de investigación es cuantitativo, teniendo en cuenta que son estudios métricos en donde se utiliza el estudio de los datos por medio de la descripción analítica. . Esta metodología permite identificar patrones, tendencias y relaciones entre publicaciones, autores y demás, brindando una comprensión profunda de la evolución y la estructura de un área de investigación como lo es la Big Data. Las fases de esta investigación se definieron por los objetivos teniendo en cuenta los indicadores bibliométricos que utilizaron como lo son los autores países, años, idiomas y la co-ocurrencia. Esta investigación brinda una perspectiva más amplia del Big data y sus características más comunes a través de las publicaciones como revistas científicas, artículos, trabajos de grado ya que son herramientas confiables de acceso para el conocimiento y así definir el alcance del Big data como tema de investigación en diferentes contextos. Al Utilizar análisis de palabras clave y técnicas de agrupamiento para identificar tendencias y temas emergentes en la Big Data ayuda a comprender la evolución del campo y las áreas de investigación en crecimiento

#### **3.1 Enfoque**

La investigación cuantitativa es aquella en la que se recogen y analizan los datos sobre variables. Se basa en la estadística, por medio de métodos de medios de datos en este caso por medio de las ciencias métricas. Ocho y Molina (2018) exponen que se utilizan variables como las características o propiedades los cuales se pueden presentar en diferentes grados de intensidad y tiene caracteres numéricos para realizar su medición y así poder realizar el análisis de información.

Este enfoque de investigaciones es deductivo, ya que es una investigación que va desde la general a lo concreto. Utiliza variables que son características que tiene al menos dos variables y los atributos que son los valores que componen las variables para lo cual se utilizan autores, países, años, idiomas y coocurrencias.

Las variables pueden ser continuas que son aquellas que pueden tomar cualquier valor dentro de un intervalo o las variables discontinuas también llamadas discretas que son las que no pueden tomar valores intermedios entre otros dos valores dados, han de hacerlo siempre con valores enteros.

Por medio de un enfoque cuantitativo se busca utilizar la estadística, teniendo en cuenta variables que tengan relación con los objetivos del análisis bibliométrico que se busca desarrollar en la base de datos Scopus . Las variables son los autores, años de producción, países e idiomas. Se busca determinar un balance de la producción científica en el campo de Big Data.

### **3.2 Tipo de Investigación**

El tipo de investigación es descriptiva analítica teniendo en cuenta que por medio de ella se analiza la información sobre el tema objeto del estudio a partir de fuentes documentales en este caso las bases de datos. Utiliza la estadística para describir datos y características de objeto de estudio, su meta es la recolección de datos e identificar las relaciones que existen entre las variables, la investigación descriptiva tiene diferentes métodos uno de ellos es la observación cuantitativa, “es la recopilación objetiva de datos que se centran principalmente en números y valores” (Muguira,Sf, pag1)

Se realizó la implantación de fórmulas que permitan la creación de gráficos relacionados con la base de datos Scopus, la cual tiene información relacionada con la archivística y en este caso con el big data que es nuestro objeto de estudio. Como menciona Scopus (2022), estas bases de datos permiten realizar las ilustraciones y tablas para determinar características de la producción académica relacionada al Big data lo que también genera la posibilidad de realizar el análisis de datos.

. El análisis de estos datos es necesario para determinar la relación respecto con la producción académica, el Big data y la archivística. Se debe aplicar variables caras que fundamente el análisis de resultados y así lograr coherencias entre las ilustraciones, tables y el análisis de estas.

### **3.3 Método de Investigación**

El método de investigación que se utilizará es la investigación documental, teniendo en cuenta que los objetivos de este trabajo se tratan del análisis de información en la base de datos Scopus. La investigación documental como menciona Contantino Tancara “La investigación documental como una serie de métodos y técnicas de búsqueda, procesamiento y almacenamiento de la información contenida en los documentos, en primera instancia. Y la prestación sistemática. Coherente y suficiente argumentada de nueva información en un documento científico, en segunda instancia” (1987. - Pág. 4)

Las unidades de datos del ejercicio serán la base de datos Scopus e implicarán la investigación documental que desarrollan los investigadores con la recopilación de información para ser analizada. La recopilación de la información es crucial para este trabajo, así mismo mirar

el contexto en que se desarrollan los resultados. El contexto de los resultados demuestra las causas de la variación de datos y se evidencia el porqué es necesaria la investigación.

Otro de los factores es que el trabajo se desarrolla con las ciencias métricas las cuales se basan en la investigación documental y la aplicación de indicadores bibliométricos para dar más claridad de los resultados que se pueden obtener en este trabajo.

La investigación documental no es solo la recolección de datos si no el análisis de estos lo que genera que sea el método ideal para este trabajo y que se articule con las ciencias métricas para dar un contexto claro de las consecuencias de los resultados.

### **3.4 Técnicas**

La técnica de Investigación utilizada corresponde con lo señalado por las ciencias métricas las cuales son métodos con la que se analiza a la producción científica en las bases de datos. La base de Scopus permite generar ilustraciones y tablas relacionadas con la producción científica para un tema en particular con ciertos indicadores que permiten que los datos tengan menor probabilidad de error.

Las ciencias métricas también facilitan la recuperación de la información, por medio de la aplicación de las matemáticas y las estadísticas, las cuales se aplican en las ciencias sociales, en un campo específico de investigación, otra característica es que es interdisciplinar y una herramienta para los investigadores. La bibliometría es las que una ampliación de matemáticas, como menciona Pérez

La bibliometría integra los elementos descriptivos de la información escrita, se basa en los documentos y los resultados obtenidos siempre tendrán un carácter

científico... Es una herramienta capaz de determinar fenómenos, tendencias y regularidades que acontecen en el ámbito científico. (2002 ,p. 9)

La importancia de la bibliometría para el análisis de la producción científica en diferentes unidades de información y su atribución para las ciencias sociales y los investigadores. La técnica de la bibliometría garantiza que este trabajo se desarrolle adecuadamente para el análisis de los datos de la base de Scopus en donde se va a aplicar la bibliometría para generar resultados, los cuales se deben analizar para comprender cuales son las características de la producción científica que se encuentra en esta base de datos referente al Big Data. La Big Data al ser en análisis masivo de datos y su importancia en diferentes campos de investigación o de negocios crea la necesidad de determinar las características que tiene la información publicada en Scopus referente a este tema, que es parte de las tendencias de la archivística a nivel mundial.

Para localizar la información nos basamos en el autor Cruz (2020) quien plantea un método para la búsqueda de bibliografía, para esta búsqueda de datos en la base de datos Scopus se utilizan una formula en la cual se utilizan palabras claves como elemento principal y otras palabras que se utilizan como elemento contextual y por último se agregan otro aspecto para que funcione como articulador.

Una vez se arrojan los resultados mediante la base de datos se realiza el proceso de selección para los indicadores que vamos a analizar en los cuales se filtran los resultados para que solamente se considere la producción científica en el área del Big data.

### **3.5 Instrumentos**

El instrumento utilizado fueron los indicadores bibliométricos, los cuales permiten cuantificar las características del conjunto de los documentos estudiados. Estos indicadores son herramientas que permiten analizar los resultados para darle un contexto que establezca fenómenos que contribuyeron a estos resultados. Los indicadores pueden ser personales los cuales, producción, dispersión, visibilidad o impacto, colaboración obsolescencia, forma y contenido. El aporte y concepto dado por Ardanuy (2012), permite tener mayor claridad sobre el concepto de indicador bibliométrico. los indicadores bibliométricos son datos numéricos calculados a partir de características bibliográficas observadas en los documentos publicados en el mundo científico y académico, o los que utilizan los usuarios de tales documentos y que permiten el análisis de rasgos diversos de la actividad científica, vinculados tanto a la producción como al consumo de información”. (, p. 16)

Scopus cuenta con siete indicadores los cuales son por año, autor, fuente, afiliación, país, tipo, área temática, patrocinador de financiación (Institución productores). Se encuentran herramientas para filtrar por mayor citación, es importante reconocer que, aunque puede ser que el documento sea uno de los más citados no quiere decir que sea el más verídico ya que se pudo haber citado para contradecir la información que se tiene en él.

Estos indicadores deben ser procesados de tal forma que se les puedan dar un contexto para poder entender a mayor profundidad el movimiento de la producción científica en este caso en el Big Data. Los términos que se tienen en cuenta para este trabajo son Big Data, Analítica de datos, Análisis masivo de datos, estos términos se escogieron teniendo en cuenta que son relacionados con la archivística y las nuevas tecnologías.

### 3.6 Fases

#### ***1. Identificación de los autores, países y años de mayor producción científica sobre Big Data.***

Se desarrollaron dos actividades importantes, la primera fue la estructuración de la fórmula para su adecuada aplicación en las bases de datos la cual tiene como objetivo extraer la información para que se pueda comprender de forma clara, y sea útil en el ejercicio de la ilustración de esta información.

La segunda actividad se trata de la compresión de la información en cuanto al análisis de los autores, países y años de producción para poder representarlo en tablas e ilustración. Esto genera una mayor compresión de la producción científica en el campo del Big Data.

#### ***2. Determinación de los idiomas en que mayormente se publica de Big Data y las áreas temáticas conexas***

Para la determinación de los idiomas de producción requiere de dos actividades: la primera consiste en el análisis y comprensión de los datos generados en cada base de datos, los cuales se deben ilustrar. Las áreas temáticas conexas también se deben analizar y comprender para revisar en qué campo se genera más cantidad de información y así determinar las áreas de interés.

La segunda actividad consiste en la comparación de las ilustraciones de las cuatro bases de datos para determinar semejanzas o diferencias que tenga la información y realizar un análisis deductivo de los idiomas con mayor publicación y las áreas temáticas conexas que tienen

participación en Big Data. Se va a estipular cuáles son las bases de datos con mayor producción académica, teniendo en cuenta el idioma y el área temática.

### ***3. Establecimiento del balance de la producción científica sobre Big Data en las fuentes de datos seleccionadas***

El establecimiento del balance de la producción científica sobre Big Data requiere de dos actividades, la primera consiste en el análisis de toda la información obtenida y la adecuada comprensión de las tablas y las ilustraciones, mencionando las características de cada uno de los resultados que se obtuvieron con la aplicación de las variables ya planteadas.

La segunda actividad es la comparación de los resultados obtenidos en las bases de datos, describiendo según las variables aplicadas y generar las conclusiones para determinar si se alcanzaron los objetivos planteados y las características concretas de cada variable y de cada base de datos siempre haciendo referencia al campo del Big Data.



## Capítulo IV. Resultados.

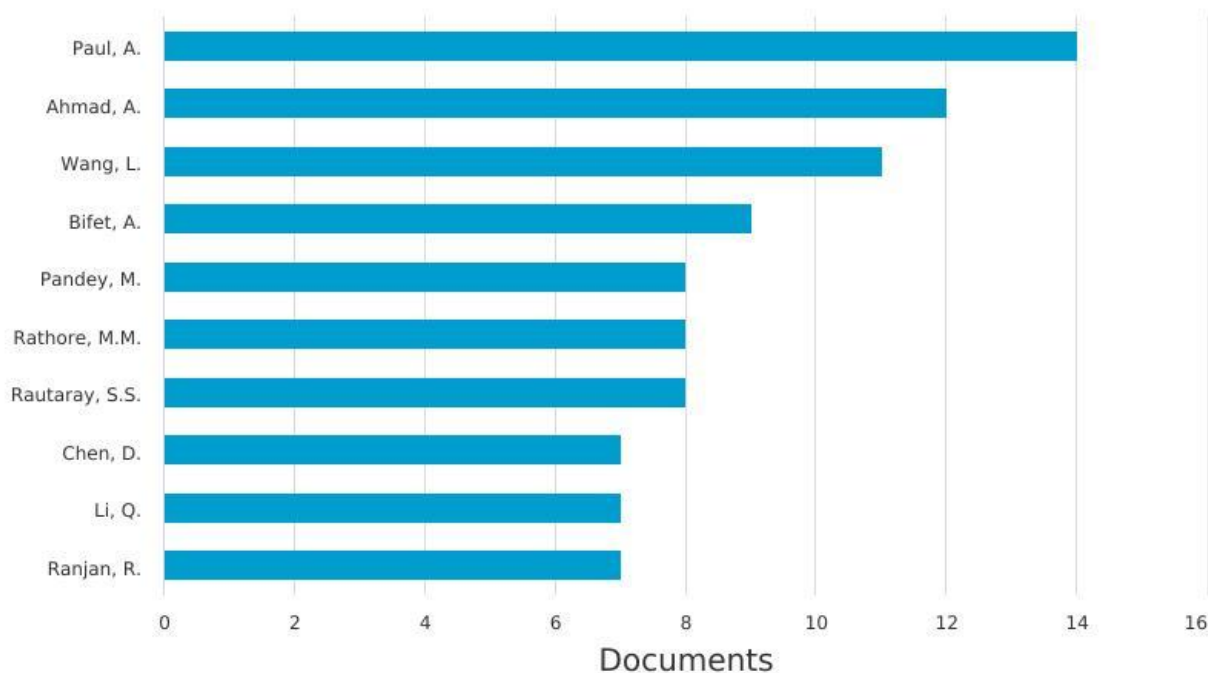
Para los resultados utilizaron ocho indicadores bibliométricos a los cuales se les realizó su respectiva ilustración para comprender los resultados y dar un contexto de cada uno de ellos. Los indicadores que se utilizaron fueron: autores, países, producción científica a través de fuentes, años, filiación institucional, tipo de documento, área temática y patrocinador de financiación.

### 4.1 Documentos por Autores

#### Documents by author

Scopus

Compare the document counts for up to 15 authors.



Copyright © 2022 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

De acuerdo con la producción científica que nos brinda Scopus a través de las gráficas, se identifican los autores más destacados en la Big data desde el año 2014 a 2022, esto evidencia que el análisis de la información a través de Big Data está tomando más importancia ya que muchos de estos autores orientan el tema para analizar patrones o tendencias en la investigación a través de revistas científicas, bibliotecas digitales o tesis que aportan información en masa con la

cual se pueden tomar decisiones en diferentes ámbitos.

Uno de los autores que más ha contribuido a la investigación es Paul Anand un investigador norcoreano que se ha enfocado en la Big data relacionándolo a varios temas como la salud, uno de los artículos más interesantes se titula: ‘‘Toma de decisiones basada en gráficos multinivel en Big Scholarly Data: un enfoque para identificar revisores expertos, encontrar factores de impacto de calidad, clasificar revistas e investigadores’’. En este documento se habla de la importancia de los gráficos para procesar eficientemente la información, lo cual presenta un mecanismo para identificar las necesidades que tiene actualmente la comunidad académica hoy en día, a través de algoritmos los cuales permiten clasificar las revistas para que brinden los campos de investigación que más se han popularizado, esta información es posible debido a la medición del número de citas que tiene cada artículo.

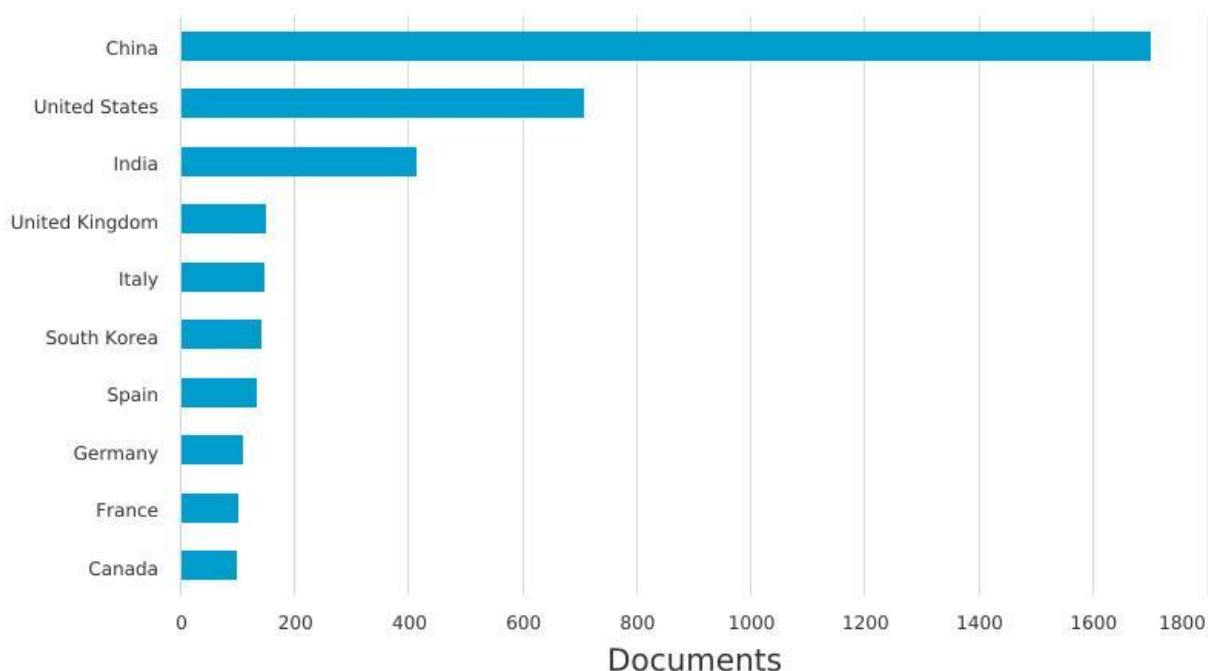
Paul Anand junto a otros investigadores realizaron un estudio en cuanto al análisis de Big data de medios geosociales para la planificación y decisiones en tiempo real, en este artículo, se propuso un sistema eficiente para explorar Redes Geosociales mientras recolectamos datos, así como la información de ubicación del usuario. Se propone una arquitectura de sistema que procesa una cantidad abundante de diversas redes sociales, datos para monitorear eventos terrestres, incidentes, enfermedades médicas, tendencias de usuarios y vistas para tomar decisiones futuras en tiempo real y facilitar la planificación futura.

## 4.2 Documentos producidos por país

### Documents by country or territory

Scopus

Compare the document counts for up to 15 countries/territories.



Copyright © 2022 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

Los países que más han contribuido a la investigación en Big data son China y Estados Unidos, de acuerdo con la base de datos Scopus, China ha producido poco más de 1.700 documentos relacionados al análisis de información, al ser este un país que ha invertido una cantidad importante de recursos en la producción científica, esto evidencia el compromiso que tiene con el desarrollo del Big data en el país, debido al potencial y las alternativas que permite conocer en cuanto a factores económicos, sociales y políticos, el análisis predictivo que ofrece en diferentes entornos les permite tener una mejor perspectiva y tomar de decisiones en base a la información, gran parte de los artículos que producen están enfocados en las ciencias de la computación y la ingeniería, con ello se puede concluir el gran impacto que ha tenido la Big data en la comunidad académica y científica de China.

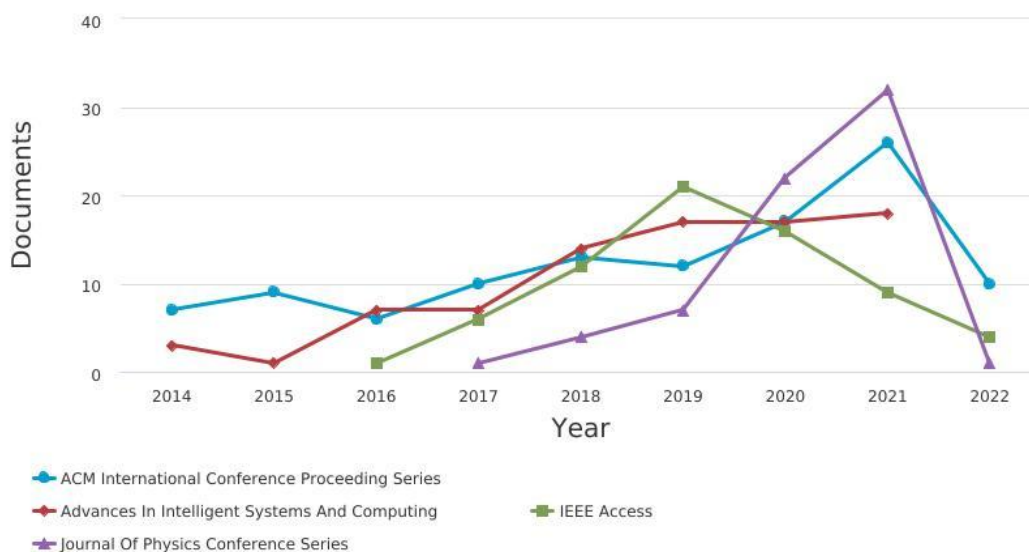
Por otra parte, Estados Unidos es otro de los países que están aportando al conocimiento científico con más de 700 documentos relacionados al análisis de la información y la Big data. Su campo científico está orientado a los avances tecnológicos que se están desarrollando a través de patrones y experiencias que brindan el análisis de información en masa, también han desarrollado modelos de información para su aplicación en problemáticas actuales como es el caso del cambio climático y así poder medir el impacto negativo que está teniendo actualmente, a través de los datos crean sistemas automatizados que permiten analizar los riesgos y poder tomar decisiones que solucionen estos problemas.

#### 4.3 Producción científica a través de fuentes

##### Documents per year by source

Compare the document counts for up to 10 sources. Compare sources and view CiteScore, SJR, and SNIP data

Scopus



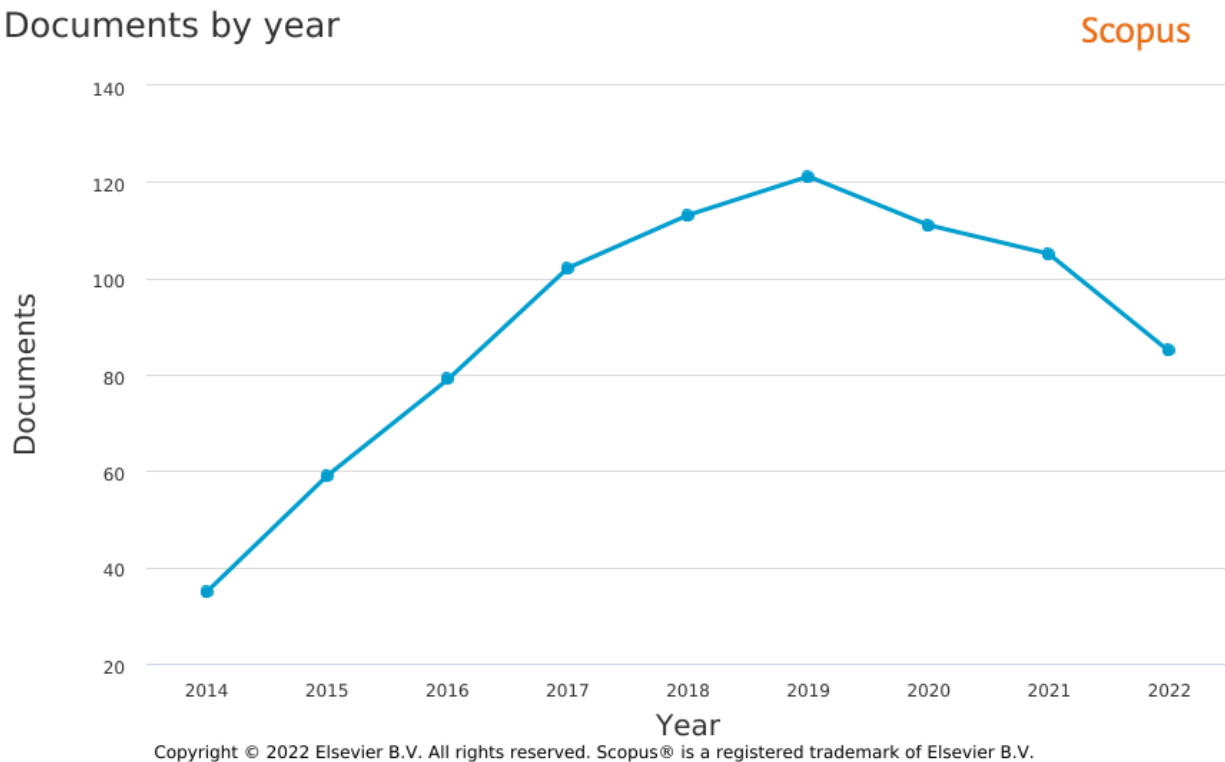
Copyright © 2022 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

Según la anterior gráfica de las fuentes que más producen información referentes al Big data son la ACM International Conference proceeding series, esto permite ver que mucha de esta información que se encuentra en Scopus sobre el análisis de la información ha sido aportado por conferencias internacionales donde se debate sobre diferentes puntos de vista y proyectos que se

están llevando a cabo con la ayuda de la Big data, en estas conferencias se analizan temas como la inteligencia artificial que está muy ligado al estudio de información en grandes cantidades con la cual permita responder a las necesidades que se van presentando.

#### 4.4 Documentos por año.

Documents by year



Los documentos por año que son acorde a la producción científica relacionada con el Big Data muestran una curva de crecimiento desde el 2014 en donde su punto más alto corresponde a 2019, y se evidencia una declinación para los años 2020, 2021 y 2022. El crecimiento en la información recopilada se debe a la necesidad que se presentó a nivel mundial por las nuevas tecnologías y su aplicación a la vida laboral por causa de la pandemia del COVID- 19. Esta nueva tecnología orientada al análisis masivo de datos permitió que las ventas online y las actividades técnicas de la empresa se desarrollaran en espacios diferentes a la oficina. Las ventas

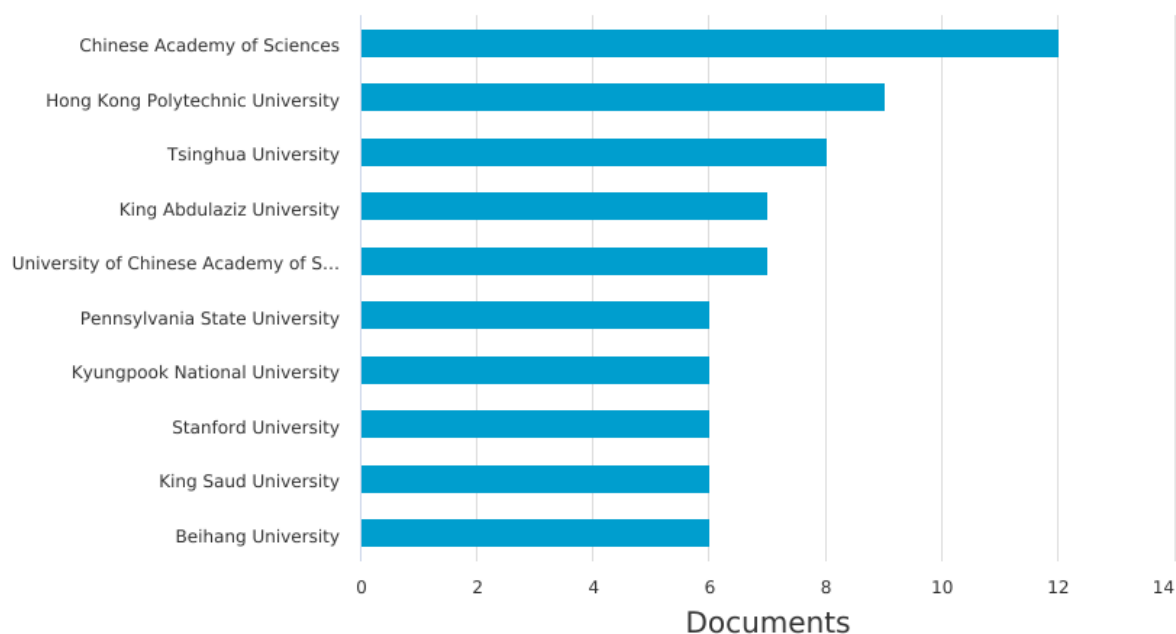
en línea también se vieron intervenidas por el Big Data ya que se analizaban las necesidades de los usuarios para crear publicidad personalizada y mejorará las ventas.

#### 4.5 Documentos por Afiliación.

##### Documents by affiliation

Scopus

Compare the document counts for up to 15 affiliations.



Copyright © 2022 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

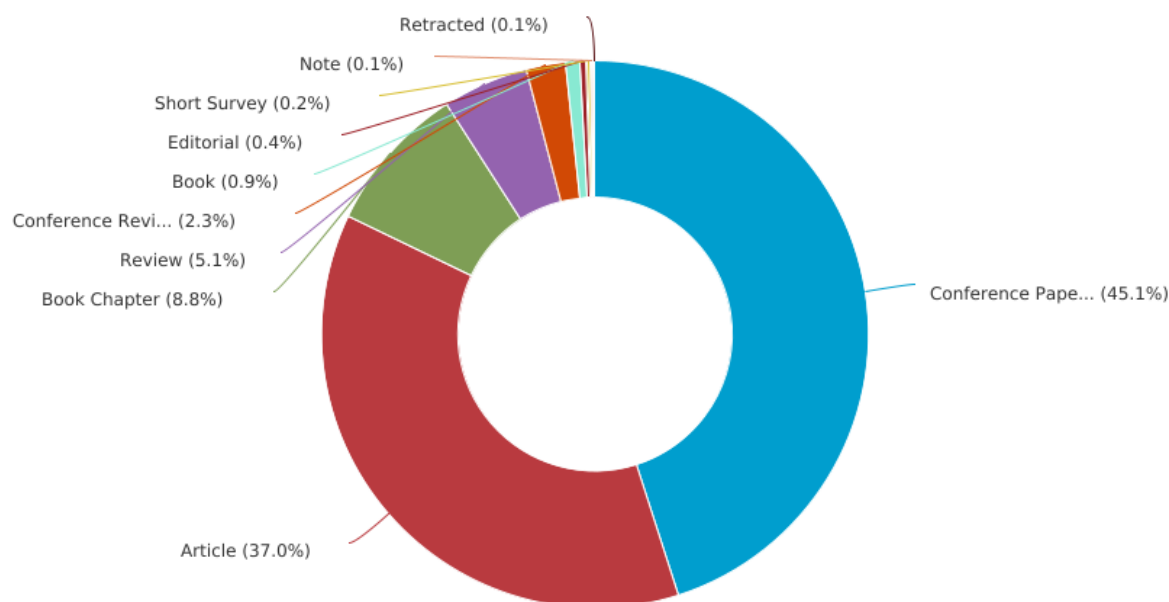
Estas instituciones que tienen mayor cantidad de documentos se encuentran situadas en China y Hong Kong. Los documentos por afiliación en donde se representa la institución de Chinese Academy of Sciences, es donde se presenta mayor producción, esta academia es considerada el eje del impulso de China para explorar y aprovechar la alta tecnología y las ciencias naturales. En segundo lugar, se encuentra The Hong Kong Polytechnic University la cual se caracteriza por ser una institución multidisciplinaria que busca formar líderes en el mundo para desafíos sociales importantes. En el tercer lugar, se encuentra Tsinghua University en donde se

identifica el énfasis en las ciencias y las humanidades, otra de las características es que esta institución ha formado una cantidad de once presidentes. .

#### 4.6 Documentos por Tipo.

##### Documents by type

Scopus



Copyright © 2022 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

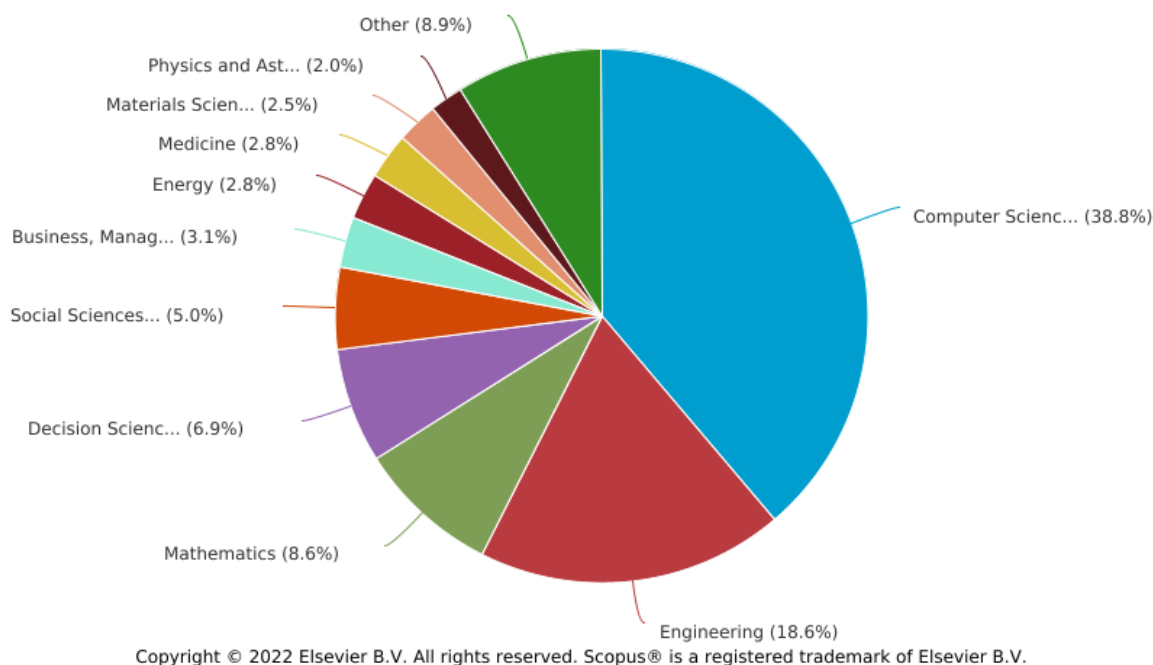
Los documentos por tipo se organizan por conferencias, artículos, capítulos de libros, revistas, libro, editorial, encuesta corta, nota y retratados. En primer lugar, se encuentran las conferencias, en el 2021 siete conferencias principales y en el 2019 con un total de doce. La segunda posición es de los artículos en donde el artículo con mayor citación es del articulo de Boyd, Danah 2012 Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon, donde se desarrollan preguntas relacionadas con el campo interdisciplinario del Big Data. En tercer lugar, esta los capítulos de libros, donde se encuentra el capítulo de Recommendation Systems for personalized technology – enhanced learning con 41 citas del autor Khribi, Mohamed Koutheair del 2015, en donde se desarrolla el tema de comercio

electrónico hasta el aprendizaje electrónico, la minería de datos, analítica de datos y grandes datos.

#### 4.7 Documentos por Área Temática.

Documents by subject area

Scopus



Los documentos por área temática se dividen por un total de 24 áreas temáticas, las cuales se desarrollan la medicina, las tecnologías, economía, agricultura, ciencias y medicamentos. En donde se tiene mayor producción científica son las ciencias de la computación, ingeniería, matemáticas y economía, se debe resaltar que a nivel de la medicina también se tiene una gran participación que se dividen por especialización como es el caso de la psicología o la neurología. Las ciencias de la computación se asocian a la archivística y tienen un total de 616 producciones, el documento con más citación es Deep learning applications and challenges in Big data analytics con un total de citas 1252, en donde se desarrolla la importancia de estos dos conceptos para que las computadoras puedan desarrollar actividades como los humanos aprendiendo de



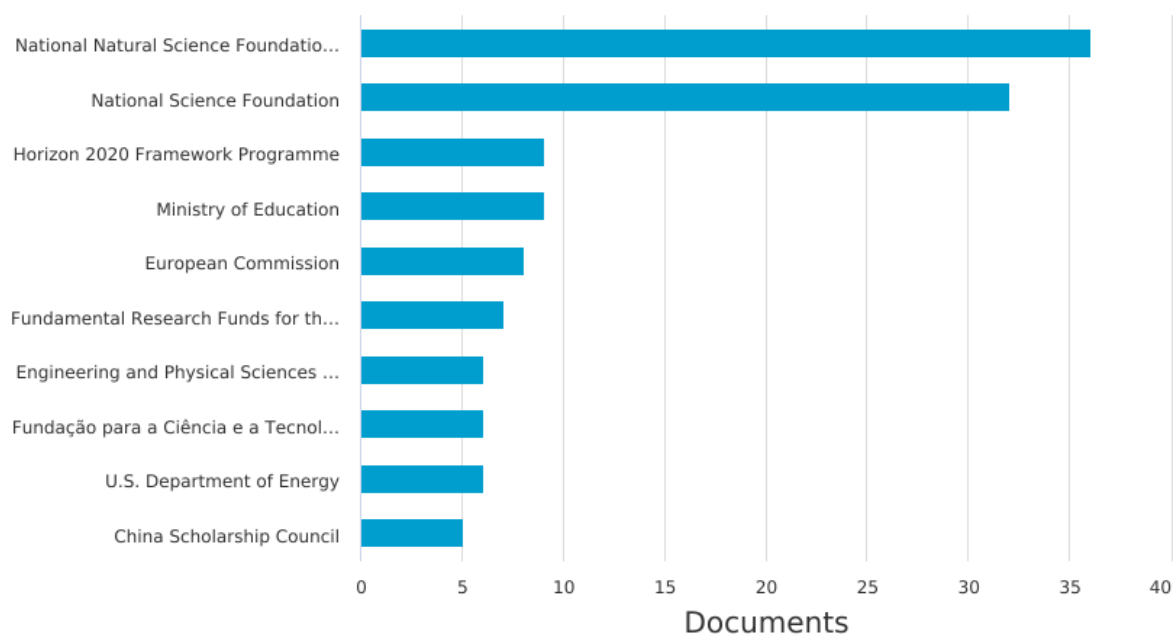
configuración de parámetros básicos acerca de los datos, por eso es la importancia del big data. Esto se puede aplicar para diferentes actividades como algunas relacionadas con gestión documental para mejorar los procesos y disminuir el margen de error, en procesos técnicos.

#### 4.8 Documentos por patrocinador de financiación.

##### Documents by funding sponsor

Scopus

Compare the document counts for up to 15 funding sponsors.



Los documentos por patrocinador financiero, aunque se desarrollen diferentes patrocinadores, pero tiene en cuenta los patrocinadores a partir de cinco producciones científicas. Los dos patrocinadores con mayor participación son National Natural Science Foundation of China y National Science Foundation, los cuales tiene más de 30 producciones científicas las demás no superan más de 10 producciones. El documento con mayor cantidad Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges la citación es de 536

citaciones en donde se desarrollan los temas de la analítica de datos, las relaciones con las redes eléctricas y la transición de los sistemas energéticos.

## CONCLUSIONES.

En la identificación de los autores se reconocieron los tres principales que son Paul, Ahmad y Wang, los cuales tienen mayor producción científica frente a la big data. En cuanto a los años con mayor producción científica fue tomado a partir del año 2014 en donde se evidencia un incremento a partir del 2020 pero se encuentra un declive en el 2022, esto se caracteriza por la pandemia del covid-19. Los países con mayor producción son China y Estados Unidos en donde se evidencia el desarrollo de diferentes áreas temáticas como la computación y la ingeniería, esto evidencia el apoyo que se brinda en estos países a la investigación y las publicaciones científicas. Los documentos con mayor producción por tipos de documentos se resalta las conferencias y los artículos en donde se desarrollan temas relacionados con el Big Data, la analítica de datos o el manejo masivo de datos.

En la determinación de las producciones a través de fuentes, se identifica el que tiene mayor producción es el de ACM International Conference Proceeding, la cual es una conferencia en donde se desarrolla temas como el análisis de información construyendo conocimiento por medio de debates. En los documentos por afiliación se desarrollan entidades que se concentran en China o Hong Kong, esto es evidencia de la importancia que le están dando al Big Data en las potencias mundiales como China. Los documentos por áreas temáticas se caracterizan por medio de las ciencias de datos, computación, ingeniería y otras áreas como la medicina en diferentes campos de aplicación. Los documentos por patrocinador financiero, en donde se evidencia dos entidades principales y las demás se desarrollan con menos participación. En cuanto a las áreas temáticas se resalta la reacción que tiene el Big Data, con las ciencias de tecnologías y las ciencias sociales, los patrocinadores pueden ampliar el campo de aplicación orientado a estas áreas.

El Big Data se desarrolla en muchas áreas temáticas en donde se evidencia su aplicación en cualquier campo de estudio. La producción científica referente al Big Data se desarrolla en mayor medida en China en donde se puede evidenciar mayor participación por patrocinador financiero y países, debido a que invierten más recursos enfocados en el análisis a través de la big data. Esto puede resaltar la curva de crecimiento por la pandemia mundial del 2019 lo que generó un mayor interés por el Big Data. Lo anterior a causa de las medidas para combatir el aumento de contagios, lo cual fue una herramienta que cumplió su objetivo.

La importancia del Big Data creció potencialmente pero actualmente tiene un declive el cual se debe estudiar para determinar su causa. El Big Data se desarrolla en muchas áreas temáticas, está relacionada con la archivística en donde se buscan alternativas de aprovechamiento de datos. La archivística y el Big Data se relacionan para facilitar la lógica de los datos y poder crear conocimiento lo que facilite la toma de decisiones. La economía es una de las áreas temáticas se debe reforzar por causa de que es una oportunidad para incrementar el crecimiento de pequeños y grandes negocios.

## RECOMENDACIONES.

Teniendo en cuenta el análisis de la base de datos Scopus y sus resultados arrojados mediante gráficas en cuanto al Big data, es necesario tener en cuenta que en varios países se está consolidando el análisis mediante el Big data en varios ámbitos de estudio, cabe recalcar que se deben invertir más recursos que se orienten a la investigación de la Big data.

Cómo recomendación las universidades deben garantizar el acceso a las bases de investigación para que la Big data sea más utilizada en las organizaciones en aspectos económicos, sociales y políticos que relacionen directamente a los ciudadanos, para garantizar sus derechos a la información que se encuentra en las bases de datos.

En cuanto al estudio bibliométrico es un método que ha tenido un gran avance en las bases de datos, pero se deben desarrollar en el ámbito académico para que así mismo se utilicen como referencia para tesis y artículos de investigación. Estos indicadores pueden resaltar esos vacíos que se encuentran en la producción científica, por lo tanto se pueden fortalecer estos temas a través de la creación de artículos y publicaciones que permitan tener un mejor avance en distintos campos de la investigación.

La archivística como profesión debe hacer énfasis en los estudios bibliométricos, ya que estos brindaran conocimiento que son el fundamento base para la toma de decisiones en las organizaciones, y así mismo crear otras perspectivas en cuanto al análisis de la información

El semillero YUARDI, el cual tiene en cuenta la evolución de las tecnologías en el campo de la archivística, en donde se puede desarrollar proceso de investigación orientados hacia el big data teniendo en eta este estudio bibliométrico para comenzar a aumentar la producción científica en esta área, resaltando la relación con la archivística para impulsar la aplicación de conceptos y diferentes procesos.

## REFERENCIAS

6 to Emdata. (2021). <https://encuentromundialbigdata.com/2021/conferencistas/>

Ardila Pérez, J., (2021). *Análisis bibliométrico de la visibilidad de la producción científica de América latina en ciencia de la información, propuesta para revisar la interdisciplinariedad* [Tesis de pregrado, Pontificia Universidad Javeriana]. Repository Javeriana.

<https://repository.javeriana.edu.co/bitstream/handle/10554/58717/Trabajo%20de%20Grado%20-%20Juli%20c3%a0n%20Ardila.pdf?sequence=1&isAllowed=y>

Vargas Bolaños, R., & Madrigal Vargas, Y. (2007). Estudios métricos como aliado en el trabajo del profesional de la información.

<https://www.kerwa.ucr.ac.cr/handle/10669/734>

Camargo Vega J., Camargo ortega J., Joyanes Aguilar L. (2014). *Conocimiento Big Data, Vol 24*. 1-15. <http://www.scielo.org.co/scielo.php?pid=S0121->

Cárdenas Isaza J., (2019). *Visibilidad de la producción científica y académica de los grupos de investigación adscritos al centro de investigación sociales y humanas, desde la ruta verde del repositorio institucional de la Universidad de Antioquia* [Tesis pregrado, Universidad de Antioquia]. Biblioteca Digital.

[https://bibliotecadigital.udea.edu.co/bitstream/10495/16833/1/CardenasJuan\\_2019\\_VisibilidadProduccionCientifica.pdf](https://bibliotecadigital.udea.edu.co/bitstream/10495/16833/1/CardenasJuan_2019_VisibilidadProduccionCientifica.pdf)

Cárdenas Reyes Y., (2020). *Análisis bibliométrico aplicado a la ciencia abierta 2010-2019* [Trabajo de grado pregrado, Universidad de La Salle]. Archivo Digital.

[https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=2027&context=sistemas\\_informacion\\_documentacion](https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=2027&context=sistemas_informacion_documentacion)

Chaviano O., Limaymanta C. y Lopez. E, (2020). *Análisis bibliométrico de la producción científica latinoamericana sobre covid -19, Vol 40.* 1-12.

[http://www.scielo.org.co/scielo.php?pid=S0120-41572020000600104&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S0120-41572020000600104&script=sci_arttext)

Galiana P., (2019). *25 terminos del Big Dat básicos que debes conocer.* IEBS.

<https://www.iebschool.com/blog/glosario-big-data/#:~:text=Bien%20conocido%20por%20todos%20es.que%20corre%20por%20la%20Red>

Goyzueta S. (2015). *Big data Marketing: una aproximación, Vol 35.* 1-12.

[http://www.scielo.org.bo/scielo.php?pid=S1994-37332015000100007&script=sci\\_arttext](http://www.scielo.org.bo/scielo.php?pid=S1994-37332015000100007&script=sci_arttext)

Hernandez E., Leal N., Duque D. y Moreno J. (2017). *Big Data: una exploración de investigación tecnologías y casos de aplicación, Vol 20.* 1-24.

<http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf>

Hernández J., Alvares E. y Cantón R. (2021). Tendencia del big data y cloud computing:

Bibliometria del 2010 al 2020, *Ciencia latina revista científica multidisciplinar, Vol 5.* 1

<https://ciencialatina.org/index.php/cienciala/article/view/601/768>

Wu J., Wang J., Nicholas S., Maitland E. y Fan Q. (2020). *Applition of Big Data technology for covid- 19 prevention and control in China: lessons and recommendations, Vol 22.* 1.

<https://www.jmir.org/2020/10/e21980/>

Ladino L., (2020). *Evaluación bibliométrica y temática de la revista Colombiana Internacional 2011-2020*, [Trabajo de grado pregrado, Pontificia Universidad Javeriana] Repository



Joveriana.

<https://repository.javeriana.edu.co/bitstream/handle/10554/54367/Johana%20Marcela%20Ladino%20Lemus%20-%20Trabajo%20de%20grado.pdf?sequence=2&isAllowed=y>

León L., Socorro A., Cáceres M. y Pérez M. (2020). *Producción científica en América Latina y el caribe en el periodo 1996-2019, Vol 49*. 1-10.

[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0138-65572020000300013#:~:text=Los%20resultados%20muestran%20que%20la,poco%20reconocimiento%20a%20escala%20global](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-65572020000300013#:~:text=Los%20resultados%20muestran%20que%20la,poco%20reconocimiento%20a%20escala%20global)

Márquez J., (2020). *Inteligencia artificial y Big data como soluciones frente a la covid -19, Vol 50*. 1-17.

[https://scielo.isciii.es/scielo.php?pid=S1886-58872020000300019&script=sci\\_arttext&tln\\_g=pt](https://scielo.isciii.es/scielo.php?pid=S1886-58872020000300019&script=sci_arttext&tln_g=pt)

Marx V. (2013). The big challenges of big data. Vol 498. *Technology feature*.

<https://www.nature.com/articles/498255a#citeas>

Muguirra (S.f). *¿Qué es la investigación descriptiva?* Questionpro.

<https://www.questionpro.com/blog/es/investigacion-descriptiva/>

Ochoa S. y Molina M. (2018). *Estadísticas. Tipos de variables. Escalas de medida*. 1-5.

[https://evidenciasenpediatria.es/files/41-13363-RUTA/Fundamentos\\_29.pdf](https://evidenciasenpediatria.es/files/41-13363-RUTA/Fundamentos_29.pdf)

Reyes, C., y Martínez, R. (2021). *Alfabetización digital en la educación. Revisión sistemática de la producción científica en Scopus. Revista de Educación a Distancia*.

<https://doi.org/10.6018/red.444751>

Ríos C. y Herrero V. (2005). *La producción científica y la ciencia mundial: una revisión bibliográfica (1989-2003)* Vol 28. 46-61

<http://www.scielo.org.co/pdf/rib/v28n1/v28n1a03.pdf>

Romera, I., Jesús, M., ( 1992). *Potencialidad de la bibliometría para el estudio de la ciencia: Aplicación a la educación especial*. Revista de educación.

<https://redined.educacion.gob.es/xmlui/handle/11162/81717>

Valero, J. (2022) *Bibliometría. Origen y evolución*. Hospital a domicilio

<https://scielo.isciii.es/pdf/had/v6n3/2530-5115-had-6-03-105.pdf>

Scopus. (2022)

[https://service.elsevier.com/app/answers/detail/a\\_id/15534/supporthub/scopus/#tips](https://service.elsevier.com/app/answers/detail/a_id/15534/supporthub/scopus/#tips)

Spinak, E., (1996). *Diccionario enciclopédico de bibliometría, cienciometría e informetría*.

Unesco

<https://unesdoc.unesco.org/ark:/48223/pf0000243329/PDF/243329spa.pdf.multi>

Pérez, L., Hernández, J., (2014). *Big Data analytics: Oportunidades, retos y tendencias*. Universidad de San Buenaventura.

<https://gc.scalahed.com/recursos/files/r161r/w25569w/Big%20Data%20Analytics.pdf>

Gorriz, V., Castera, V. (2018) *La bibliometría en la evaluación de la actividad científica*. Universidad de Valencia

<https://revistahad.eu/index.php/revistahad/article/view/51/40>

Universidad externado de Colombia. (S.F) *Los avances de Colombia en la Big Data*

<https://www.uexternado.edu.co/derecho/los-avances-de-colombia-en-el-big-data/>