



UNIVERSIDAD DE
COSTA RICA

EBCI

Escuela de
Bibliotecología y Ciencias
de la Información

e-Ciencias de la Información

La informetría y el análisis del discurso aplicados a la producción científica en la Ciencia de Datos y Ciencia de la información

Celso Martínez Musiño

Recibido: 23/12/2020 | Corregido: 20/05/2021 | Aceptado: 26/05/2021

e-Ciencias de la Información, volumen 11, número 2, Ene-Jun 2021

DOI: <http://dx.doi.org/10.15517/eci.v11i2.45234>

ISSN: 1649-4142



¿Cómo citar este artículo?

Martínez Musiño, C. (2021). La informetría y el análisis del discurso aplicados a la producción científica en la Ciencia de Datos y Ciencia de la información, *e-ciencias de la información* 11(2). doi: [10.15517/eci.v11i2.45234](https://doi.org/10.15517/eci.v11i2.45234)

La informetría y el análisis del discurso aplicados a la producción científica en la Ciencia de Datos y Ciencia de la información

Informetrics and discourse analysis applied to scientific production in Data Science and Information Science

Celso Martínez Musiño¹ 

RESUMEN

Objetivo. Investigar la producción científica en la intersección temática de la Ciencia de Datos (CD) y la Ciencia de la Información (CI). **Método.** Estudio informétrico, descriptivo, y de primera incursión, en el análisis del discurso escrito de los textos académicos incluidos en la Web of Science (WoS), periodo 1900 al 6 de noviembre de 2020, y cuya cobertura de búsqueda fue en las bases de datos: SCI-Expanded, SSCI, A&HCI, ESCI, CPCI-S, CPCI-SSH, BKCI-S y BKCI-SSH. **Resultados.** Se recuperaron y analizaron 49 documentos representados en 38 artículos, 7 textos de memorias de congresos, 2 capítulos de libro y 2 materiales editoriales. El conjunto de las investigaciones que tratan el tema de la CD y la CI sumaron 128 citas, 2.6 citas por documento e índice H: 7. **Discusión.** Conceptualmente, se encontró que para la CD y la CI su origen son los “datos” y que ambas disciplinas son predominantemente de carácter práctico. En aquellas investigaciones con mayor visibilidad hay participación multiautoral. La CD y la CI son áreas del conocimiento recientes en las cuales las tecnologías de la información son indispensables para el análisis de grandes cantidades de datos e información. **Conclusiones.** La CD y CI tienen un carácter intra y multi y transdisciplinar y se caracterizan por utilizar las tecnologías de la información para el análisis de grandes cantidades de datos e información.

Palabras Clave: Ciencia de datos; Ciencias de la información; Informetría; Análisis del discurso escrito.

ABSTRACT

Objective. To research scientific production at the thematic intersection of Data Science (DS) and Information Science (IS). **Method.** Informetric, descriptive study, and of first incursion, in the analysis of the written discourse of the academic texts included in the Web of Science (WoS), period 1900 to November 6, 2020, and coverage in the databases: SCI-Expanded, SSCI, A & HCI, ESCI, CPCI-S, CPCI-SSH, BKCI-S and BKCI-SSH. **Results.** 49 documents represented in 38 articles, 7 conference memoirs, 2 book chapters and 2 editorial materials were retrieved and analyzed. The set of investigations that deal with the subject of DS and IS added 128 citations, 2.6 citations per document and H index:

¹ Dokumenta Research, Ciudad de México, MÉXICO, correo cmartinez@colmex.mx

7. **Discussion.** Conceptually, it was found that for DS and IS their origin is the “data” and that both disciplines are predominantly practical. In those investigations with greater visibility there are more than an author. DS and IS are recent areas of knowledge in which information technologies are indispensable for the analysis of large amounts of data and information. **Conclusions.** DS and IS have an intra- and multi and transdisciplinary character and they are characterized by the use of information technologies for the analysis of large amounts of data and information.

Keywords: *Data science; Information sciences; Informetrics; Written Speech Analysis.*

1. INTRODUCCIÓN

En este Siglo XXI, cuando con las Tecnologías de la Comunicación y la Información (TIC) nos permiten acceder a basta cantidad de datos, información y conocimiento, así como a gran variedad de servicios mediante sofisticados equipos de cómputo o la versatilidad de aplicaciones en los *smartphones*, es el contexto en el que se desarrolla este artículo. En esta situación de generación de grandes datos e información también participan las comunidades científicas. El conocimiento se desarrolla en todos los espacios creados por la naturaleza y por el ser humano, pero los saberes científicos se originan por las preguntas de los individuos colectivos o instituciones, y su forma de transmitirse, entre otras maneras, es con sus publicaciones. Así, los editores preparan materiales que compilan y reflejan, mediante libros, revistas, memorias de congresos, por mencionar algunos, para ser consultados y compartidos por las comunidades. Otro espacio donde confluyen autores a través de sus productos investigativos son las bases de datos especializadas, la *Web of Science* (WoS) por ejemplo.

Entonces, este estudio, se guían por los cuestionamientos siguientes: ¿Es posible mediante herramientas informétricas conocer la producción científica en las Ciencia de Datos (CD) y de la Ciencia de la Información (CI)? ¿Cuáles son los autores, las publicaciones y las temáticas o áreas del conocimiento que se intersectan con la CD y la CI? ¿es posible con el análisis del discurso escrito describir la evolución cronológica de la CD y la CI? Estas preguntas son el detonador para enumerar las motivaciones para desarrollar esta investigación: a) la CI y la CD son áreas del conocimiento recientes; b) ambas disciplinas mantienen un carácter intra y multidisciplinario; c) la CD y la CI, debido a la masificación en el uso y aplicación de las TIC sostienen un crecimiento en la publicación documental científica; y, d) en las definiciones conceptuales de la CD y la CI hay una intrínseca convergencia a través de los datos.

Al buscar antecedentes de investigaciones que atienden las inquietudes de este estudio, podemos decir que, en estricto sentido, no las hay. Sí existen trabajos que abordan la temática de manera amplia, es decir, sumando la búsqueda y análisis de los documentos científicos que tratan la CI y la CBI con la CD. Por ejemplo, se localizaron los artículos, “*Data science and its relationship to library and information science: a content análisis*” (Virkus y Garoufallou, 2020) y “*Data science from a library and information science perspective*” (Virkus y Garoufallou, 2019); y, una editorial, “*Ciência de Dados e Ciência da Informação*” (Freire y Freire, 2019).

Los dos artículos ofrecen resultados del análisis de publicaciones que exploran la CD desde la perspectiva de la Ciencia Bibliotecaria y de la Información (CBI). El primer artículo presenta resultado de 80 documentos analizados (Virkus y Garoufallou, 2020, p. 643), en tanto que el segundo, además de buscar documentos de CD, se incluyeron otros temas como “e-science”, “e-research”, “data curation” o “research data management” en WoS (Virkus y Garoufallou, 2019, p. 422). Por su parte, la nota editorial encuentra en la producción de datos y la capacidad de su procesamiento una oportunidad de investigación en la CI (Freire y Freire, 2019, p. 3).

La ausencia de investigaciones que atienden específicamente los tópicos de la CD y la CI es uno de los problemas que se atienden con este artículo. Además, la intra y la multidisciplinariedad de la CD y la CI dificultan la clasificación temática, pero con los estudios informétricos es posible: esquematizar y conocer la calidad, el análisis del discurso escrito y establecer mapas de los tópicos mayormente tratados y aquellos con menor presencia. Cabe señalar que la informetría es una herramienta metodológica que “abarca el estudio de los aspectos cuantitativos de la información, independientemente de la forma en que aparezca registrada y del modo en que se genere” (Tague-Sutcliffe, 1994: s.p.; Almind y Ingwersen, 1997, p. 405).

Finalmente, traducidos los cuestionamientos investigativos en objetivos, nos proponemos: investigar y analizar, informétricamente, la producción científica en la intersección temática de la Ciencia de Datos (CD) y la Ciencia de la Información (CI), e incursionar en el análisis del discurso. Para cumplir con los objetivos, el artículo se estructura de la manera siguiente: anotaciones conceptuales: CI y CD; método; resultados; discusión, y; conclusiones y propuestas para futuras investigaciones.

2. ANOTACIONES CONCEPTUALES: CI Y CD

Posteriormente a la Segunda Guerra Mundial la CI, al igual que otras áreas del conocimiento, por ejemplo, la ciencia computacional y las investigaciones de operaciones, tiene su origen por las innovaciones tecnológicas (Saracevic, 1996, p. 42) y la consecuente producción de datos e información documental científica y tecnológica. La explosión de información durante los periodos de guerra, detonan la necesidad de organizarla y sistematizarla.

La CI se encuentra en un continuo esfuerzo para definirla o redefinirla, e incluso para establecer sus alcances y su relación con otras áreas del conocimiento. Por ejemplo, actualmente, se analiza el mapeo de la CI en bases de datos como LISA (Figuerola, Marco & Pinto, 2017) o bien se indaga la relación de la CD y las Ciencias Bibliotecaria y de la Información (Virkus y Garoufallou, 2020; 2019), o se buscan las equivalencias de la CD y la CI para la actualización curricular (Wang, 2018; Ortiz-Repiso, Greenberg y Calzada-Prado, 2018); incluso, se indaga cómo se aplica la CI en contextos geográficos específicos, China por ejemplo (Zhou, Yu, y Zhang, 2020). Dada la inestabilidad para la adopción conceptual de la CI, a continuación, se seleccionan algunos autores que nos remiten a los primeros esfuerzos conceptuales, definiciones con las cuales contextualizan y se apegan a los objetivos de esta investigación.

2.1 Ciencia de la información (CI)

La CI, es un área del conocimiento con características o naturaleza interdisciplinaria (Saracevic, 1995; Bicalho y Oliveira, 2011), y adicionalmente, puede ser multi, y transdisciplinaria (Martínez Musiño, 2012, p. 8.). Sin embargo, la CI y sus conceptualizaciones contienen variantes según la época o el grado de desarrollo de otros campos del conocimiento (Paul, Bhuimali y Aithal, 2017, p. 57). Por lo tanto, dada la amplitud y diversidad de definiciones de la CI, se seleccionaron solo algunas que, a nuestra consideración, frecen un mejor soporte para el cumplimiento de este artículo.

En primer lugar, H. Borko (1968, p. 3) encuentra que la CI investiga las propiedades y el comportamiento de la información, su uso y transmisión, así como su procesamiento para su óptimo almacenamiento y recuperación. En segundo término, Tefko Saracevic (1999, p. 1052), señala que la CI mantiene una vinculación con la tecnología, pero también tiene una dimensión social y humana, afirmando, además, que la CI está lejos de terminar su evolución. Finalmente, la tercera propuesta seleccionada es la de (Zins, 2007, p. 338), quien encuentra que la CI es el estudio de los datos, la información y el conocimiento, y cómo son utilizados por los individuos.

En las propuestas conceptuales de la CI vertidas en esta sección se observa la suma escalonada de elementos, por ejemplo, H. Borko y Tefko Saracevic señalan a la información como elemento aglutinador, en tanto que Chaim Zins al término información le añade los términos datos y conocimiento. En las definiciones de los tres autores seleccionados se aprecia una practicidad, es decir, a la información se le asocia con el uso o la transmisión, el estudio o el análisis, o bien a la tecnología y los individuos.

2.2 Ciencia de Datos (CD)

La CD, al igual que la CI tiene sus antecedentes en la aplicación de la tecnología en el análisis y procesamiento de los datos, así, Sirje Virkus y Emmanouel Garoufallou mencionan Peter Naur, quien en 1960 señalaba que la CD es la extracción generalizable del conocimiento a partir de los datos Virkus y Garoufallou, 2019, p. 424). No obstante, las definiciones son de reciente cuño. La CD esboza un campo multidisciplinario que ha experimentado un crecimiento considerable en años recientes (Da Sylva, 2017, p. 28).

Al igual que para la CI, seleccionamos algunas definiciones, por ejemplo: 1) la CD es el campo de conocimiento teórico y práctico que habla no solo sobre el origen de los datos / información, sino también sobre la representación de los datos (Paul y Dey, 2017, p. 2); 2) la CD se entiende generalmente como el análisis computacional y cuantitativo de grandes conjuntos de datos para crear información y conocimiento [...] se deriva del uso de marcos metodológicos, procesos y herramientas que se utilizan para analizar datos y obtener conocimientos (Ortiz-Repiso, Greenberg y Calzada-Prado, 2018, p. 770), y; 3) la CD es un campo interdisciplinario emergente que se ocupa de identificar y extraer patrones valiosos de Big data, convertirlos en información y conocimiento a través del análisis y la minería de datos (Wang, 2018, p. 1244)

Lo que observa en estas definiciones es la vinculación de grandes cantidades de datos para generar información y conocimiento, por supuesto, mediante su organización, sistematización y el análisis e interpretación. Además, como todas las ciencias, la CD requiere de marcos metodológicos y teóricos, y por supuesto, el auxilio de las herramientas computacionales.

3. METODOLOGÍA

3.1 Enfoque y universo de estudio

Enfoque: Estudio descriptivo, de tipo cuantitativo, y de primera aproximación en el análisis del discurso escrito. *Universo de estudio:* Se analizaron todos los registros recuperados de la "Colección principal" WoS mediante la configuración de búsqueda "Data science" and "information science", en la sección "por tema".

3.2 Recolección de datos

Al momento de ejecutar la búsqueda en WoS, se recuperaron 49 referencias documentales de la "Colección principal" en el periodo comprendido desde 1900 al 6 de noviembre de 2020 y abarcaron las siguientes bases de datos: *Science Citation Index Expanded (SCI-Expanded)*, *Social Sciences Citation Index (SSCI)*, *Arts & Humanities Citation Index (A&HCI)*, *Emerging Sources Citation Index (ESCI)*, *Conference Proceedings Citation Index - Science (CPCI-S)*, *Conference Proceedings Citation Index - Social Sciences & Humanities (CPCI-SSH)*, *Book Citation Index- Science (BKCI-S)* y *Book Citation Index-Social Sciences & Humanities (BKCI-SSH)*.

3.3 Procedimiento de análisis

Para la extracción de los 49 registros bibliográficos se utilizaron las funciones de WoS, trasladándose posteriormente a una hoja de cálculo (*Microsoft Excel*). Dichos registros se prepararon para el análisis y elaborar las gráficas respectivas. Posteriormente, se utilizó el programa *Voyant-Tools* para el análisis del discurso escrito y su representación gráfica. Finalmente, se procedió a la redacción de los resultados, la discusión y las conclusiones.



3.4 Alcances y limitaciones del estudio

WoS reporta distintos datos, sin embargo, para el desarrollo de este artículo se seleccionaron las variables: a) cantidad y tipo de documentos (*Document Type*); b) año de publicación (*Publication Year*); c) títulos de recursos (*Source Title*); d) citas recibidas (*Times Cited, WoS Core*); e) adscripción institucional (*Addresses*); f) Idioma de las investigaciones (*Language*); g) cantidad de autores por documento; h) primera categoría temática (*WoS Categories*), e; i). Índice H, el cual representa “un indicador que a partir del balance entre el número de publicaciones y las citas a éstas” (Hirsch y Buela-Casal, 2014, p 161). En tanto que, para el análisis del discurso se seleccionó la totalidad de investigaciones recuperadas de la variable Título del artículo (*Article Title*) y como criterio de homogeneidad se seleccionaron los títulos en el idioma inglés.

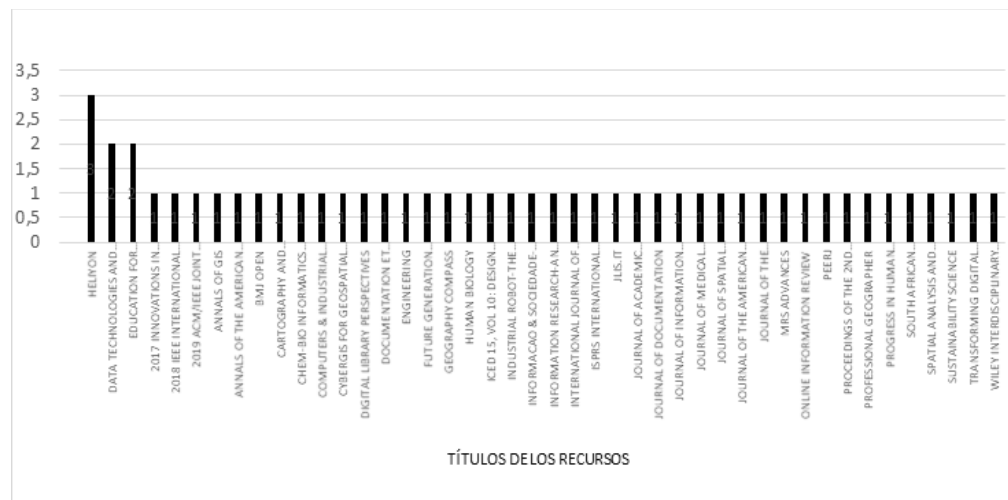
4. RESULTADOS

4.1 Métricas generales

Después de ejecutar la búsqueda de registros documentales que tratan los tópicos en CD y CI en la “Colección principal” de WoS se recuperaron 49 referencias (Anexo), mismas que se publicaron cronológicamente de la manera siguiente: 2015, 1; 2016, 2; 2017, 5; 2018, 12; 2019, 13, y; 2020, 16. En el periodo 2015-2017 la cantidad de investigaciones se duplicó cada año y la proporción de estudios publicados en los años 2020 y 2015 es de 16 a 1. Según su tipo de documento, se recuperaron 38 artículos (78 %) en los cuales se incluyen revisiones y notas editoriales, 7 textos publicados en memorias de congresos (14 %), 2 capítulos de libro (4 %) y 2 materiales editoriales (4 %).

Por otro lado, del total de las investigaciones que representan la fusión de la CD y CI, se han publicado en 45 títulos de recursos documentales (Figura 1). Aquellos títulos en los cuales hay repetibilidad de artículos incluidos son: *Heliyon*, *Data Technologies and Applications* y *Education for Information* con tres investigaciones (Tucker, Capps y Shamir, 2020; Arnaboldi y Azzone, 2020; Grimaldi, Diaz, Arboleda y Fernandez, 2019), con dos estudios (Virkus y Garoufallou, 2020; Virkus y Garoufallou, 2019) y 2 textos (Hagen, 2020; Casarosa, Ruggieri, Salvatori, Simi y Turbanti, 2020). Estas tres publicaciones con sus siete investigaciones incluidas recibieron dos citas, y representan solo el 1.56 %; los demás títulos de revistas o memorias de congresos cuentan con una publicación.

FIGURA 1.
DISTRIBUCIÓN DE LAS INVESTIGACIONES POR TÍTULO DE LOS RECURSOS
SEGÚN WOS, 2020

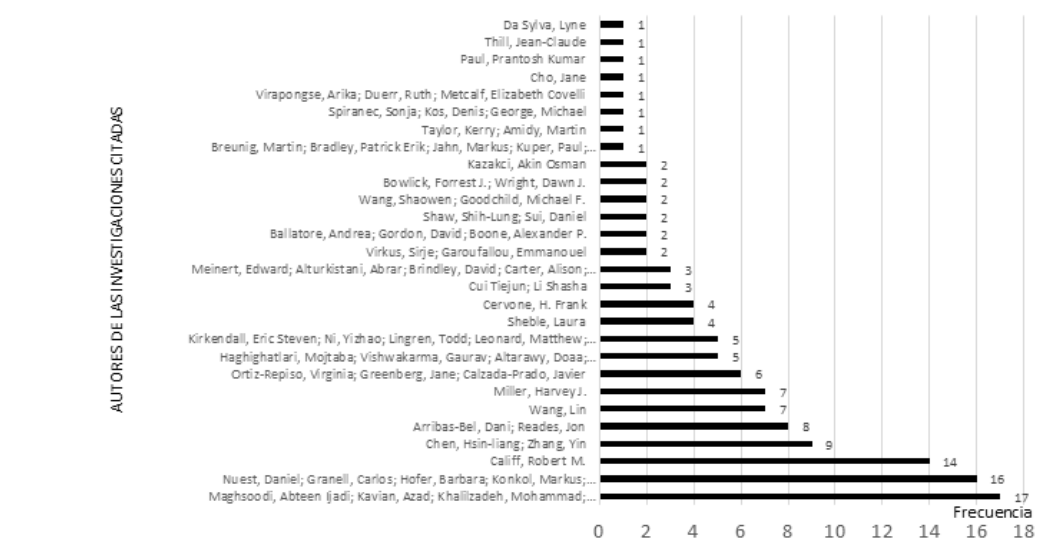


Fuente: Elaboración propia, 2020.

4.2 Análisis de citas

Las 49 investigaciones que tratan los temas CD y CI han recibido 128 citas (Figura 2 y Anexo), el conjunto de éstas, contienen una representatividad, de acuerdo con el Índice H, igual a 7. El promedio de citas recibidas en la totalidad de estudios recuperados es de 2.6. Cabe mencionar que no todas las publicaciones cuentan con citas, del total, 28 textos recibieron el 57.14 %, en tanto que los otros 21 estudios, 42.86 %, aún no cuentan con cita alguna. Con relación a la cantidad de citas, el artículo "*CLUS-MCDA: A novel framework based on cluster analysis and multiple criteria decision theory in a supplier selection problem*" (Maghsoodi et al. 2018) es aquel que recibió la mayor cantidad: 17; le sigue el documento "*Reproducible research and GIScience: an evaluation using AGILE conference papers*" (Nüst et al, 2018) con 16 citas y en tercer lugar, la investigación "*Future of Personalized Cardiovascular Medicine JACC State-of-the-Art Review*" (Califf, 2018) con 14 citas. Las 47 citas recibidas en estas tres investigaciones representan el 36.72 %, o en términos de proporciones es de casi la tercera parte; las 46 investigaciones restantes han recibido 81 citas, o lo que es lo mismo, representan el 73.28 %.

FIGURA 2.
DISTRIBUCIÓN DE LAS INVESTIGACIONES Y LA CANTIDAD DE CITAS RECIBIDAS, SEGÚN WOS, 2020



Fuente: Elaboración propia, 2020.

Ahorabien, la adscripción institucional de los autores de las tres investigaciones con la mayor cantidad de citas son las siguientes. En el artículo de Abteen Ijadi Maghsoodi et al (2018) participan autores con adhesión de dos países: Irán (Islamic Azad University y Amirkabir University of Technology) y Bélgica (University of Antwerp). En cuanto al artículo de Daniel Nüst et al (2018) se observa mayor cantidad de autores, países e instituciones participantes: Alemania (University of Münster), España (Universitat Jaume I de Castellón), Australia (University of Salzburg) y Holanda (University of Twente y Delft University of Technology). Finalmente, el autor del artículo de Robert M. Califf mantiene su adscripción en Duke University School of Medicine, Durham, North Carolina, Estados Unidos de América.

4.3 Autoría e idioma

Los 28 textos que abordan la CD y la CI y recibieron citas, han sido elaborados de manera colectiva (67.2 %) o individual (32.8 %). Por un lado, aquellos artículos con más integrantes participantes y citas recibidas son: “*Geospatial Data Management Research: Progress and Future Directions*” (Breunig et al., 2020) con nueve participantes y 1 cita recibida; “*ChemML: A machine learning*

and informatics program package for the analysis, mining, and modeling of chemical and materials data" (Haghighatlari et al., 2020), con ocho investigadores; 5 citas recibidas, y; *"Reproducible research and GIScience: an evaluation using AGILE conference papers"* (Nüst et al, 2018) con siete firmantes y 16 citas recibidas.

Por otra parte, los autores que prefieren escribir de manera individual son los siguientes: Robert M. Califf (2018), 14 citas; Lin Wang (2018), 7 citas; Harvey J. Miller (2017), 7 citas; Laura Sheble (2016), 4 citas; H. Frank Cervone (2016), 4 citas; Akin Osman Kazakci (2015), 2 citas. En tanto que los autores Jane Cho (2019), Prantosh Kumar Paul (2018), Jean-Claude Thill (2018) y Lyne Da Sylva (2017) recibieron una cita, respectivamente. Entonces, con los resultados expuestos apreciamos que el trabajo colectivo y la firma multiautoral ofrecen mejores dividendos que las investigaciones individuales, es decir, obtuvieron mayor cantidad de citas y por lo tanto mayor presencia en la documentación científica que tratan el tópico CD y CI.

Respecto al idioma de los documentos científicos localizados, cuatro de los 49 registros, representan el 8.2%, están escritos en idioma distinto al inglés. Estos estudios son *"Ciência de Dados e Ciência da Informação"* de Gustavo Henrique de Araújo Freire e Isa Maria Freire (2019), *"Les données et leurs impacts théoriques et pratiques sur les professionnels de l'information"* de Lyne Da Sylva (2017), *Exploration of Education and Research on Drug Informatics* de Mayumi Mochizuki (2020) y *"Birth and Development of Data Librarianship"* de Rossana Morriello (2020), redactados en portugués, francés, japonés e italiano. Aunque hay un predominio de la lengua inglesa como medio natural de la comunicación científica, se observa la presencia de idiomas distintos que cumplen con las exigencias y requerimientos normativos de las revistas y las bases de datos especializadas.

4.4 Análisis del discurso

El título representa de manera transversal, el o los contenidos, de las investigaciones, generalmente asignado por los autores o con cambios sugeridos por los revisores (árbitros) o los editores. Con los títulos de los registros documentales, es posible, también, profundizar en la obtención de métricas. Así, al seleccionar los títulos por año y analizarlos obtuvimos los siguientes resultados. En el año 2015 el promedio de palabras asignado al título es de ocho; en 2016 los 2 documentos contienen en promedio 11 palabras; para 2017 para la suma de los 5 títulos, el promedio es de 10.6 palabras; en tanto que los 12 títulos publicados en 2018 contienen en promedio 12.4 palabras, y, finalmente, los 13 estudios contienen en promedio 12.41 en sus títulos (Tabla 1).

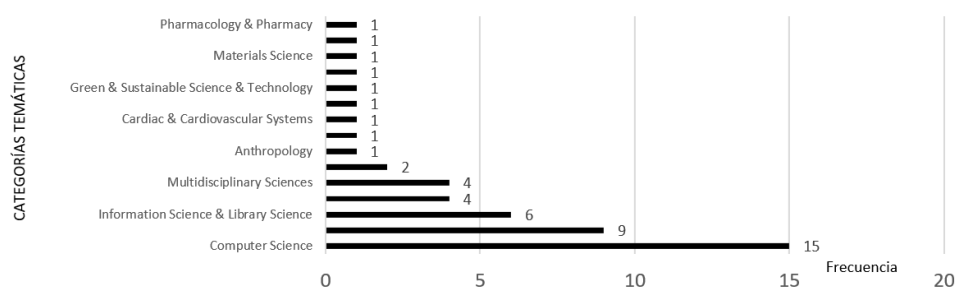
TABLA 1.
MÉTRICA DE LOS TÍTULOS DE LAS INVESTIGACIONES EN CD Y CI, EXTRAÍDOS DE WOS, PERIODO 2015-2020

AÑO	TÍTULOS	PALABRAS	PROMEDIO PALABRAS / TÍTULO / AÑO	PALABRAS MEJOR POSICIONADAS (FRECUENCIA)
2015	1	8	8.00	a (1); as (1); data (1); design (1); for (1)
2016	2	22	11.00	information (2); science (2); data (1); diffusion (1); informatics (1)
2017	5	53	10.60	data (5); information (3); analysis (2); geographic (2); professionals (2)
2018	12	149	12.41	data (4); information (4); science (4); analysis (3); evaluation (2)
2019	13	127	9.76	data (11); science (6); information (5); analysis (3); dimensions (2)
2020	16	186	11.62	data (10); science (8); information (4); informatics (3); materials (3)
Total	49	545		

Fuente: elaboración propia con datos de WoS, 2020.

Las palabras (y su frecuencia) con mayor aparición en los títulos es la siguiente: 2016, *information* (2); *science* (2); *data* (1); *diffusion* (1); *informatics* (1), y; 2017, *data* (5); *information* (3); *analysis* (2); *geographic* (2); *professionals* (2), En estos dos años se observa una alternancia por el primer lugar de los términos *information* y *data* (Figura 3). En periodo 2018-2020, con más títulos publicados, los primeros términos en orden de aparición son los siguientes: 2018, *data* (4); *information* (4); *science* (4); *analysis* (3); *evaluation* (2); 2019, *data* (11); *science* (6); *information* (5); *analysis* (3); *dimensions* (2), y; 2020, *data* (10); *science* (8); *information* (4); *informatics* (3); *materials* (3). En la etapa 2017-2020 se observa una superposición de la palabra *data* al término *information* el cual forma parte de la dupla CI (Figura 3).

FIGURA 4.
DISTRIBUCIÓN DE LOS DOCUMENTOS POR CATEGORÍA TEMÁTICA PRINCIPAL, SEGÚN WOS, 2020



Fuente: Elaboración propia, 2020.

5. DISCUSIÓN

Conceptualmente, los datos son el origen de la información, por supuesto, después de su análisis respectivo, y así, el exceso de la información para su compilación, organización y el análisis como estudio fenomenal son el ámbito de investigación de la CI. Por otra parte, los datos, también, en grandes cantidades son el objeto de estudio de la CD, claro, con el apoyo de herramientas computacionales. Además, los datos y la información se asocian a un tercer término, el *conocimiento*, que, también, son objeto de estudio, tanto de la CD y la CI.

Con hallazgos encontrados a partir del análisis de los conceptos de la CD y la CI y los resultados en esta investigación es que, ambas, son áreas del conocimiento, intra, multi e intra disciplinares. También, la CD y la CI mantienen un carácter eminentemente práctico e instrumental y requieren de herramientas computacionales (hardware y software) para el análisis e interpretación de grandes volúmenes de datos e información.

Por otra parte, se observa que hay diferencia temporal en la conceptualización, primeramente, aparecen definiciones de la CI y posteriormente los conceptos de la CD. Sin embargo, de acuerdo con los hallazgos y el análisis de las referencias documentales, y del análisis del discurso, actualmente aparecen más los términos datos y ciencia, más que información y ciencia, con lo cual se puede inferir que la CD proporcionará mayor cantidad de documentos científicos que la CI.

No obstante, con la diferencia de palabras encontradas, producto del análisis del discurso, la cantidad de definiciones o las publicaciones científicas, coincidimos que existe una frontera de estudios en el campo de la información [... y a la vez], un vínculo entre CD y la comunicación científica, objeto de interés en el campo de los estudios de la CI (Freire y Freire, 2019, p. 3).

6. CONCLUSIONES

Con la aplicación de las herramientas de búsqueda de WoS y la aplicación de técnicas que la informetría ofrece, es posible: 1) conocer la producción científica en las cuales se intersectan temáticamente la CD y de la CI; 2) identificar a los autores y sus publicaciones más relevantes o de mayor calidad, según su impacto en el circuito de las comunidades científicas, a través de las citas recibidas; 3) reafirmar que la comunicación científica que tratan la CD y la CI circula en lengua inglesa, pero no de manera exclusiva, también hay documentos en portugués, francés y japonés; 4) saber que no siempre a mayor cantidad de investigaciones por autor se refleja mayor cantidad de citas recibidas.

Por otra parte, es posible confirmar cuáles son las temáticas o áreas del conocimiento con las cuales se intersectan la CD y la CI. Así, concluimos que, 5) las investigaciones y los tópicos con mayor visibilidad son las ciencias computacionales, la geografía y la CBI; 6) el conocimiento de esta situación permite apuntalar estas temáticas o impulsar aquellas con menor cantidad de estudios; 7) la CD y la CI son áreas del conocimiento eminentemente prácticas, asociadas a verbos como coleccionar o compilar, gestionar, analizar o interpretar y tomar decisiones; 8) ambas disciplinas, la CD y la CI, también han creado teorías, métodos y técnicas que las impulsan para categorizarse y solidificar su estatus de ciencia; 8) a la CD y la CI se les puede considerar metaciencias en la medida que realizan coberturas, intra, trans y multidisciplinarias porque muestran flexibilidad y transversalidad con otras áreas del conocimiento.

La comunicación de saberes científicos la proporciona y facilitan servicios ofrecidos por WoS, y otros aglutinadores de bases de datos científicas. Con WoS, según los resultados obtenidos, es posible incursionar en el análisis del discurso escrito. Así, aunque el periodo de análisis de los términos que componen los títulos de las investigaciones analizadas es corto, es posible inferir que, en un futuro cercano habrá una superposición de los estudios de la CD sobre aquellos de la CI.

6.1 Propuestas para futuras investigaciones

WoS es una fuente de información que aglutina las referencias documentales de investigaciones de gran valor académico y científico, pero no es la única, hay otras posibilidades. WoS es un punto de partida de las comunidades académicas como marco de referencia para la evaluación de los posicionamientos científicos, pero hay otros similares. WoS utiliza un lenguaje que facilita la comunicación, el inglés, pero en términos de agrupar e incluir en sus índices documentos de la CD y la CI en otros idiomas (portugués, italiano, francés y japonés) nos muestra alternancias lingüísticas en la comunicación científica.

Para este artículo, se seleccionaron solo algunas variables para el análisis informétrico, también, sería posible seleccionar otras y ejecutar el estudio de otras, o bien realizar el cruce entre ellas. WoS es un aglutinador de bases de datos con registros bibliográficos referenciales, por lo tanto, sería factible recuperar los documentos con sus contenidos completos y aplicar el análisis

del discurso, ya sea de sus contenidos escritos, o bien en el uso de imágenes, por ejemplo.

La informetría genera, principalmente, datos para el análisis cuantitativo, también, en cada área del conocimiento, los especialistas pueden profundizar en el análisis de los contenidos propios. En este artículo se consultó y recuperaron referencias documentales de WoS, sin embargo, hay otras fuentes de información, *Scopus*, *Schoolar Google* o *Microsoft Academic*, por mencionar algunas, en las cuales se pueden realizar estudios informétricos y del análisis del discurso.

7. REFERENCIAS

- Almind, T.C. e Ingwersen, P. (1997). Informetric Analyses on the World Wide Web: Methodological Approaches to Webometrics. *Journal of documentation*, 53(4), 404-426.
- Arnaboldi, M., y Azzone, G. (2020). Data Science in the Design of Public Policies: Dispelling the Obscurity in Matching Policy Demand and Data Offer. *Heliyon*, 6(6), 13, e04300. DOI: <https://doi.org/10.1016/j.heliyon.2020.e04300>
- Bicalho, L., y Oliveira, M. D. (2011). A teoria e a prática da interdisciplinaridade em Ciência da Informação. *Perspectivas em Ciência da Informação*, 16(3), 47-74. Recuperado de <https://www.scielo.br/pdf/pci/v16n3/04.pdf>
- Borko, H. (1968). Information Science: What is it? *American Documentation*, 19(1), 3-5.
- Breunig, M., Bradley, P. E., Jahn, M., Kuper, P., Mazroob, N., Rosch, N., Al-Doori, M., Stefanakis, E., y Jadidi, M. (2020). Geospatial Data Management Research: Progress and Future Directions. *ISPRS International Journal of Geo-Information*, 9(2), 20, 95. DOI: <https://doi.org/10.3390/ijgi9020095>
- Califf, R. M. (2018). Future of Personalized Cardiovascular Medicine JACC State-of-the-Art Review. *Journal of the American College of Cardiology*, 72(25), 3301-3309. DOI: <https://doi.org/10.1016/j.jacc.2018.09.079>
- Casarosa, V., Ruggieri, S., Salvatori, E., Simi, M., & Turbanti, S. (2020). Educational Ecosystems for Information Science: the Case of the University of Pisa. *Education for Information*, 36(2), 119-138. DOI: <https://doi.org/10.3233/efi-190330>
- Cervone, H. F. (2016). Informatics and Data Science: An Overview for the Information Professional. *Digital Library Perspectives*, 32(1), 7-10. DOI: <https://doi.org/10.1108/dlp-10-2015-0022>
- Cho, J. (2019). Subject Analysis of LIS Data Archived in a Figshare Using Co-occurrence Analysis. *Online Information Review*, 43(2), 256-264. DOI: <https://doi.org/10.1108/oir-12-2017-0369>

- Da Sylva, L. (2017). Les données et leurs impacts théoriques et pratiques sur les professionnels de l'information [The Theoretical and Practical Impact of Data on Information Professionals]. *Documentation Et Bibliothèques*, 63(4), 5-34. DOI: <https://doi.org/10.7202/1042308ar>
- Figuerola, C. G., Marco, F. J. G., & Pinto, M. (2017). Mapping the Evolution of Library and Information Science (1978–2014) Using Topic Modeling on LISA. *Scientometrics*, 112(3), 1507-1535. DOI: <https://doi.org/10.1007/s11192-017-2432-9>
- Freire, G. H. D., y Freire, I. M. (2019). Ciência de Dados e Ciência da Informação [Data Science and Information Science]. *Informação & Sociedade: Estudos*, 29(3), 3-4. Recuperado de <https://brapci.inf.br/index.php/res/download/147968>
- Grimaldi, D., Diaz, J., Arboleda, H., y Fernandez, V. (2019). Data Maturity Analysis and Business Performance. A Colombian Case Study. *Heliyon*, 5(8), 9, e02195. DOI: <https://doi.org/10.1016/j.heliyon.2019.e02195>
- Hagen, L. (2020). Teaching Undergraduate Data Science for Information Schools. *Education for Information*, 36(2), 109-117. DOI: <https://doi.org/10.3233/efi-200372>
- Haghighatlari, M., Vishwakarma, G., Altarawy, D., Subramanian, R., Kota, B. U., Sonpal, A., Setlur, S., y Hachmann, J. (2020). ChemML: A Machine Learning and Informatics Program Package for the Analysis, Mining, and Modeling of Chemical and Materials Data. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 10(4), 10. DOI: <https://doi.org/10.1002/wcms.1458>
- Hirsch, J. E., y Bucla-Casal, G. (2014). The Meaning of the H-index. *International Journal of Clinical and Health Psychology*, 14(2), 161-164. DOI: [https://doi.org/10.1016/S1697-2600\(14\)70050-X](https://doi.org/10.1016/S1697-2600(14)70050-X)
- Kazakci, A. O. (2015). Data Science as a New Frontier for Design. En C. Weber, S. Husung, M. Cantamessa, G. Cascini, D. Marjanovic, & S. Venkataraman (Eds.), *Design (ICED 15) Vol 10: Design Information and Knowledge Management Milan, Italy, 27-30.07.15*. Congreso organizado por Mines ParisTech, Milan, Italia. Recuperado de <https://www.designsociety.org/download-publication/37969/DATA+SCIENCE+AS+A+NEW+FRONTIER+FOR+DESIGN>
- Maghsoodi, A. I., Kavian, A., Khalilzadeh, M., & Brauers, W. K. M. (2018). CLUS-MCDA: A Novel Framework Based on Cluster Analysis and Multiple Criteria Decision Theory in a Supplier Selection Problem. *Computers & Industrial Engineering*, 118, 409-422. DOI: <https://doi.org/10.1016/j.cie.2018.03.011>
- Martínez Musiño, C. (2012). La ciencia de la información como plataforma para potenciar el estudio de los flujos de la información en las organizaciones. *e-Ciencias de la Información*, 2(1), 1-14. DOI: <https://DOI.10.15517/ECI.V2I1.1212>
- Miller, H. J. (2017). Geographic Information Science I: Geographic Information Observatories and Opportunistic GIScience. *Progress in Human Geography*, 41(4), 489-500. DOI: <https://doi.org/10.1177/0309132517710741>

- Mochizuki, M. (2020). Exploration of Education and Research on Drug Informatics. *Yakugaku Zasshi-Journal of the Pharmaceutical Society of Japan*, 140(4), 543-554. DOI: <https://doi.org/10.1248/yakushi.19-00250>
- Morriello, R. (2020). Birth and Development of Data Librarianship. *Jlis.It*, 11(3), 1-15. DOI: <https://doi.org/10.4403/jlis.it-12653>
- Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., y Cerutti, V. (2018). Reproducible Research and GIScience: An Evaluation Using AGILE Conference Papers. *Peerj*, 6(23), e5072. DOI: <https://doi.org/10.7717/peerj.5072>
- Ortiz-Repiso, V., Greenberg, J., y Calzada-Prado, J. (2018). A Cross-Institutional Analysis of Data-Related Curricula in Information Science Programmes: A Focused Look at the iSchools. *Journal of Information Science*, 44(6), 768-784. DOI: <https://doi.org/10.1177/0165551517748149>
- Paul, P., Bhuimali, A., y Aithal, P. S. (2017). Information Science: Science or Social Science? *International Journal on Recent Researches in Science, Engineering & Technology*, 5(9), 54-65. Recuperado de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3042150
- Paul, P. K. (2018). The Context of IST for Solid Information Retrieval and Infrastructure Building: Study of Developing Country. *International Journal of Information Retrieval Research*, 8(1), 86-100. DOI: <https://doi.org/10.4018/ijirr.2018010106>
- Paul, P. K. y Dey, J. L. (2017). Data Science Vis-a-Vis Efficient Healthcare and Medical Systems: A Techno-Managerial Perspective. En *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, 2017, Congreso organizado por IEEE, Vellore, India. DOI: [10.1109/IPACT.2017.8245148](https://doi.org/10.1109/IPACT.2017.8245148).
- Saracevic, T. (1999). Information Science. *Journal of the American Society for Information Science*, 50(12), 1051-1063. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:12<1051::AID-ASIS2>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(1999)50:12<1051::AID-ASIS2>3.0.CO;2-Z)
- Saracevic, T. (1996). Ciência de Informação: Origem, evolução e relações. Perspectivas em Ciência da Informação, *Belo Horizonte*, 1(1), 41-46. Recuperado de <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>
- Saracevic, T. (1995). Interdisciplinary Nature of Information Science. *Ciência da informação*, 24(1), 36-41. Recuperado de https://brapci.inf.br/repositorio/2010/03/pdf_dd085d2c4b_0008887.pdf
- Sheble, L. (2016). Research Synthesis Methods and Library and Information Science: Shared Problems, Limited Diffusion. *Journal of the Association for Information Science and Technology*, 67(8), 1990-2008. DOI: <https://doi.org/10.1002/asi.23499>
- Tague-Sutcliffe, J. 1994. Introducción a la informetría. *Acimed*, 2(3), 26-35.
- Thill, J. C. (2018). Innovations in GIS&T, Spatial Analysis, and Location Modeling. En J. C. Thill (Ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems* (pp. 1-6). Alemania: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-642-37896-6_1

- Tucker, E. C., Capps, C. J., y Shamir, L. (2020). A Data Science Approach to 138 Years of Congressional Speeches. *Heliyon*, 6(8), 8, e04417. DOI: <https://doi.org/10.1016/j.heliyon.2020.e04417>
- Virkus, S., y Garoufallou, E. (2020). Data Science and Its Relationship to Library and Information Science: A Content Analysis. *Data Technologies and Applications*, 21. DOI: <https://doi.org/10.1108/dta-07-2020-0167>
- Virkus, S., y Garoufallou, E. (2019). Data Science from a Library and Information Science Perspective. *Data Technologies and Applications*, 53(4), 422-441. DOI: <https://doi.org/10.1108/dta-05-2019-00760>
- Wang, L. (2018). Twinning Data Science with Information Science in Schools of Library and Information Science. *Journal of Documentation*, 74(6), 1243-1257. DOI: <https://doi.org/10.1108/JD-02-2018-0036>
- Zhou, X., Yu, X. y Zhang, L. (2020). The Definitions of Information Science in China. En *Multidisciplinary Digital Publishing Institute Proceedings*, 47(1). DOI: <https://doi.org/10.3390/proceedings2020047045>
- Zins, C. (2007). Conceptions of Information Science. *Journal of the American Society for Information Science and Technology*, 58(3), 335-350.

