

Mapeando la datificación: Un análisis bibliométrico de perspectivas y patrones en la comunicación científica

Gastón Becerra¹

gaston.becerra@uflouniversidad.edu.ar

Joaquín Mezzadra²

joaquin.mezzadra@uflouniversidad.edu.ar

Bruno Gambino³

bgambino@sociales.uba.ar

Fecha de recepción: 11 de abril de 2025

Fecha de aceptación: 21 de mayo de 2025

ARK CAYCYT: <https://id.caicyt.gov.ar/ark://hovrgbg8a>

Resumen

Este artículo emplea una perspectiva bibliométrica para analizar la comunicación científica sobre la datificación. Nuestro corpus se compone de 904 documentos recuperados de la base de Scopus a través de los criterios “datafication” o “datafied” y limitando a sólo artículos en revistas y en idioma inglés. En diálogo con la revisión sistemática de Flensburg y Lomborg (2023) como principal antecedente, y respondiendo a cuatro preguntas de investigación, exploramos temas, trabajos influyentes, canales de comunicación, contribuciones geográficas y colaboraciones entre autores.



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/deed.es)

¹ Conicet / Universidad de Flores (Ciudad de Buenos Aires, Argentina)
<https://orcid.org/0000-0001-9432-8848>

² Universidad de Flores (Ciudad de Buenos Aires, Argentina) / Universidad de Valencia (Valencia, España)
<https://orcid.org/0009-0000-7962-4307>

³ Universidad de Buenos Aires (Ciudad de Buenos Aires, Argentina)
<https://orcid.org/0009-0004-6754-1925>

Nuestros análisis sugieren el predominio de una mirada crítica y deconstructiva del big data, y una relación proceso-producto con la datificación. El foco desarrollado por los referentes más citados respalda esta tendencia, destacando el trasfondo sociopolítico de la datificación. El análisis geográfico revela una concentración de la producción en Estados Unidos y Europa, por lo que subrayamos la necesidad de ampliar la diversidad de voces y contextos para comprender la datificación como un proceso global con devenires locales. En conclusión, este estudio contribuye a una mirada más compleja de la literatura sobre datificación, ofreciendo una actualización hasta 2023, destacando la relevancia de la crítica en el campo y abogando por la inclusión de perspectivas diversas para enriquecer la investigación futura.

Palabras clave: Datificación; Big data; Bibliométricos; Estudios críticos de datos

Abstract

This article employs a bibliometric perspective to analyze scientific communication on datafication. Our corpus consists of 904 documents retrieved from the Scopus database using the criteria "datafication" or "datafied" and limited to articles in journals and in the English language. In dialogue with the systematic review by Flensburg and Lomborg (2023) as the main reference, and addressing four research questions, we explore themes, influential works, communication channels, geographical contributions, and author collaborations. Our analyses suggest the predominance of a critical and deconstructive view of big data, along with a process-product relationship with datafication. The focus developed by the most cited references supports this trend, highlighting the socio-political background of datafication. Geographical analysis reveals a concentration of production in the United States and Europe, underscoring the need to broaden the diversity of voices and contexts to understand datafication as a global process with local variations. In conclusion, this study contributes to a more nuanced understanding of the literature on datafication, providing an update until 2023, emphasizing the relevance of critique in the field, and advocating for the inclusion of diverse perspectives to enrich future research.

Keywords: datafication; big data; bibliometrics; critical data studies

Resumo

Este artigo adota uma perspectiva bibliométrica para analisar a comunicação científica sobre dataficação. Nosso corpus é composto por 904 documentos extraídos da base de dados Scopus, utilizando os critérios "datafication" ou "datafied" e limitado a artigos em periódicos e em língua inglesa. Em diálogo com a revisão sistemática de Flensburg e Lomborg (2023) como referência central, e abordando quatro questões de pesquisa, exploramos temas, obras influentes, canais de comunicação, contribuições geográficas e colaborações entre autores. Nossas análises sugerem a predominância de uma visão crítica e desconstrutiva sobre big data, além de uma relação processo-produto com a dataficação. O enfoque desenvolvido pelas referências mais citadas reforça essa tendência, destacando o pano de fundo sociopolítico da dataficação. A análise geográfica revela uma concentração da produção nos Estados Unidos e Europa, evidenciando a necessidade de ampliar a diversidade de vozes e contextos para entender a dataficação como um processo global com variações locais. Em conclusão, este estudo contribui para uma compreensão mais matizada da literatura sobre dataficação, atualizando-a até 2023, enfatizando a relevância da crítica no campo e defendendo a inclusão de perspectivas diversas para enriquecer pesquisas futuras.

Palavras-chave: dataficação; big data; bibliometria; estudos críticos de dados

Introducción⁴

En este trabajo ponemos el foco en un proceso que, si bien se puede rastrear hasta la modernidad, se ha comenzado a investigar con más fuerza en las últimas décadas: nos referimos a la datificación –“datafication” en inglés– por el cual distintas esferas de nuestra vida social y subjetiva adoptan la forma de datos. El concepto fue introducido y popularizado en 2013 por Viktor Mayer-Schonberger y Kenneth Cukier en su bestseller “Big data. A revolution that will transform how we live, work, and think”, y desde entonces se le han dedicado trabajos conceptuales y empíricos desde distintas áreas del conocimiento, así como revisiones sistemáticas que buscan dar una imagen general del campo de estudio. Específicamente, aquí queremos aportar a tipo de

⁴ Los autores agradecen los comentarios y sugerencias de los revisores, que ayudaron a mejorar la claridad de la presentación final del estudio.

objetivos con una caracterización de la comunicación científica en torno a la datificación, y para ello, adoptamos la perspectiva bibliométrica, que se fundamenta en el análisis cuantitativo, de alcance descriptivo, sobre los patrones de publicación científica.

El artículo se compone de las siguientes secciones. En esta introducción ofrecemos una breve presentación teórica, señalando dos usos del concepto: uno más estrecho, que lo vincula con la generación y manipulación de datos sobre un fenómeno de la realidad para extraer algún tipo de valor; y un uso más amplio que alcanza a la reflexión crítica del contexto social, político e ideológico que lo soporta, incorporando una mirada proveniente de las ciencias sociales. Luego, reseñamos nuestro principal antecedente en el objetivo de dar con una mirada de conjunto de la literatura de la datificación: la revisión sistemática realizada por Flensburg y Lomborg (2023), con quienes dialogamos para poner en contexto nuestros hallazgos. Esperamos que el fruto de este diálogo resulte en una imagen más completa de la literatura, sistematizando así tanto las ideas principales como su dimensión comunicativa. Cerramos la introducción explicitando nuestras preguntas de investigación. La siguiente sección refiere al método del estudio bibliométrico y detalla nuestras decisiones en materia de recolección y pre-procesamiento de los datos. La sección de resultados busca responder las preguntas planteadas. Y en la conclusión y discusión ponemos nuestros resultados en diálogo con los antecedentes, y explicitamos desafíos pendientes para avanzar la indagación sobre la datificación.

Sobre el concepto “datificación”

El término “datificación” se popularizó con el trabajo de Mayer-Schonberger y Cukier (2013), que lo introducen con la siguiente definición: “To datafy a phenomenon is to put it in a quantified format so it can be tabulated and analyzed” (2013, p. 48). A lo largo de la obra, los autores recorren la datificación de distintos dominios de nuestro mundo físico, social y subjetivo, argumentando que se trata de una revolución que está cambiando la naturaleza de dichas áreas. De acuerdo con los autores, se trata del proceso por el cual arribamos a la era del *big data*, que se caracterizaría por volúmenes de datos que exceden la capacidad clásica de cómputo, vincular fuentes varias y menos estructuradas, y permiten un manejo del tiempo distinto, al ofrecer bases para predicciones y manipulaciones que extraigan valor del dato en tiempo real. Este uso en tanto

proceso, cuyo resultado es el fenómeno “datificado” en vistas de un objetivo práctico, también se encuentra en trabajos más recientes, como el de Bibri y Krogstie (2020), quienes refieren al caso de las ciudades inteligentes [smart cities] o manejadas con datos [data-driven cities] para optimizar sus operaciones, funciones, servicios y políticas: “En términos generales, la datificación se refiere a las herramientas, procesos, métodos, técnicas y tecnologías colectivas que se utilizan para transformar una ciudad en una empresa basada en datos.” (2020, p. 5). Este tipo de definiciones se caracterizan por remarcar el potencial o la “promesa” del análisis de los grandes datos.

Otros autores han avanzado la mirada, observando al interior de este proceso, diseccionando las operaciones epistémicas necesarias para construir el dato (Baack, 2015; Lycett, 2013). Por caso, Lycett refiere a tres momentos –desmaterialización, licuificación, y densidad–, por los cuales los fenómenos son descontextualizados y abstraídos del mundo físico, manipulados y circulados para servir distintos propósitos, y re combinados para lograr un cierto output en un momento particular. En última instancia, la datificación es una forma de dar sentido al mundo a través de la tecnología de información.

Ampliar el foco de este proceso micro, a uno más macro que incluya el contexto que lo sustenta es el fundamento de una mirada más crítica de la datificación, como la que propone José van Dijck (2014). Esta autora explícitamente busca tomar distancia de la definición de Mayer-Schonberger y Cukier, al desnaturalizar el proceso de la datificación y presentarla como una estrategia de extracción, sostenida por un discurso ideológico con reclamos epistémicos cuestionables. Así, introduce la noción de “dataísmo”, aclarando que se trata de una “ideología”, y que define como:

“una creencia generalizada en la cuantificación objetiva y el seguimiento potencial de todo tipo de comportamiento humano y socialidad a través de tecnologías de medios en línea. Además, el dataísmo también implica confianza en los agentes (institucionales) que recopilan, interpretan y comparten (meta)datos extraídos de las redes sociales, plataformas de Internet y otras tecnologías de comunicación”. (van Dijck, 2014, p. 198).

En esta vena se encuentran también los trabajos que equiparan la datificación como un mandato o imperativo de convertir todos los fenómenos posibles en datos, o el supuesto acerca de que todo puede ser efectivamente medido y convertido en este formato para extraer valor (Kitchin,

2014b; Lupton & Williamson, 2017; Sadowski, 2019). Así, poniendo el foco en el trasfondo social de este mandato, Sadowski plantea una analogía entre la datificación y la acumulación originaria del capital y la lógica financiera: “De hecho, siguiendo los pasos de otras empresas extractivas a lo largo de la historia del capitalismo, como la apropiación de tierras y la minería de recursos... muchas de las prácticas ahora comunes de acumulación de datos deberían entenderse en términos de la práctica más contundente de extracción de datos, en la que los datos se toman sin consentimiento significativo y compensación justa para los productores y fuentes de esos datos” (2019, p. 7).

En estos casos el problema no es económico sino político, es decir, de relaciones sociales de dominación. Así, también Zuboff (2019a) –quien no usa explícitamente el término– se refiere a un “capitalismo de vigilancia” en el que la experiencia humana es el material que produce datos que pueden ser explotados para influenciar y predecir nuestras acciones, y así insertarnos como productos en un mercado de comportamientos. También, en clave geopolítica, Mejías y Couldry (2019a), identifican la datificación con el colonialismo de datos, el cual combinaría las prácticas predatorias y extractivas del colonialismo histórico, con la abstracción de los métodos de cuantificación y su manipulación a través del cómputo: “Comprender Big Data del Sur Global significa comprender la dependencia actual del capitalismo de este nuevo tipo de apropiación que funciona en todos los puntos del espacio donde las personas o las cosas están unidas a las infraestructuras de conexión actuales.” (Couldry & Mejias, 2019, p. 2). Siguiendo esta lógica, los autores llegan a utilizar el término para referir a las transformaciones sociales, económicas y políticas de la vida humana que habilitan estas prácticas extractivas: “Dado que la mayoría de los que escriben sobre datos también se preocupan por lo que sucede con la vida humana, el término “datificación” ha adquirido rápidamente un significado adicional: la transformación más amplia de la vida humana para que sus elementos puedan ser una fuente continua de datos.” (Mejias & Couldry, 2019, p. 2).

En resumen, desde su masificación el concepto parece haber recorrido un arco que va desde una mirada relativamente acrítica que pone el foco en su potencial económico y/o comercial, pasando por una lectura “micro” que descompone el proceso como un acto epistémico, hasta llegar a una mirada más “macro” que lo ubica en un contexto social y lo inserta en relaciones de poder,

explotación y dominación, creencias ideológicas y grandes transformaciones sociopolíticas.

La datificación como campo de estudios

La proliferación de sentidos en los que se utiliza la noción de datificación es indicativo de un mayor interés por dicho proceso, y con ello, la posible emergencia de un campo de estudios. Este es el punto de partida del trabajo de Sofie Flensburg y Stine Lomborg (2023), quienes realizaron una revisión sistemática de la literatura en la que mapean temas y campos, identifican perspectivas teóricas y preguntas empíricas, así como métodos de indagación, y sistematizan sus principales unidades de análisis y casos. Este será nuestro principal antecedente y el interlocutor para nuestros hallazgos, ya que, hasta donde conocemos, no hay estudios bibliométricos de la datificación – aunque sí contamos con varios antecedentes centrados en el *big data* (Ahmad et al. 2020; Liu et al. 2019; Kalantari et al. 2017), además de uno propio (Becerra & Ratovicius, 2022). Nuestro trabajo apunta a complementar el de estas autoras, ofreciendo una mirada bibliométrica que caracteriza a la literatura desde el punto de vista de sus comunicaciones, sus ámbitos de producción y colaboración, espacios de publicación, y principales referentes.

El estudio de Flensburg y Lomborg trabaja con un corpus de 463 publicaciones recuperadas en octubre del 2020 de los buscadores Web of Science y Scopus, bajo los criterios de “datafication” o “datafied*”, y quedándose sólo con documentos en idioma inglés. Sus principales preguntas son: ¿Cuáles son los campos de investigación dominantes y las áreas de actualidad para estudiar la datificación? ¿Qué tienen en común y en qué pueden diferir?; ¿Qué tipo de perspectiva teórica y preguntas de investigación se persiguen y con qué métodos y enfoques?; ¿Cuáles son los focos analíticos y las unidades de análisis más comunes para comprender la datificación?; y finalmente, ¿cómo podría nuestra comprensión del impacto de la datificación en la existencia humana y la vida social avanzar en investigaciones futuras mediante la polinización cruzada de conocimientos en el campo? Para responder estas preguntas, siguen un proceso analítico que supone dos momentos: por un lado, para responder la primera pregunta realizan un análisis de redes exploratorio de la información bibliográfica del corpus con el software VosViewer (van Eck & Waltman, 2010); para responder las siguientes avanzan con un análisis de contenido de cada artículo, que supuso una codificación manual en términos de sus campos de investigación, tipos de publicación, enfoque

metodológico, así como fuentes, casos y unidades de análisis.

Con respecto a sus resultados, las autoras mapean las 115 keywords totales de los artículos en forma de red, e identifican cinco clusters o grupos de trabajos: el primero corresponde a los estudios de medios y comunicaciones, donde la datificación refiere mayormente al concepto de “plataformización”; el segundo, proviene del ámbito educativo, para el cual la datificación es una evolución natural de las tareas evaluativas del campo; el tercero, es el de los trabajos que adoptan una mirada más social, indagando sus distintas implicaciones a través de varios campos; en el cuarto, destacan las miradas desde el derecho y las prácticas de datos, indagando la vigilancia y el control; mientras que el quinto, pone el foco en el problema de los medios y el tratamiento de las audiencias.

Luego, se disponen a describir los enfoques y métodos empleados, señalando el predominio de los trabajos teóricos, con un acumulado de 232/n=463; siguen los trabajos empíricos 115/463, que codificaron manualmente para detallar su diseño; otros tantos de corte metodológico, con 71/463; finalmente un conjunto marginal de trabajos de revisión sistemática, editoriales u otras. Volviendo a los empíricos: el 70% adoptan un diseño cualitativo, el 9% cuantitativos, y el restante 16% opta por diseños mixtos. El predominio de los estudios cualitativos se explica por el interés en observar actores, sentidos y prácticas en torno a los datos. Acerca de las unidades de análisis de los trabajos, las autoras señalan que hay dos grandes grupos: el de la infraestructura y la analítica de la datificación (que equiparan con el trabajo de Mayer-Schonberger y Cukier), y el de sus implicancias sociales. Las unidades de análisis recortadas varían en estos casos; en el primero, el foco se pone en la infraestructura, mientras que en el segundo se posa la mirada en los usuarios: este segundo es el que predomina con 326/463 artículos.

Estos resultados permiten caracterizar a la investigación sobre la datificación como campo en el que conviven dos miradas, que coinciden a grandes rasgos con nuestro arqueológico conceptual: una centrada en la técnica y la analítica (y sus promesas), y una interesada por su trasfondo social y político. También proponen que se trata de la emergencia de un tema de investigación –y un concepto– que cruza disciplinas. Como contraparte, entienden que la definición del concepto es poco clara y fragmentada, la investigación empírica se encuentra desbalanceada, y más importante, que existe una escisión en el campo. Cómo cruzar esta escisión entre humanismo/técnica, y cómo

acercar a los campos de estudio, son el principal desafío que se plantea para el campo, por lo que abogan por un desarrollo conceptual –desde la teoría de la comunicación– para unir las dimensiones de la infraestructura y la del usuario. Así, en sus conclusiones, abogan por:

“Una conceptualización comunicativa integral de la datificación, que considere tanto la tecnología material, los significados generados a través de los datos y las consecuencias socioculturales de la ubicua datificación de nuestra vida cotidiana, podría ofrecer un próximo paso prometedor en la integración y el avance de la investigación actual sobre la datificación. , además de fundamentar la datificación de manera más sistemática en una trayectoria histórica de configuración mutua de la tecnología de la comunicación y la sociedad” (Flensburg & Lomborg, 2023, p. 1466)

Nuestras preguntas de investigación

A nuestro entender, el trabajo de Flensburg y Lomborg (2023) ofrece un buen primer pantallazo global del estado, alcance, variedad y desafíos de la investigación sobre datificación. No obstante, su mirada tiene algunas limitaciones que debemos considerar para decidir un curso de acción para el nuestro.

Una primera limitación reside en que la caracterización ofrecida por las autoras no excede al producto de las investigaciones, y por ello no ofrece pistas sobre su proceso de producción. Es decir, se reportan tendencias temáticas, metodológicas, etc., pero no se dice nada acerca de quienes producen, desde donde, o en qué revistas. Tampoco se nos aclara cuáles son los autores más citados que utilizan el concepto, o los más referenciados de una literatura más amplia que no utiliza el concepto. Así, la dimensión comunicativa de la investigación de la datificación queda oculta. Por ello, en este trabajo apostamos por avanzar sobre esta limitación, adoptando una perspectiva bibliométrica, que atiende a las siguientes preguntas:

(RQ1) Qué temas y conceptos se identifican;

(RQ2) Cuáles son los trabajos más influyentes, y cuáles son las principales referencias;

(RQ3) Cuáles son los canales de comunicación;

(RQ4) Qué países y autores contribuyen a la literatura, y cómo colaboran.

Responder estas preguntas permitirá aportar a una mirada más completa sobre la producción y la comunicación científica de la datificación, identificando los temas y los aportes de mayor impacto en el campo, y mapeando las contribuciones y los canales de comunicación desde una perspectiva más amplia, a partir de un interés geográfico que nos permita ubicar ambas dimensiones del campo.

La segunda limitación que encontramos en el trabajo de Flensburg y Lomborg es que las dimensiones y preguntas planteadas por las autoras no se cruzan. Es decir, se nos informa de dos grandes concepciones de la datificación, de cinco grandes áreas o espacios, así como de distintos tipos de investigación, pero no se nos aclara cómo se relacionan estas miradas. Así, no se sugiere qué concepciones son predominantes en cada campo, o si en algún área varían diferencialmente los tipos de investigación. Esta falta de una mirada comparativa resulta en que la imagen totalizante de grandes tendencias en el campo, sin subdivisiones. Sin pretender desestimar la necesidad de atender a esta limitación, este es un objetivo que en el presente trabajo apenas empezamos a ensayar, dejándolo pendiente para futuros trabajos que combinen la perspectiva de la revisión sistemática y la bibliométrica.

Método

Construcción de corpus

Siguiendo la estrategia de Flensburg y Lomborg (2023), se realizó una búsqueda en Scopus con los keywords “datafication” o “datafied*”, limitando a sólo artículos en revistas y en idioma inglés, para poder asegurar una base de metadatos consistente. El criterio utilizado fue: *(TITLE-ABS-KEY (datafication) OR TITLE-ABS-KEY (datafied*)) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "english"))*. Esta estrategia excluye publicaciones que no incluyen los términos explícitamente en los títulos, *abstracts* o *keywords*. Seleccionamos la base de datos Scopus porque ofrece metadatos completos y curados, a diferencias de otras bases –como Jstor o Ebsco– que también fueron consultadas pero que devuelven muy pocos resultados y con menos metadatos. La búsqueda se realizó en diciembre de 2023. Al trabajar con una sola fuente (Scopus) no nos vimos obligados a tratar con duplicados, ni eliminar documentos por irrelevantes o incompletos. El resultado es un corpus de 904 documentos.

Preprocesamiento y análisis

Para responder nuestras preguntas no se necesitó más que estadísticas descriptivas y visualizaciones, las cuales fueron realizadas enteramente con el software R (Team R Core, 2018). Para los análisis de relaciones entre temas y conceptos (**RQ1**) usamos dos caminos. En un primer lugar, generamos redes de *keywords* (proporcionadas por los autores) considerando su co-presencia en los artículos. Tal como hicieron Flensburg y Lomborg, para evitar una red con “datafication” en el centro se lo excluyó del análisis, al igual que los términos “big data”, “big”, y “data”, “online”, y “literature”, y se definió un mínimo de frecuencia para la inclusión de los *keywords* de 5. Así, pasamos de una base de 2.718 *keywords* únicos, a una de 129. Luego, se detectaron comunidades, utilizando el algoritmo *cluster_louvain*⁵, implementado en el paquete *igraph* (Newman, 2006).

En segundo lugar, realizamos un modelado de tópicos a partir de los *abstracts*. Para ello, usamos el paquete *stm* (Roberts et al., 2019) el cual supone una distribución probable del vocabulario del corpus a partir de analizar las correlaciones en los documentos. Este procedimiento supone que se introduzca el número de temas a detectar como un parámetro (K). Luego de realizar distintas pruebas estadísticas y de coherencia, optamos por un número de tópicos K=10. Los resultados fueron explorados a través del análisis cualitativo y la lectura pormenorizada de los *abstracts* y *keywords* más representativos de cada tópico.

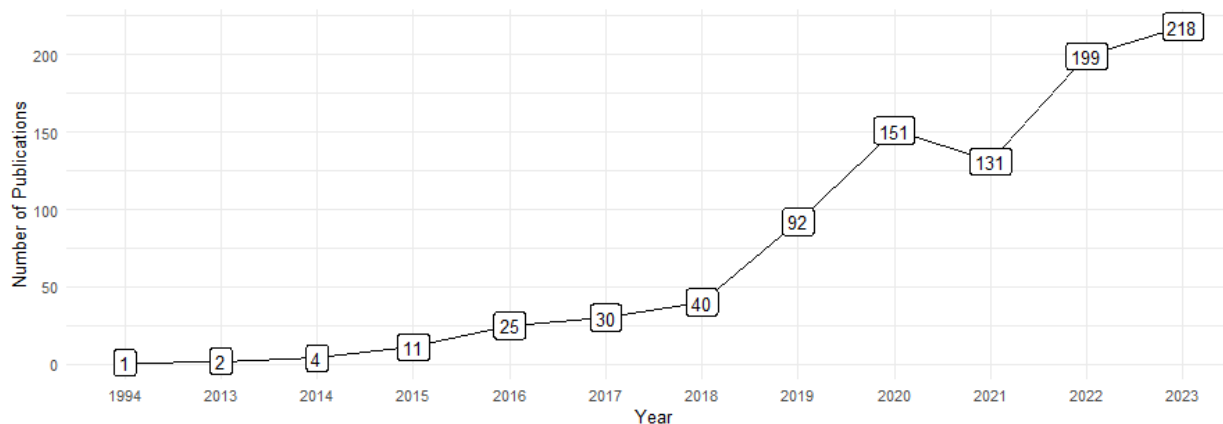
Para identificar referentes y referencias (**RQ2**) utilizamos el software *Anystyle* (<https://anystyle.io/>) en su implementación para *RubyGem*. Los resultados de este procesamiento dependen de la consistencia de las referencias que, lamentablemente, varían de formato y calidad en cada documento, de modo que deben ser tomados con recaudo. Los datos para responder las preguntas sobre revistas y autoría/países (**RQ3** y **RQ4**) no requirieron mayores procesamientos. Aquí hemos realizado tablas ordenadas por frecuencias y redes de co-ocurrencia para observar rankings y grupos.

⁵ Es un método heurístico utilizado para detectar comunidades dentro de redes complejas o grafos. Utiliza el agrupamiento jerárquico basado en la optimización de una métrica llamada *e*, que mide la densidad de enlaces dentro de comunidades en comparación con enlaces entre comunidades.

Resultados

En nuestro dataset se observa un crecimiento continuado de la producción sobre datificación desde los primeros artículos que lo incluyen (Figura 1). El primer artículo en utilizar el término es el de John Hankins (1994) que lo incluye en su título “The CICNet rural datafication project: Extending network access and services”. Luego, ya en 2013, Monica Brannon (2013) lo incluye en el *abstract* para referir a la producción de datos satelitales. En ninguno de estos casos el término se utiliza con una definición explícita, sino más bien metafóricamente y entre comillas. Finalmente, también de 2013 se encuentra el trabajo de van Dijck y Poell (2013) que ya lo utilizan referenciando a la obra de Mayer-Schonberger y Cukier.

Fig. 1: Evolución de la producción científica

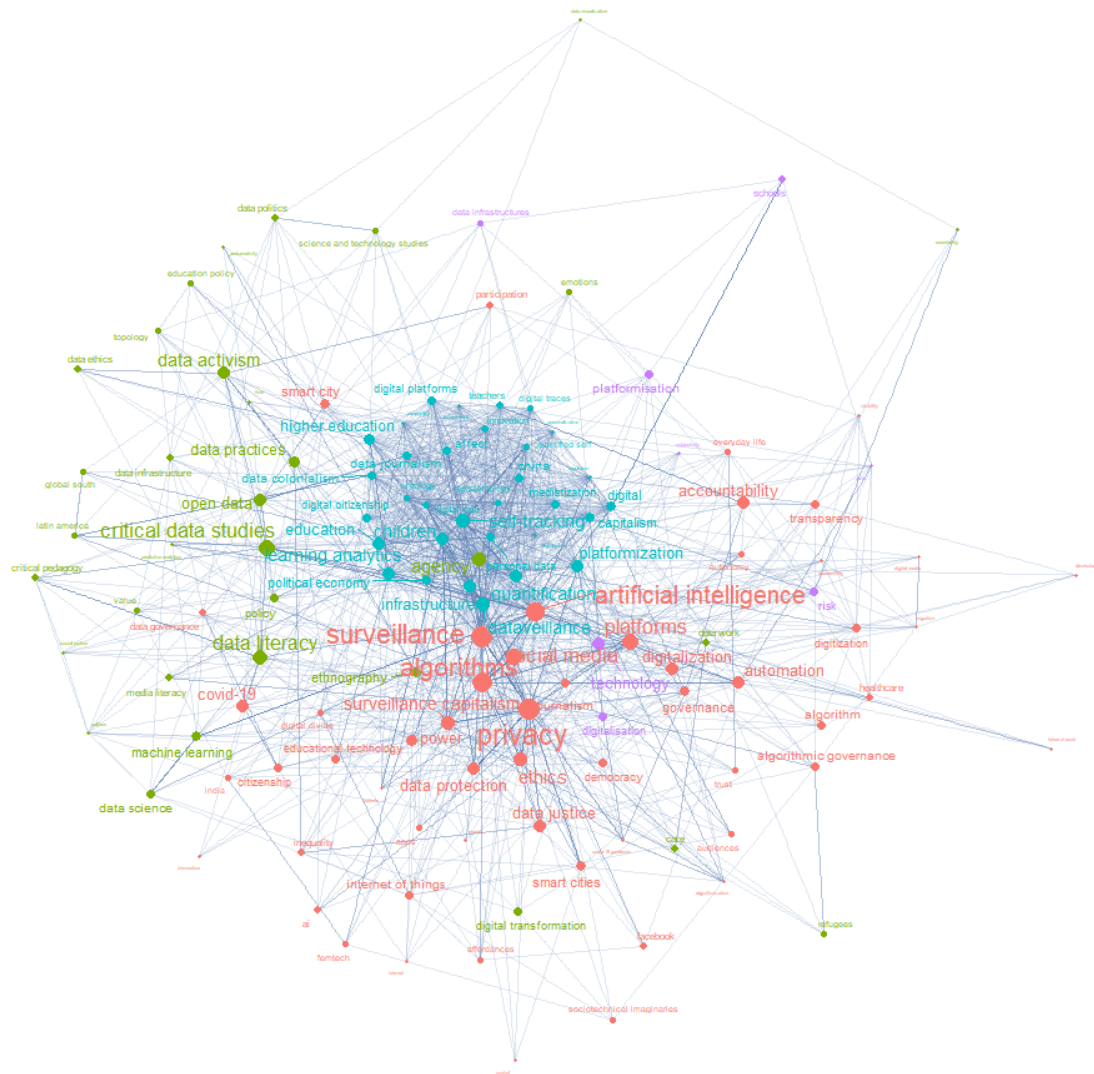


Fuente: producción propia

RQ1: temas y conceptos

Nuestra primera pregunta se interesa por los temas y conceptos que se trabajan en relación a la datificación. Para ello, generamos una red de *keywords* según su co-presencia en los artículos (Figura 2), detectando cuatro comunidades.

Fig. 2: Co-ocurrencia de keywords (n>5)



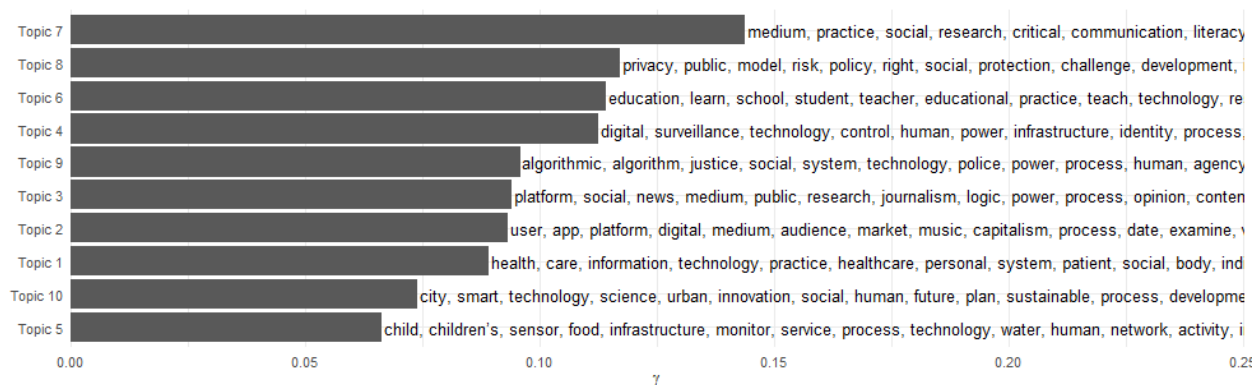
Fuente: Elaboración propia.

La comunidad más grande (56 elementos) incluye los *keywords* más frecuentes: privacidad, vigilancia, algoritmos, inteligencia artificial, redes sociales, plataformas, entre otros. La segunda comunidad (33 elementos) incluye *keywords* vinculados al enfoque de los estudios críticos de datos, que ponen el foco en la naturaleza política de estas prácticas y que buscan la discusión de sus supuestos para favorecer el entendimiento del público (Dalton & Thatcher, 2014; Kitchin &

Lauriault, 2014). El tercer grupo (32 elementos) se encuentra altamente integrado, y sus *keywords* remiten a la datificación y cuantificación de espacios de la vida social y subjetiva como el auto-trackeo –también conocido como *quantified self*– o la educación. Estos son espacios que tienen una larga tradición de producir datos para el monitoreo y la evaluación. En el último grupo tiene sólo ocho elementos y trata con conceptos generales –incluyendo tecnología, riesgos, digitalización, subjetividad o trabajo–, sin una coherencia semántica clara, y bien puede ser residual.

Una segunda manera de explorar y clasificar los temas del corpus es a través de modelar los tópicos de los abstracts, considerando la distribución probable de su vocabulario. Este tipo de análisis permite una mayor granularidad. La Figura 3 muestra los resultados para 10 tópicos (filtrando otra vez los *keywords* “datafication”, “big data” y “data”).

Fig. 3: modelado de tópicos en abstracts (K=10)



Fuente: Elaboración propia.

Los primeros dos tópicos que más prevalecen (#7 y #8) y el cuarto (tópico # 4) incluyen algunos términos vinculados con la perspectiva de los estudios críticos de datos, pero también términos genéricos del vocabulario de las ciencias sociales. Esta generalidad nos advierte a tomar estos tópicos con recaudo, ya que podrían ser un reflejo de una perspectiva común en el corpus. Si buscamos coincidencias entre términos del *abstract* y *keywords* vemos prevalencia de un interés

por el control, la vigilancia, la protección, la regulación, y la crítica.

Más claros son los tópicos siguientes: el #6 trata con educación, el #9 con el sistema jurídico y policial, el #3 con plataformas y periodismo, el #2 con plataformas y audiencias, el #1 con el sistema de salud, el #10 con la *smart city* y el urbanismo, y el #5 con los niños y padres. En todos ellos hay coherencia entre los resultados del análisis de *abstracts* y de *keywords*. Keywords como “privacidad”, “ética”, “gobierno” (governance), “plataformas” son una constante también en estos tópicos. Todos estos son buenos candidatos para ser tomados como sub-campos de estudios, aunque esto debería ser confirmado por otro tipo de estudios.

RQ2: referencias y referentes

Si tomamos las citas como un *proxy* para la visibilidad e influencia de una comunicación, podemos observar que nuestro corpus cuenta con claros referentes. Entre los 10 artículos más citados (Tabla 1) se repiten los nombres de José van Dijck (Universidad de Utrecht), Thomas Poell (Universidad de Amsterdam), y David Nieborg (Universidad de Toronto), investigadores del área de la cultura y los medios, y coautores entre sí de trabajos sobre el proceso de la “plataformización”. Luego, vemos trabajos en áreas específicas antes referidas: *smart city*, infancias, justicia, educación, entre otras. También observamos trabajos que introducen el concepto de datificación y que ofrecen una interpretación teórica y una discusión crítica desde la sociología, la política, y los estudios críticos de datos.

Tabla. 1: Artículos más citados

Citas	Autores	Título
999	van Dijck J.	Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology
731	van Dijck J.; Poell T.	Understanding social media logic
468	Nieborg D.B.; Poell T.	The platformization of cultural production: Theorizing the contingent cultural commodity
409	Sadowski J.	When data is capital: Datafication, accumulation, and

		extraction
269	Bossetta M.	The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election
254	Lupton D.; Williamson B.	The datafied child: The dataveillance of children and implications for their rights
226	Kitchin R.	The ethics of smart cities and urban science
190	Poell T.; Nieborg D.; van Dijck J.	Platformisation
177	Baack S.	Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism
174	Milan S.; Treré E.	Big Data from the South(s): Beyond Data Universalism
174	Dencik L.; Hintz A.; Cable J.	Towards data justice? The ambiguity of anti-surveillance resistance in political activism

Fuente: Elaboración propia.

El trabajo más citado es “Datafication, dataism and dataveillance” de van Dijck (2014), en el que la autora busca dilucidar críticamente los supuestos ideológicos y epistémicos de Mayer-Schoenberger y Cukier (2013), señalando que la datificación y el *big data* se sustentan en la creencia en la cuantificación objetiva y el potencial de las prácticas de registro sobre todos los aspectos de la vida social y subjetiva, y que esta confianza se transfiere a agentes institucionales y corporativos. Además, discute la idea de la datificación como un proceso natural o neutral de registro, caracterizándose más bien como una estrategia de extracción y comercialización, con reclamos epistemológicos discutibles, tales como la supuesta exhaustividad del *big data*, su naturalidad, o su confiabilidad. Un antecedente central de este último punto –referenciado por van Dijck, y resultando en uno de los trabajos más citados del campo– lo ofrecen boyd y Crawford (2012) con el paper titulado “Critical questions for big data”. Estas autoras estrictamente no hablan de “datificación” sino de *big data*, al cual definen como un fenómeno cultural, tecnológico y académico que se constituye en el interjuego entre tecnología, análisis y mitología.

Si observamos ahora las referencias incluidas en los artículos del corpus (Tabla 2), podemos tener una idea de otros trabajos influyentes, sin limitarnos a nuestro corpus y al formato de artículo. En primer lugar, volvemos a ver el trabajo de van Dijck (2014) y en tercero el de Mayer-Schoenberger y Cukier (2013), lo que sugiere la centralidad de esta discusión para enmarcar el concepto. Otros trabajos destacados y muy referenciados, tanto dentro como fuera de este corpus, son los de la investigadora Shoshana Zuboff en los que se conceptualiza sobre el capitalismo de vigilancia, al cual define –entre otras sentencias– como “Un nuevo orden económico que reivindica la experiencia humana como materia prima gratuita para prácticas comerciales ocultas de extracción, predicción y venta” (2015, 2019b) y “Una expropiación de derechos humanos críticos que se entiende mejor como un golpe desde arriba: un derrocamiento de la soberanía popular” (Zuboff, 2019b, p. 1). También observamos trabajos críticos del *big data* y el rol social de los algoritmos, como el de Rob Kitchin (2014a), Cathy O’Neil (2016), Virginia Eubanks (2018) o Frank Pasquale (2015). Todos ellos ofrecen una interesante casuística de ejemplos en los que los algoritmos tienen efectos sociales perversivos que refuerzan inequidades o generan situaciones de injusticia, al automatizar procesos de clasificación y decisión sobre grupos vulnerables, a partir de datos sesgados. Todos ellos también coinciden en avanzar la línea de discusión epistemológica del *big data* como respuesta a dicha problemática.

Tabla. 2: Referencias más citadas

Citas	Autor	Título
120	José van Dijk	Datafication, Dataism And Dataveillance: Big Data Between Scientific Paradigm And Ideology
98	Shoshana Zuboff	The Age Of Surveillance Capitalism: The Fight For A Human Future At The New Frontier Of Power
66	Kenneth Cukier, Viktor Mayer-Schonberger	Big Data: A Revolution That Will Transform How We Live, Work, And Think

66	Shoshana Zuboff	Big Other: Surveillance Capitalism And The Prospects Of An Information Civilization
57	Rob Kitchin	The Data Revolution: Big Data, Open Data, Data Infrastructures And Their Consequences
55	Virginia Eubanks	Automating Inequality: How High-Tech Tools Profile, Police, And Punish The Poor
51	Jathan Sadowski	When Data Is Capital: Datafication, Accumulation, And Extraction
50	Frank Pasquale	The Black Box Society: The Secret Algorithms That Control Money And Information
49	Deborah Lupton & Ben Williamson	The Datafied Child: The Dataveillance Of Children And Implications For Their Rights
45	Catherine O'Neil	Weapons Of Math Destruction: How Big Data Increases Inequality And Threatens Democracy

Fuente: Elaboración propia.

Un dato interesante es que todos estos referentes son autores que provienen de los Estados Unidos o Europa, de modo que los casos documentados suelen referir a dichas latitudes y a las prácticas de datos de las instituciones de estas sociedades, las cuales no siempre son generalizables a otros contextos.

RQ3: revistas

Una caracterización de la comunicación sobre datificación no puede dejar de considerar los principales canales de comunicación científica. Para ello, podemos consultar cuáles son las revistas más frecuentes de nuestro corpus y cuales las más frecuentes entre las referencias de los artículos del corpus. Hemos compilado y cruzado ambos datos en la Tabla 3, e incluido la información de la editorial, las áreas disciplinares y el cuartil en el que tienen el mayor ranking, utilizando la información disponible en *Scimago Journal & Country Rank*.

Tabla. 3: revistas más frecuentes en el corpus y en las referencias del corpus

Journal	Corpus (rank / n)	References (rank / n)	Editorial / País	Áreas*	Mejor cuartil
Big Data and Society	1 / 46	1 / 804	Sage / UK	CS, DS, SS	Q1 Communication
New Media and Society	2 / 27	2 / 753	Sage / UK	SS	Q1 Communication
Learning, Media and Technology	3 / 17	91 / 37	Routledge / UK	E, SS	Q1 Education
Television and New Media	4 / 16	9 / 206	Sage / US	AH, SS	Q1 Cultural studies
International Journal of Communication	5 / 14	4 / 353	USC / US	SS	Q1 Communication
Surveillance and Society	6 / 14	6 / 305	Surveillance Studies Network / UK	SS	Q1 Safety research
Information Communication and Society	7 / 13	109 / 32	Routledge / UK	SS	Q1 Communication
Internet Policy Review	8 / 12	13 / 145	A. v Humboldt Inst. for Internet and Society / Germany	CS, ES, SS	Q1 Communication
Media and Communication	9 / 10	34 / 74	Cogitatio Press / Portugal	SS	Q1 Communication
Media, Culture and Society	10 / 10	153 / 24	Sage / UK	SS	Q1 Communication

Fuente: Elaboración propia.

* Áreas = CS: computer science, DS: decision sciences, SS: social sciences, E: engineering, AH: arts and humanities, ES: environmental sciences

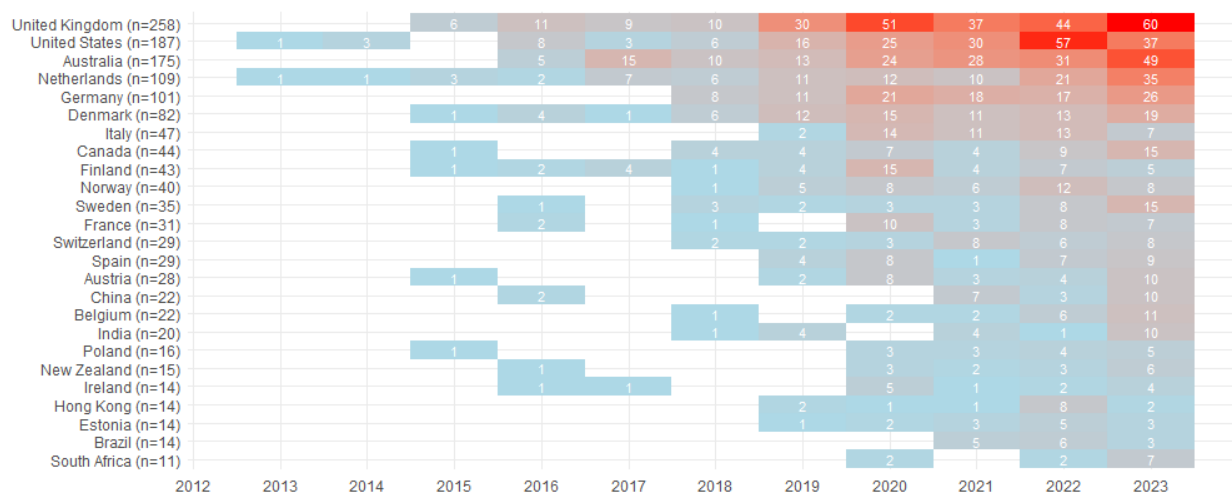
Tanto entre los artículos de nuestro corpus, como en sus referencias, las revistas de mayor presencia son “Big data & Society” y “New Media & Society”, ambas de Sage Publications y registradas en U.K. Otras revistas que comparten el ranking entre ambas listas son “International Journal of Communication” y “Surveillance and Society”, editadas por otras firmas y basadas en U.S. y U.K. Al menos en el corpus construido a través de Scopus, entre las principales no hay revistas editadas fuera de U.S. y Europa.

Si consultamos las categorías y áreas de temas en *Scimago Journal & Country Rank*, podemos notar que todas las revistas comparten la categoría de ciencias sociales. Además, la mayoría de las revistas tienen su mejor cuartil en el área de comunicaciones, incluso en el caso de las revistas que comparten otras áreas de interés del *big data*, como podrían ser las ciencias de la computación, o la ingeniería, o las ciencias vinculadas a la decisión y el *management*. Esto puede ser indicativo de que la mirada de la datificación como proceso (social) es propia de las ciencias sociales y la perspectiva crítica.

RQ4: producción por países y colaboración

La distribución de la producción por países y años (Figura 4), muestra la preeminencia de Reino Unido, con un conteo de *papers* que se ha ido incrementado en los últimos años, superando a Estados Unidos y Países Bajos, desde donde se publicaron los primeros trabajos. Otros países como Australia, Dinamarca, Alemania o Italia comenzaron a publicar en estos temas sólo en los últimos años pero con un ritmo sostenido. Se debe notar que China y Hong Kong se encuentran en una posición más baja entre los que más producen, a pesar de ser países líderes en la investigación en *big data*. Por otro lado, a pesar de que teórica y conceptualmente se incorpora una mirada decolonial o que pretende retomar la posición del “sur global” en la datificación –algo que se hace evidente incluso por tener a “Big Data from the South(s): Beyond Data Universalism” de Milan y Treré (2019) entre los más citados–, entre los principales países sólo se observa a Brasil y Sudáfrica.

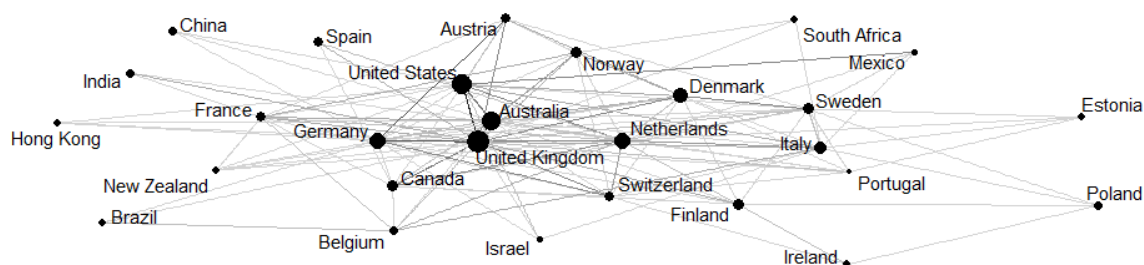
Fig. 4: producción por países y años (producción total > 10)



Fuente: Elaboración propia.

Las colaboraciones más fuertes se observan también entre los países que más producen, al menos en términos de coautoría (Figura 5).

Fig. 5: colaboración entre países, según coautoría (países > 5 artículos)

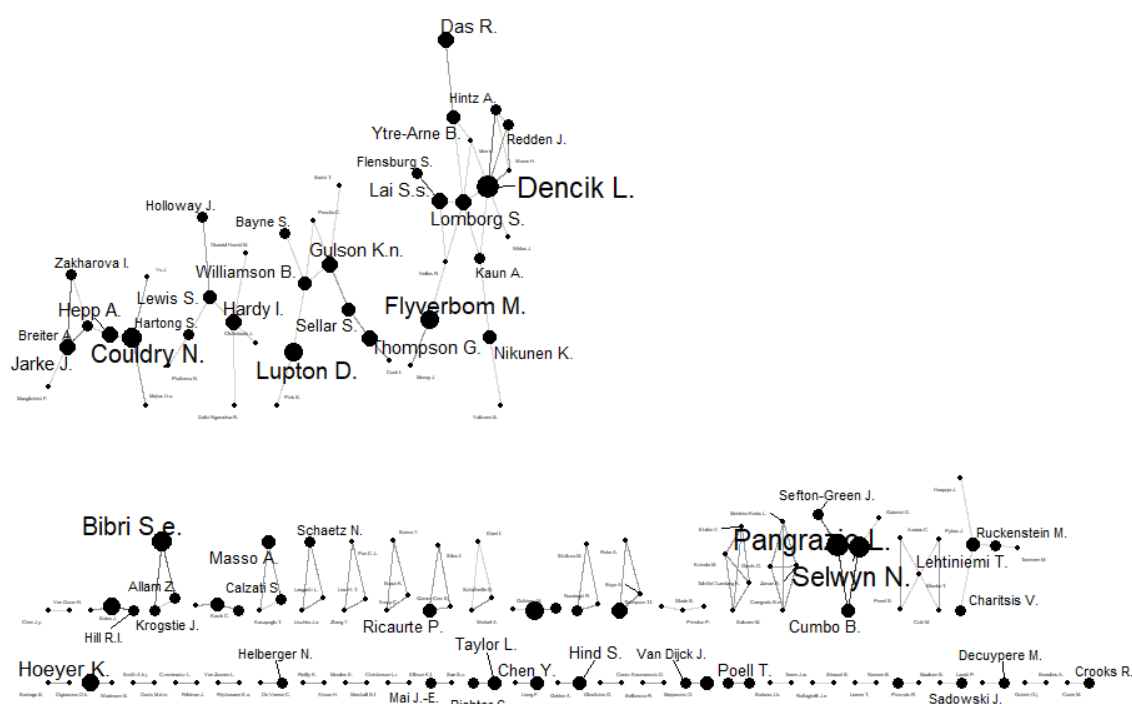


Fuente: Elaboración propia

Finalmente, si observamos la colaboración entre autores se puede observar una comunidad integrada con más de 20 autores colaborando, entre los que se encuentran autores de renombre y

gran presencia en nuestro corpus, como Dencik, Couldry o Lupton. Luego, hay varios autores publicando con otro autor, u otros 2 autores, aunque sin integración a otros grupos.

Fig. 6: colaboración entre autores, según coautoría (autores > 1 artículo)



Fuente: Elaboración propia.

Conclusiones

En esta sección sintetizamos nuestros resultados, los ponemos en diálogo con la literatura, e indicamos futuras líneas de indagación. En nuestro recuento de temas y conceptos (**RQ1**) identificamos tres grandes grupos de keywords concurrentes que se pueden diferenciar, y un cuarto menos claro: el primero centrado en los *media studies*, el segundo en la perspectiva de los *critical data studies*, y el tercero en el *quantified self* y la datificación del mundo educativo. A grandes

rasgos, este mapeo de temas coincide con el realizado por Flensburg y Lomborg (2023) que también identificaron estos tres componentes, aunque propusieron otros adicionales de menor peso. Si bien esta metodología debe ser considerada exploratoria, los resultados sugieren una red altamente conectada, con muchos *keywords* que sirven de puentes entre los distintos grupos de artículos. En esta caracterización coincidimos con la lectura realizada por Flensburg y Lomborg sobre una red similar, y también con la afirmación de que el “El campo de los estudios de medios y comunicación está fuertemente representado y podría decirse que es un impulsor clave en la investigación sobre datificación.” (2023, p. 1457). Luego, nosotros complementamos este análisis con otra técnica y otra fuente (*topic modeling* y *abstracts*) que, si bien arrojó resultados con mayor granularidad, estos son coherentes con lo encontrado vía análisis de *keywords*.

En la introducción conceptual referíamos a dos grandes ideas de la datificación: una centrada en el proceso de recolección de datos y las posibilidades o promesas de su analítica, que identificamos con el uso de Mayer-Schonberger y Cukier; y otra más amplia que privilegia cuestionar sus fundamentos sociales, políticos y epistémicos, en la línea de la crítica del dataísmo que propone van Dijck. El recuento conceptual de Flensburg y Lomborg (2023) también identifica estas dos ideas tensionadas. Las autoras sugieren que las dos ideas conviven en todos los temas, tal vez en proporción distinta. De hecho, a partir del análisis de la red de *keywords*, quisiéramos matizar la idea de las autoras acerca de que en el campo se encuentran balanceados el interés crítico/sociológico, y el interés por lo tecnológico/*business-oriented*. Más bien creemos que se puede hipotetizar una clara hegemonía de la perspectiva crítica en torno a este concepto. O en otros términos: creemos que la datificación en el sentido más trivial y menos crítico es, tal vez, marginal, y que cuando se habla de “datificación” la mirada predominante es la crítica. Cabe destacar que casi no hay presencia de *keywords* que supongan un interés netamente comercial, computacional o informático, como sí hemos observado en otras revisiones centradas en el concepto *big data*.

Entre ambos conceptos –*big data* y *datificación*– parece haber una relación de proceso-producto, en la que la mirada sobre el producto es menos crítica, y la mirada en el proceso se caracteriza por ser deconstructiva. Los principales referentes (RQ2) parecen apoyar esta idea: los autores y estudios más citados aportan miradas críticas de la datificación y el *big data* que

privilegian su trasfondo social y su relación con el capitalismo y sus actores.

Por otro lado, el análisis sobre países de autores (**RQ4**) y revistas (**RQ3**) da cuenta de una comunicación en donde predominan las voces de Estados Unidos y Europa, lo cual impacta en la casuística sobre la que se elabora la reflexión. Sin embargo, se debe considerar que nuestros criterios de búsqueda limitaban el corpus a artículos en inglés y de revistas indexadas en Scopus, lo cual supone cierta capacidad e infraestructura editorial, a veces difícil de lograr para revistas de Latinoamérica y otros países emergentes. No hay duda que los actores y la lógica de la datificación suponen un capitalismo transnacional, pero es importante señalar que esto puede resultar en impactos y recorridos diferenciados en distintas latitudes. Como bien señalan Milan y Treré (2019), en torno al *big data* y la datificación parece operar un cierto “universalismo digital” que tiende a absorber la heterogeneidad de diversos contextos sobre las diferencias y especificidades culturales. Para contrarrestarlo nos resulta importante contar con mayor pluralidad de voces que permitan estudiar las dinámicas relacionadas con la datificación desde otros contextos históricos, sociopolíticos, regulatorios, culturales y económicos, y considerando distintas historias de problemas como la desigualdad social, la desposesión y apropiación de recursos, la manera en que las ideologías se encarnan en las idiosincrasias culturales, entre otros (D’Ignazio & Klein, 2020). Así, la estrategia colaborativa que abogamos ofrece una tendencia de estudios críticos y líneas de investigación sobre los datos que cuestionan la centralidad de las interpretaciones “occidentalistas”. Al igual que Milán y Treré (2019), apostamos a que la colaboración entre comunidades de conocimiento pueda ser un factor importante para avanzar sobre este punto crítico.

Con este trabajo hemos buscado generar una descripción de la comunicación científica sobre la datificación, ofreciendo una mirada bibliométrica complementaria que observa el campo en relación a sus temáticas, ámbitos de producción y colaboración, espacios de publicación y principales referentes, avanzando el estado del conocimiento en relación a nuestro principal antecedente, que es el trabajo de Flensburg y Lomborg (2023). Entre nuestros principales aportes destacamos, por un lado, haber actualizado el corpus, ya que el de dichas autoras abarca hasta octubre del 2020, y el nuestro hasta diciembre de 2023. Además, hemos podido brindar una imagen en la que se pone de manifiesto la dimensión geopolítica de los procesos de construcción de conocimiento en el área. Así, si bien no se optó por una revisión sistemática, los objetivos

alcanzados brindan un cruce de variables e información que alumbran el proceso de producción científica en relación a quiénes producen, desde qué lugares del mundo y en qué tipo de revistas científicas.

En última instancia, creemos que el aporte de este trabajo no sólo es actualizar el estado de conocimiento en relación a la última revisión sistemática, sino además abrir la discusión sobre las asimetrías que condicionan la producción científica sobre el proceso de la datificación. Sin embargo, este recorrido tuvo limitaciones que sería interesante que futuras investigaciones consideren. Por caso, sería necesario incorporar otras fuentes de datos aparte de Scopus y/o Web Of Science para ampliar el corpus considerado, así como explorar otros criterios de búsqueda que reflejen la tematización de otras miradas disciplinares. También, sería importante incluir otros motores de búsqueda regionales, y/o ampliar a criterios en otros idiomas. Asimismo, que se integren ciertas preguntas de investigación uniendo la mirada de la revisión sistemática con la bibliométrica, a fin de profundizar los datos y complejizar el debate bajo esa complementación metodológica. Dado que se trabajó únicamente con artículos científicos para la construcción del corpus, sería importante incorporar otros *venues* o canales, como las conferencias u otros espacios de la comunidad científica, a fin de enriquecer las fuentes y llegar a otros discursos reconocidos en el ámbito de la investigación. También, siguiendo el ejemplo de Jobin et al. (2019) para la inteligencia artificial, sería un aporte sustancial conocer si existe, para el caso de la datificación, una distribución diferencial de intereses/problemáticas planteadas por investigaciones de países desarrollados respecto de las “periferias” o del “tercer mundo”. En cualquier caso, se destaca el riesgo que supone abandonar el conocimiento local, el pluralismo cultural y ciertas exigencias éticas en las discusiones actuales sobre la datificación, como una forma de cuestionar el universalismo digital, de cara a futuras investigaciones. Por último, en relación a los resultados del presente trabajo (**RQ2**), sería necesario seguir profundizando los resultados de estudios bibliométricos que trabajen la datificación y el *big data*, para conocer cuál es el marco conceptual que predomina en el abordaje de estos fenómenos.

Referencias

Ahmad, I., Ahmed, G., Shah, S. A. A., & Ahmed, E. (2020). A decade of big data literature:

- analysis of trends in light of bibliometrics. *The Journal of Supercomputing*, 76(5), 3555–3571.
<https://doi.org/10.1007/s11227-018-2714-x>
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data and Society*, 2(2), 1–11.
<https://doi.org/10.1177/2053951715594634>
- Becerra, G., & Ratovicus, C. (2022). Social sciences and humanities on big data: a bibliometric analysis. *Journal of Information Systems and Technology Management*, 19(e202219011).
<https://doi.org/10.4301/S1807-1775202219011>
- Bibri, S. E., & Krogstie, J. (2020). The emerging data-driven Smart City and its innovative applied solutions for sustainability: the cases of London and Barcelona. *Energy Informatics*, 3(1).
<https://doi.org/10.1186/s42162-020-00108-6>
- boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. [https://doi.org/https://doi.org/10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878)
- Brannon, M. M. (2013). Standardized spaces: Satellite imagery in the age of big data. *Configurations*, 21(3), 271–299. <https://doi.org/10.1353/con.2013.0021>
- Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television and New Media*, 20(4), 336–349.
<https://doi.org/10.1177/1527476418796632>
- D'Ignazio, C., & Klein, L. (2020). *Data feminism*. The MIT Press.
- Dalton, J., & Thatcher, J. (2014). What does a critical data studies look like , and why do we care? Seven points for a critical approach to 'big data.' *Society and Space*, 1–12.
<http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/>
- Eubanks, V. (2018). *Automating inequality*. Macmillan Publishers.
- Flensburg, S., & Lomborg, S. (2023). Datafication research: Mapping the field for a future agenda. *New Media and Society*, 25(6), 1451–1469. <https://doi.org/10.1177/14614448211046616>
- Hankins, J. L. (1994). The CICNet rural datafication project: Extending network access and services. *Internet Research*, 4(1), 71–74.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature*

- Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kalantari, A., Kamsin, A., Kamaruddin, H. S., Ale Ebrahim, N., Gani, A., Ebrahimi, A., & Shamshirband, S. (2017). A bibliometric approach to tracking big data research trends. *Journal of Big Data*, 4(1), 1–18. <https://doi.org/10.1186/s40537-017-0088-1>
- Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <http://journals.sagepub.com/doi/10.1177/2053951714528481>
- Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage. <https://doi.org/10.4135/9781473909472>
- Kitchin, R., & Lauriault, T. P. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. *Geoweb and Big Data*, 1–19. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Liu, X., Sun, R., Wang, S., & Wu, Y. J. (2019). The research landscape of big data: a bibliometric analysis. *Library Hi Tech*, 38(2), 367–384. <https://doi.org/10.1108/LHT-01-2019-0024>
- Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media and Society*, 19(5), 780–794. <https://doi.org/10.1177/1461444816686328>
- Lycett, M. (2013). “Datafication”: Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381–386. <https://doi.org/10.1057/ejis.2013.10>
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work, and think*. Eamon Dolan/Houghton Mifflin Harcourt.
- Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, 8(4), 1–10. <https://doi.org/10.14763/2019.4.1428>
- Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television and New Media*, 20(4), 319–335. <https://doi.org/10.1177/1527476419837739>
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(3), 1–19. <https://doi.org/10.1103/PhysRevE.74.036104>
- O’Neil, C. (2016). *Weapons of math destruction*. Crown/Archetype.
- Pasquale, F. (2015). *The Black Box Society*. Harvard University Press.

<https://doi.org/10.4159/harvard.9780674736061>

- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, 91. <https://doi.org/10.18637/jss.v091.i02>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data and Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718820549>
- Team R Core. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*, 12(2), 197–208.
- van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication*, 1(1), 2–14. <https://doi.org/10.12924/mac2013.01010002>
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89. <https://doi.org/10.1057/jit.2015.5>
- Zuboff, S. (2019a). Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum*, 28(1), 10–29. <https://doi.org/10.1177/1095796018819461>
- Zuboff, S. (2019b). *The Age of Surveillance Capitalism*. PublicAffairs. <https://doi.org/https://dl.acm.org/doi/10.5555/3350443>