

DATIFICACIÓN CRÍTICA DE LA VIOLENCIA. APORTACIONES
A LOS ESTUDIOS CRÍTICOS DE DATOS DESDE CONTEXTOS
LATINOAMERICANOS

***Critical Datification of Violence. Contributions to Critical Data
Studies from Latin American Contexts***

*Datificação Crítica da Violência. Contribuições para os Estudos Críticos
de Dados a partir de Contextos Latino-Americanos*

VÍCTOR HUGO ABREGO MOLINA¹

Recibido: 13 de noviembre de 2023.

Corregido: 14 de agosto de 2024.

Aceptado: 2 de septiembre de 2024.

Resumen

En diálogo con los estudios críticos de datos, este artículo propone la noción de datificación dominante, constituida por una ontología numérica, una racionalidad predictiva, una gubernamentalidad de vigilancia y control, una ética comercial y una epistemología universalista. Revisa las áreas del conocimiento y los énfasis que están construyendo este campo y agrega aspectos del escenario latinoamericano, donde los procesos de datificación de la realidad están insertos en contextos de crisis de violaciones a los derechos humanos y de fragilidad institucional. Bajo este marco, se proponen insumos conceptuales para una datificación crítica de la violencia, que colocan a lo numérico fuera del centro del análisis, buscan evidenciar versiones invisibilizadas de registros históricos de violaciones graves a los derechos humanos, toman al dato como una infraestructura simbólica intervenible y abordan los archivos masivos como recortes políticamente situados; se utilizan herramientas para el diseño metodológico desde áreas como los métodos digitales y la lingüística de corpus.

Palabras clave: Estudios críticos de datos, datificación, violencia, Latinoamérica, derechos humanos.

¹ Licenciado en Comunicación por la Benemérita Universidad Autónoma de Puebla. Maestro en Comunicación por la Universidad de Guadalajara. Doctorado en Comunicación, lenguajes e información en la Pontificia Universidad Javeriana. Profesor del Departamento de Estudios Socioculturales. ITESO. Coordinador ejecutivo de Signa_Lab. Líneas de investigación: Estudios críticos de datos Comunicación y cultura digital. Correo electrónico: abregation@iteso.mx

Abstract

In dialogue with critical data studies, this article proposes the notion of dominant datafication, constituted by a numerical ontology, a predictive rationality, a governmentality of surveillance and control, a commercial ethic, and a universalist epistemology. It reviews the areas of knowledge and the emphases that are building this field and adds aspects of the Latin American scenario, where the processes of datafication of reality are inserted in contexts of crisis of human rights violations and institutional fragility. Within this framework, conceptual inputs are proposed for a critical datafication of violence that place the numerical outside the center of the analysis, seek to reveal invisible versions of historical records of serious human rights violations, take data as an intervenable symbolic infrastructure and address massive archives as politically situated clippings; tools are used for methodological design from areas such as digital methods and corpus linguistics.

Keywords: Critical data studies, datafication, violence, Latin America, human rights.

Resumo

Em diálogo com os estudos críticos de dados, este artigo propõe a noção de dataficação dominante, constituída por uma ontologia numérica, uma racionalidade preditiva, uma governamentalidade de vigilância e controle, uma ética comercial e uma epistemologia universalista. Revisa as áreas do conhecimento e enfatiza que estão construindo esse campo e acrescenta aspectos do cenário latino-americano, onde os processos de dataficação da realidade estão inseridos em contextos de crise de violações de direitos humanos e de fragilidade institucional. Neste quadro, são propostos insumos conceituais para uma dataficação crítica da violência que coloque o numérico fora do centro da análise, procure destacar versões invisíveis de registros históricos de graves violações dos direitos humanos, tome os dados como uma infraestrutura simbólica intervencionável e eles abordar os arquivos de massa como recortes politicamente situados; Ferramentas são utilizadas para o desenho metodológico de áreas como métodos digitais e linguística de corpus.

Palavras-chave: Estudos críticos de dados, dataficação, violência, América Latina, direitos humanos.

Introducción. Datificación de la realidad

La datificación de la realidad es una condición sociodigital que incide de distintos modos y con mayor o menor intensidad en los procesos de subjetivación, en las relaciones sociales y en las formas de producción del conocimiento en la actualidad. Atraviesa y condiciona las relaciones de poder y contribuye a modular y a modelar los límites de lo posible. Desde distintos ámbitos de la investigación social se intenta desarrollar andamiajes teórico-metodológicos que permitan comprender la incidencia de estas dinámicas en prácticas, sujetos y discursos concretos. Este artículo se inserta en esa amplia conversación e intenta aportar algunos elementos teóricos y principios

para el diseño metodológico alrededor de la datificación de la realidad, con un énfasis en datos acerca de violaciones a los derechos humanos.

Durante la última década, los estudios críticos de datos (*critical data studies*) se perfilan como campo de investigación en consolidación desde donde se llama a la producción y experimentación constantes de andamiajes teórico-metodológicos que alimenten tanto la conceptualización y teorización de las implicaciones y consecuencias de procesos globales de datificación de la realidad, como la construcción de imaginarios críticos y alternativos alrededor de estos procesos.

Estudios críticos de datos

El campo de los estudios críticos de datos emerge como un conjunto de preguntas alrededor de las implicaciones políticas, culturales, epistemológicas e históricas que tiene el avance de procesos cotidianos de datificación de la realidad, con particular atención en la toma de decisiones desde gobiernos y empresas a partir de procesos de recolección y análisis de grandes volúmenes de datos, en formas de apropiación y trabajo con datos desde la sociedad civil y diversas organizaciones y en el desarrollo de áreas del conocimiento especializadas en el manejo de datos. En conjunto, este enfoque desarrolla una postura crítica frente al pragmatismo y confianza en los análisis de datos que llegan a suprimir la necesidad de teorías por considerar que tienen acceso a “ideas nacidas de los datos” (Kitchin, 2014) y a confundir lectura numérica de la información con objetividad en la lectura de ésta.

Kitchin y Lauriault (2014) señalan que alrededor de estos avances de una realidad filtrada y validada mediante el manejo de cada vez más volúmenes de información, se configura un imaginario sociotécnico dominante que no sólo confía demasiado en los datos, sino que lo hace alrededor de un tipo de datos en específico, por una suerte de arrastre histórico-epistemológico que tiende a legitimar de por sí, o que suele plantar menos objeciones, al conocimiento producido a partir de análisis numéricos. Así, el manejo de grandes flujos de información supone dar a los datos masivos y a su lectura numérica una supuesta carga de “neutralidad”, como si los datos por sí mismos fueran “significativos independientemente del formato, el medio, el lenguaje, el productor y el contexto” –en el que fueron generados– y como si “reflejaran el mundo tal cual es” (Kitchin y Lauriault, 2014).

Frente a esta visión, los estudios críticos de datos buscan construir una agenda de investigación que pone al centro el análisis y la problematización de distintas implicaciones y consecuencias de un manejo y una concepción de los datos como la señalada. La siguiente es una breve clasificación abierta y no normativa de los énfasis y conceptos principales de este campo en construcción, busca ofrecer un panorama general mas no encasillar a ninguno de los conceptos o autores, pues en muchas investigaciones se mezclan. Se trata de un pequeño mapa de navegación del campo.

- *Panópticos*: Con énfasis en señalar el aumento en las posibilidades de vigilancia y control ejercidos por gobiernos y empresas privadas a través de la toma de decisiones mediante procesos de datificación de la realidad y generación de inteligencia digital. Esta corriente de los estudios críticos de datos propone conceptos como capitalismo de la vigilancia o *dataveillance* (van Dijck 2014; Yeung 2018), o el de gubernamentalidad algorítmica (Rouvroy y Berns, 2018) para señalar la racionalidad predictiva y el espíritu normativo de la datificación desde estos frentes.
- *Decoloniales y neomarxistas*: Alimentados por la tradición marxista y por lecturas históricas desde la perspectiva decolonial, esta corriente propone leer a la datificación de la realidad como una práctica en continuidad histórica con los procesos de extracción de recursos naturales llevados a cabo durante los periodos de conquista y colonización de América por parte de Europa (Mejias & Couldry, 2019; Ricaurte, 2019). Identifican cómo imaginarios sociales dominantes acerca de la información y de los datos llegan a atribuirles a ambos agencia propia, trabajan en descifrar nuevas maneras de fetichización de la información y en intuir fuentes de potenciales lógicas actualizadas de enajenación de los sujetos desde estas dinámicas (Ramos-Martín, 2022). Al mismo tiempo, desde una visión que busca colocar al centro la agencia de los sujetos y no sólo el avance de las estructuras de control a través de los datos, esta vertiente invita a una “decolonización de los datos” (Thatcher, *et al.* 2016; Milan & Treré, 2019; Ricaurte 2019). Utiliza conceptos como fetichismo de la información, *big data* desde el sur, colonialismo y colonialidad de datos.
- *Apropiaciones críticas*: Desde esta perspectiva, el énfasis está en las posibilidades no sólo de desarrollo conceptual crítico frente a la mirada gubernamental y comercial sobre los datos, sino en experiencias y metodologías de apropiación crítica de herramientas de manejo de datos masivos y “small data”. Ya sea desde un cuestionamiento al

pragmatismo en las formas actuales de enseñanza de la analítica de datos (Selwyn, 2020), o desde el llamado a emprender tanto formas de alfabetización con datos como cuestionamientos políticos a la utilización dominante de estas herramientas para crear literacidades en infraestructuras de datos (Gray, Gerlitz, y Bounegru, 2018); o desde el listado de distintas formas de “resistencia productiva” con los datos (Ettliger, 2018), que incluyen formas artísticas y hacktivismos, así como estrategias individuales de uso y resistencia frente a los algoritmos.

- *Tensiones agencia-estructura*: Esta corriente parte de la necesidad de señalar la ampliación de asimetrías entre quienes producen datos, quienes tienen acceso a los datos y quienes pueden tomar decisiones con los datos; pone al centro una serie de “ambivalencias” con los datos, entendidas como tensiones entre estructuras o instituciones y agencias o formas de resistencia colectiva de sujetos o grupos concretos (Hepp, Jarke, y Kramp, 2022), por ejemplo, entre acciones colectivas y acciones individuales o entre el estado neoliberal y la justicia con datos. En esta categoría también entran discusiones acerca de gobernanza y rendición de cuentas con datos.
- *Infraestructurales*: El énfasis desde esta línea está en identificar las formas en que las infraestructuras materiales y simbólicas alrededor de los procesos de datificación de la realidad inciden y condicionan modos de ejercer control y resistencia mediante el uso de datos. Identifican los andamiajes materiales (servidores, mecanismos de enfriamiento, almacenes, etcétera) detrás de la inmaterialidad (la “nube”) colocada en el último tramo de acceso a internet (Pickren, 2018), ubican los contextos y las dinámicas históricas en las que se insertan nuevas formas de ejercer control desde los gobiernos mediante el uso de datos y las susceptibilidades de que éstas funcionen o sean intervenidas por apropiaciones críticas de las infraestructuras desde la ciudadanía (Barreneche, 2019; Barreneche, *et al.*, 2021). Desde esta perspectiva, también se hace mención al concepto de ensamblajes de datos para referirse a aparatos (actores sociales) y elementos (discursos, imaginarios y materialidades alrededor de los aparatos) que interactúan entre sí, cuyo rol central es producir un circuito material, simbólico, político y epistemológico alrededor de los datos (Kitchin y Lauriault, 2014).

Este panorama general de los principales énfasis de los estudios críticos de datos da cuenta de la floreciente conceptualización alrededor de los

procesos de datificación de la realidad y busca dialogar con otras clasificaciones, como la de Kitchin, centrada en aspectos tales como ensamblajes, ecosistemas, prácticas, trayectorias y fricciones, imaginarios, galerías, política, poder, capitalismo, ética, justicia y activismo con datos (Kitchin, 2022). El desarrollo conceptual del campo, sin embargo, no va a la par de un desarrollo metodológico igualmente prolífico, sino más bien de “invitaciones” a desarrollar más trabajo en este ámbito, específicamente en trabajos que además de problematizar el uso de los datos masivos, ellos mismos usen datos masivos, ya que los análisis de datos propios de la mayoría de estas investigaciones tampoco son una constante en la literatura revisada hasta ahora, sino la revisión de trabajos con datos llevados a cabo por terceros, sobre todo artistas, hacktivistas, periodistas y activistas.

Datificación dominante

En conjunto, los énfasis hallados en los estudios críticos de datos delimitan y cuestionan lo que podría nombrarse como un tipo de “datificación dominante”, que consta de al menos cinco características alrededor de los procesos de datificación de la realidad. *Una ontología numérica*, que remite a un abarcamiento de “lo real” del dato centrado o más bien agotado en lo numérico, es decir una cualidad mediante la cual el dato es capaz de aportar al conocimiento únicamente en tanto “mide” algo; un tipo de *racionalidad predictiva*, que remite a una cualidad particular del análisis numérico de la realidad con *big data*, que más allá de controlar la ingente cantidad de información a la que tiene acceso y produce, busca controlar “lo por venir” en lo social a partir de esos datos; un tipo de *gubernamentalidad de vigilancia y control*, que remite a un vicio institucional del uso de inteligencia digital desde gobiernos y empresas privadas que amplía las posibilidades de perfilamiento de usuarios de tecnologías digitales con intenciones de producir a los sujetos como sospechosos o como consumidores, y de agotarlos en función de los registros digitales que se tienen de múltiples prácticas cotidianas llevadas a cabo por los sujetos; una *ética comercial*, que remite a una utilidad de los datos cifrada, sobre todo, en las posibilidades de generar rendimiento económico no sólo desde las grandes empresas tecnológicas, sino desde consultoras, universidades, medios de comunicación, *think tanks* y entidades públicas que trabajan con análisis de datos masivos, colocando en segundo término las posibilidades de

aproximación crítica, de investigación o simplemente sin fines de lucro del trabajo con datos; y finalmente una *epistemología universalizante*, que remite a un reforzamiento autorreferencial (de herencia colonial) de la jerarquía ejercida por la lectura estrictamente numérica de la realidad a través de datos, por encima de otras posibilidades de producir o, en el extremo, de imaginar formas de producción de conocimiento mediante el énfasis en otros aspectos de los datos, por ejemplo, en valores categóricos.

Así, la datificación dominante se ocupa del avance en las capacidades de ejercer control sobre los sujetos, generalmente desde gobiernos y empresas privadas, en detrimento de las capacidades de agenciamiento de la realidad a través de la gestión en la producción y el otorgamiento consentido de datos masivos de personas, grupos sociales específicos o poblaciones enteras. Frente a este conjunto de señalamientos generales y atendiendo al llamado a explorar “encuadres epistemológicos diferentes que permitan (...) diseñar insumos valiosos desde grandes volúmenes de datos situados y reflexivos” (Kitchin, 2014), hay que recordar que lo que se tiene en una base de datos no es al sujeto capturado, ni siquiera la experiencia del sujeto como tal, sino al registro digital de la experiencia del sujeto.

Sobre esto último hay que poner en tensión, por un lado, los errores y sesgos de la datificación dominante y, en esa medida, las posibilidades de hacer visibles acciones que escapan al registro y a la inteligencia habilitada por la datificación dominante; y por otro, las implicaciones y consecuencias individuales y colectivas, culturales y políticas de las ramificaciones constantes de la inteligencia digital utilizada para ampliar y sofisticar distintas formas de control y anticipación de la acción mediante una lectura capilar de lo real sin precedentes.

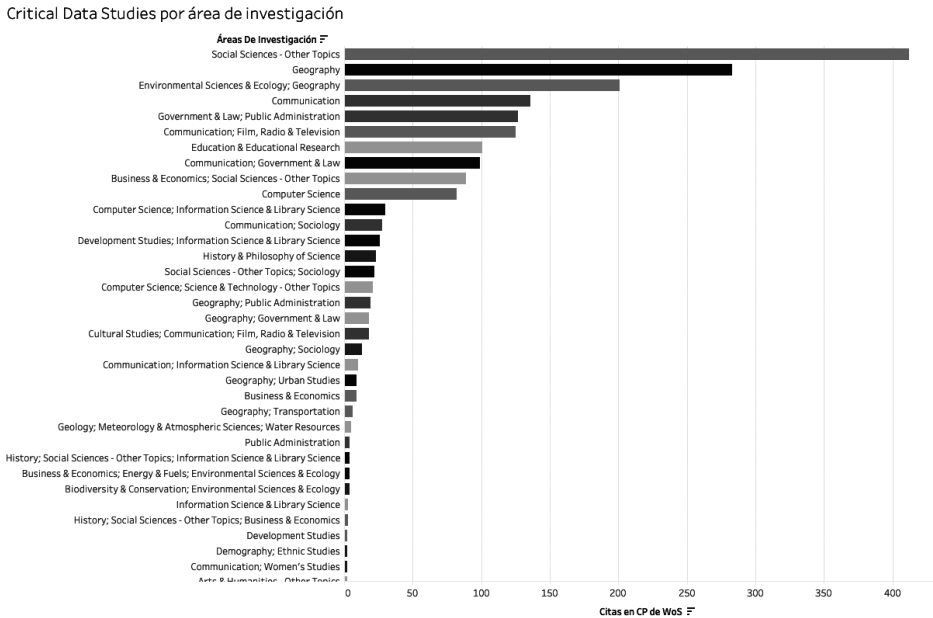
En resumen, frente a la datificación dominante hay que colocar una objeción general y una aceptación general: el sujeto no se agota en el registro numérico y digital de sus acciones; la datificación de lo real produce y retroalimenta de manera constante posibilidades de ejercer y sostener relaciones de poder dominantes.

¿Quién está construyendo el campo?

Si bien es posible afirmar que los estudios críticos de datos son un campo interdisciplinario que está siendo construido desde la teoría social en general,

la revisión de repositorios acerca de este campo en catálogos internacionales de ciencia nos puede ayudar a saber quiénes están haciendo las preguntas y llevando a cabo las investigaciones que mayor incidencia están teniendo en el mismo. Una búsqueda en el catálogo de la *Web of Science* llevada a cabo en agosto de 2023 arrojó 147 resultados de publicaciones de 2016 a 2023, alrededor de los “critical data studies”. Para jerarquizar, al menos de forma preliminar, estas publicaciones, se utilizó como filtro el número de citas que cada publicación tiene dentro del mismo catálogo. Este filtro muestra que entre las áreas que mayor número de citas acumulan destacan las ciencias sociales y “otros temas”, las ciencias de la computación, la comunicación, la geografía, y la educación e investigación en educación.

FIGURA 1
Áreas de investigación con mayor número de citas acumuladas en el catálogo de la Web of Science.



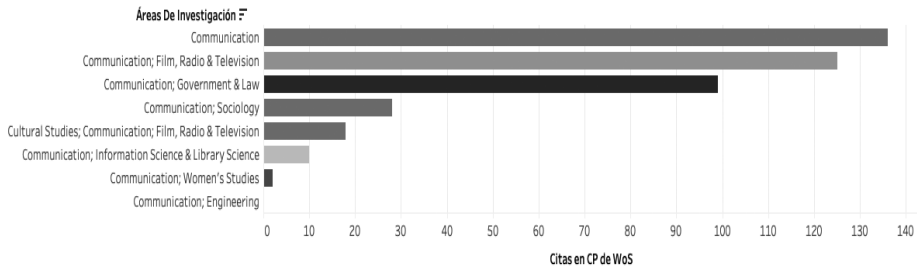
Fuente: Elaboración propia. Versión interactiva: https://public.tableau.com/app/profile/vicabr/viz/CriticalDataStudies-WebofScience_results_2016-2023/CriticalDataStudiesporreadeinvestigacin?publish=yes

Resulta de particular relevancia el peso que tiene la Comunicación en el campo de los estudios críticos de datos. La muestra del catálogo² registra siete combinaciones –el número más alto en la muestra– de este campo junto con otros en la producción de conocimiento en esta área. Las categorías Comunicación, Comunicación gobierno y leyes, Comunicación y sociología y Comunicación, cine, radio y televisión, destacan como áreas de investigación ubicadas en el top de artículos más citados en este campo desde 2016. Los estudios en comunicación han desarrollado históricamente una línea doble, tanto de comprensión como de apropiación de mediaciones técnicas de interacción social (radio, televisión, cine, prensa, etcétera), por lo que no es de extrañar su interés en aportar al estudio de datos, algoritmos y plataformas como las mediaciones técnicas de interacción social propias de la era digital; y al mismo tiempo, desde hace décadas, los estudios en comunicación han problematizado los sesgos constitutivos del registro de la realidad llevado a cabo desde mediaciones técnicas, donde quedó demostrado que los medios, por ejemplo la televisión, no “reflejan” la realidad, sino que la producen mediante procesos de “encodificación” que resultan del montaje culturalmente situado que se hace de dichos registros de la realidad, los cuales “atribuyen significado al acontecer social” (Orozco, 1987). Esto último bien puede ser tomado como un antecedente teórico de las posturas críticas frente a la supuesta neutralidad de los datos a gran escala que atañen a los estudios críticos de datos.

² Es importante ampliar con muestras de otros catálogos científicos la exploración de quién está construyendo los estudios críticos de datos.

FIGURA 2
Áreas relacionadas con comunicación

Critical Data Studies por áreas relacionadas con Comunicación



Fuente: elaboración propia.

De esta fuerte presencia en los estudios críticos de datos de disciplinas y campos como las ciencias sociales, la comunicación y la geografía, se desprende una observación clave en el desarrollo de conocimiento sobre procesos de datificación de la realidad, a saber, el impulso por no separar de anclajes teóricos críticos los desarrollos y propuestas metodológicas alrededor del estudio de datos —es decir, el desarrollo científico del campo— asume como necesario ir más allá de lo exploratorio y de lo descriptivo, e integrar como parte de sus supuestos experimentales fundamentales la identificación de tensiones, cambios y/o actualizaciones en las relaciones de poder entre actores cuyos registros digitales se estén investigando.

Hay que mencionar, además, que la centralidad de los datos, los algoritmos y las plataformas en los desarrollos teórico-metodológicos desde los estudios críticos de datos invita a la construcción de una agenda de investigación que lejos de excluir a las disciplinas y al conocimiento de lo social generado fuera de internet, abre el diálogo cercano con éstas y fortalece su pertinencia contemporánea, tanto por su conocimiento acumulado de lo social en diversos escenarios como por las posibilidades que brindan para evidenciar riesgos y oportunidades en apuestas emergentes de investigación social con datos.

Datificación de la realidad más allá del norte global

A continuación, proponemos ampliar y situar los procesos de datificación de la realidad para alimentar los alcances del estudio acerca de formas de datificación dominante y ampliar el mapa de preocupaciones pertinentes. Vistas en perspectiva, las preocupaciones generales de los estudios críticos de datos parecieran estar concentradas en la pérdida de la libertad y agencia individuales y grupales en escenarios institucionales, es decir, en la pérdida de garantías democráticas, vía la vigilancia gubernamental o de empresas privadas, a través de la captura masiva de datos personales. Estas son preocupaciones legítimas, que requieren atención desde los estudios críticos de datos y desde otros campos de estudio alrededor de la cultura digital. Sin embargo, es importante señalar que las preocupaciones por la pérdida de agencia ciudadana a nivel institucional y de garantías democráticas vienen sobre todo de sociedades en donde tales posibilidades han sido ejercidas de manera sostenida históricamente, o al menos lo suficiente como para sentir su detrimento en la época actual como “pérdida”. No obstante, si hablamos de sociedades como las latinoamericanas, las libertades democráticas y el agenciamiento ciudadano institucional no son precisamente legados históricos. Esto no significa que la defensa y el reclamo por sostener o acceder a estas libertades y garantías frente a procesos de datificación dominante no sea una parte relevante del necesario desarrollo teórico-metodológico de la investigación con y acerca de datos en todos los contextos sociales, pero sí plantea que es necesario ampliar las preguntas y los contextos para categorizar más preocupaciones pertinentes alrededor de las implicaciones de la datificación de la realidad.

Dicho lo anterior, las preocupaciones por vigilancia masiva y pérdida de libertades democráticas —a través de procesos de datificación dominante— en países ubicados en Latinoamérica, invitan a una problematización en un marco más amplio de violencias ejercidas contra grupos específicos de la población, lo cual no excluye que éstas también ocurran en contextos del norte global y que por ende también puedan integrarse a esta ampliación. Basta decir que la vigilancia y el espionaje digital en países de esta región se dan en un marco donde la persecución a periodistas, personas defensoras de derechos humanos y defensoras del medioambiente se han incrementado en los últimos años, a la par de ataques físicos y asesinatos a estos y otros actores críticos a grupos de poder locales y/o regionales. Así, la utilización de datos para la generación de inteligencia digital en

países como México, Colombia, El Salvador o Nicaragua,³ ha de ser pensada desde implicaciones más allá del acceso al derecho a la información y a cuestionar el control mercantil de datos de usuarios de plataformas digitales. Por ejemplo, en el caso mexicano, por detenernos sólo en uno de los mencionados, en 2023 el gobierno federal “actualizó” el Registro Nacional de Personas Desaparecidas y no Localizadas sin contar con el consenso de funcionarios de la Comisión Nacional de Búsqueda, de organizaciones en defensa de derechos humanos ni de colectivos de búsqueda de personas desaparecidas, lo cual causó un debate importante alrededor de los modos en que se dimensiona la desaparición de personas en el país, el tipo de atención que pueden recibir las familias de las víctimas en función de si el número de registros disminuye de manera drástica y el lugar que tiene en la agenda pública la defensa de los derechos humanos. Los datos en contextos como éste suponen un objeto de disputa política en sí.

Ampliar la mirada más allá de las preocupaciones de acceso a libertades personales y garantías institucionales alrededor de la datificación dominante es voltear a donde estas prácticas se dan en contextos de fragilidad institucional, corrupción gubernamental y crisis de violaciones a los derechos humanos. Esto contribuye de manera particular a la construcción de “derechos de los datos” y a la prevención de “daños con los datos” (Kitchin, 2022), que son parte de los supuestos teóricos y políticos de los estudios críticos de datos.

Datificación de la violencia en contextos latinoamericanos

Dentro de la contextualización y reorientación de temas de análisis relevantes para los estudios críticos de datos en contextos latinoamericanos –particularmente en países como Colombia y México– es urgente pensar en los procesos de datificación de la violencia. La conformación de flujos digitales de información y comunicación acerca de violaciones graves a los derechos humanos en estos países se da en un contexto de registros de este tipo de violencias desde comisiones de la verdad, dependencias públicas, medios alternativos y ONG’s que buscan esclarecer acontecimientos del pasado y del presente.

³ Países latinoamericanos en los cuales se han incrementado las amenazas y los ataques a periodistas en los últimos años.

A continuación, se mencionan de forma breve algunos casos recientes de datificación de la violencia en Colombia y en México, que resultan relevantes como marcos de referencia para el desarrollo teórico y metodológico de los procesos de datificación de la violencia en general.

En junio de 2023, en el marco del primer aniversario de la entrega del informe final de la Comisión de la Verdad en Colombia, el grupo de Análisis de Datos de Derechos Humanos (HRDAG, por sus siglas en inglés), publicó en su sitio oficial un enlace al archivo digital de la Comisión, enmarcándolo como “la base de datos más grande en la historia del movimiento de los derechos humanos” (“HRDAG - Human Rights Data Analysis Group” 2023). La base de datos es producto de una colaboración entre la Comisión de la Verdad, Justicia Especial para la Paz y el Grupo de Análisis de Datos de Derechos Humanos. Este archivo contiene 24 millones de registros sobre el conflicto armado en Colombia. Los registros digitales acompañan 11 tomos escritos que abordan de manera amplia el conflicto armado colombiano y buscan alimentar el diálogo nacional para el reconocimiento de las víctimas y la visibilización de los efectos actuales que ha dejado la violencia.

En 2021, Forensic Architecture publicó el proyecto *Dispossession and the memory of earth*,⁴ dentro del cual se aloja la plataforma *Despojo Urabá* (Forensic Architecture 2021), que da cuenta de procesos de despojo y acaparamiento de territorios en esta zona de Colombia desde 1950, con el fin de identificar a grupos y familias que se han beneficiado de diversas condiciones en el país para crecer y/o fortalecerse como élites locales. Este proyecto brinda insumos importantes para desarrollar formas de contramapeo (Dalton y Stallman, 2018) y es un espacio de consulta pública para la divulgación de acontecimientos relevantes a nivel local.

*El conflicto armado en cifras*⁵ es un micrositio albergado en la página del Centro Nacional de Memoria Histórica, el cual clasifica en 11 tableros de manera exhaustiva los siguientes abusos cometidos en el conflicto armado colombiano: acciones bélicas, asesinatos selectivos, ataques a la población, atentados terroristas, daño a bienes civiles, desaparición forzada, masacres, minas antipersonales, artefactos explosivos y munición sin explotar, reclutamiento y utilización de niños, niñas y adolescentes, secuestro y violencia sexual. Este sitio ofrece acceso directo a las bases de

⁴ <https://forensic-architecture.org/investigation/land-dispossession-in-nueva-colonia>

⁵ <https://micrositios.centrodememoriahistorica.gov.co/observatorio/portal-de-datos/el-conflicto-en-cifras>

datos para que sean utilizadas y analizadas por quien quiera acercarse a ellas para estudiar el conflicto armado en Colombia. Cada uno de los tipos de abusos mencionados contiene dos bases de datos: la de registros del abuso y la de víctimas del abuso.

En el caso de México, la versión pública del Registro Nacional de Personas Desaparecidas y No Localizadas⁶ (RNPDNO) contuvo, hasta octubre de 2023, 112,485 registros de personas desaparecidas a nivel nacional. Este registro ha sido utilizado por medios de comunicación, academia y por colectivos de personas en búsqueda de sus familiares, para hacer consultas, estudios e investigaciones y en ocasiones para evidenciar sesgos, subregistros o errores en el archivo.

En mayo de 2022, Quinto Elemento Lab publicó la investigación *Fragmentos de la desaparición*,⁷ en el marco de la llegada del registro oficial a 100 mil personas desaparecidas en México. En esta investigación se segmentan los datos sobre personas desaparecidas para analizar el crecimiento del registro durante tres sexenios en México, la emergencia de protestas y de colectivos –sobre todo de madres– de búsqueda de personas desaparecidas en el país y se identifican algunos patrones a nivel regional y nacional para intentar comprender las causas y ampliar el contexto en el que ocurren las desapariciones.

Por último, en 2023 Data Cívica e INTERSECTA, ambas organizaciones no gubernamentales, publicaron el trabajo *Con Copia Oculta*⁸ que consiste en la obtención y análisis de una base de datos de más de 50 mil registros de enfrentamientos de las fuerzas armadas con grupos del crimen organizado. La existencia de esta base de datos fue negada por el gobierno de Andrés Manuel López Obrador, pero una vez realizado el hackeo internacional masivo por parte del grupo Guacamaya, estas organizaciones buscaron y encontraron estos registros en los correos masivos filtrados, lo cuales dan cuenta de sesgos en la captura de bajas y en el perfilamiento, sistemáticamente de sospecha o culpabilidad, de las personas que perdieron la vida en estos enfrentamientos con las fuerzas armadas. Las organizaciones anonimizaron la base de datos y ofrecen acceso abierto a la misma, como una invitación a realizar más investigaciones con ella.

⁶ <https://versionpublicarnpdno.segob.gob.mx/Dashboard/Index>

⁷ <https://quintoelab.org/fragmentos/cien-mil-despariciones-mexico/>

⁸ <https://concopiaoculta.org/>

Con lo expuesto hasta ahora, es posible afirmar que en países como Colombia y México el registro y análisis de violaciones graves a los derechos humanos está enmarcado en la exploración de formas de apropiación crítica de prácticas de datificación de la violencia por parte de actores como medios alternativos, dependencias públicas, comisiones de la verdad y ONGs. Se evidencian, además, posibilidades posteriores de estudiar estos registros partiendo del señalamiento crítico de cada trabajo a cada uno de los aspectos de la datificación dominante, mencionados más arriba.

Antecedentes como estos son fundamentales para el desarrollo de aproximaciones alternativas al estudio de la datificación de la realidad, que contribuyan a un tipo de problematización –anclada a contextos latinoamericanos en particular– y a una datificación crítica de la violencia en general. Hacia allá construiremos algunos insumos en los siguientes apartados.

Principios conceptuales para una datificación crítica de la violencia

Problematización y densificación del dato

De acuerdo con lo planteado hasta ahora, conviene empezar a delinear aspectos conceptuales y preguntas generales hacia la construcción de una datificación alternativa o crítica de la violencia, centrada en elementos cualitativos de los datos, fundada a partir de experiencias de trabajo con archivos de procesos históricos y/o registros de experiencias de trauma social, y con conceptos que partan de andamiajes teóricos cuidadosos de poner en perspectiva la posición epistemológica –sus límites y sus alcances– de la que emergen. Esto no quiere decir que esta propuesta busca excluir exploraciones e interpretaciones que vengan de análisis numéricos de los datos, sino que busca colocar a éstos fuera del centro del análisis y en muchas ocasiones utilizarlos como punto de entrada para realizar parte del análisis, mas no como el objetivo principal. En cierta medida, lo que se busca es desestabilizar la noción dominante de dato y datificación centrada en lo numérico, mediante el señalamiento y la acumulación de cualidades que hay en los datos en tanto registro de algo que no se agota en medir, es decir, construir condiciones de posibilidad para hacer preguntas a las bases de datos de registros de violaciones a los derechos humanos desde sus aspectos valores categóricos con fines críticos.

Desde los estudios de *performance*, para Diana Taylor (2011) el *archivo* está conformado por todo el material tangible o “núcleo duradero de registros” (fotografías, huellas dactilares, libretas, documentos digitales, hallazgos forenses, etcétera) relacionados con la violencia ejercida para borrar, en casos como los de dictaduras militares, vidas y memorias de las víctimas; en tanto que el *repertorio* “almacena y recrea la memoria ‘encarnada’ –los ‘estremecimientos’ traumáticos o catárticos, gestos, movimientos orales, bailes, canciones– en resumen, todos los actos que suelen considerarse de conocimiento o “en vivo”, efímero, no reproducible” (Taylor y Fuentes, 2011). Es en las posibilidades de conjugación entre lo tangible y lo no reproducible, entre lo articulable y lo inconceptualizable que los estudios del *performance* llevan a cabo ejercicios para evidenciar los intersticios en las relaciones de poder dominantes. ¿De qué preguntas tendría que partir un proceso de datificación que busque integrar lo “no reproducible” del repertorio performático señalado por Taylor? ¿Es posible transmitir la memoria traumática a través de los datos masivos que registran la violencia? ¿Es posible construir un dato que tenga estos fines? ¿Cómo se representaría este dato? ¿Qué intersticios en las relaciones de poder sería posible evidenciar a través de los registros datificados de acontecimientos traumáticos? ¿En qué tendría que centrarse una memoria traumática datificada, además de en la reivindicación de la dignidad de las víctimas, en la de sus familias y en señalar a los culpables de las atrocidades? Estos cuestionamientos preliminares resultan útiles para comenzar a delimitar una agenda de investigación y de construcción de insumos para una datificación crítica de la violencia.

Así como Taylor contrapone un tipo de imagen social que aleja a las víctimas de desaparición del papel de “criminales o subversivos” en el que los colocan las imágenes “aisladas, precisas, no artísticas” de los registros oficiales, podríamos pensar en un tipo de “dato anclado” o “dato social”, que a diferencia del dato crudo (*raw data*) (Kitchin y Lauriault, 2014) del que parte la datificación dominante (cuantitativa-predictiva-vigilante-comercial-universalizante), puede asumir como su lugar de enunciación el desafío a las asimetrías en el pronunciamiento sobre una realidad traumática proveniente de violaciones graves a los derechos humanos. Podemos pensar en un dato social o anclado que parte de nombrar lo real desde una versión *invisibilizada o invalidada* por las autoridades o por una porción de la opinión pública, y que busca hacer visibles y legítimas otras versiones –las de las víctimas– y/o señalar a quien les causó el daño.

Esta densificación del dato toma en cuenta estos aspectos y permite que nos preguntemos por el peso que pueden tener distintos grupos sociales relevantes (además de gobiernos, medios de comunicación y empresas) alrededor del significado y de los relatos contados con los datos disponibles acerca de violaciones a los derechos humanos, por las posibilidades en la construcción de imaginarios acerca de periodos y prácticas sociales específicas a partir de estos relatos, así como por la exploración de toma de decisiones alternativas con análisis no centrados ni agotados en lo numérico de los datos ni en versiones que se empeñen en invisibilizar a las víctimas. Es decir, esta densificación del dato permite pensar en formas de datificación crítica que dialoguen y se alimenten de aspectos presentes en los proyectos acerca de la violencia en Colombia y México vistos más arriba.

Aunado a lo anterior, la integración de otras miradas que contribuyan al análisis y al registro de un tipo de datos masivos centrado en lo cualitativo, ha de tomar en cuenta el registro de la oralidad, de lo popular o del silencio inscrito en la documentación disponible. Así, podemos avanzar en más cuestionamientos preliminares como: ¿Qué tipo de registros datificados de la violencia en una región, o sobre una población en específico, se pueden construir, tomando en cuenta no únicamente la racionalidad occidental, sino la de la oralidad local, las tradiciones populares o los mitos regionales?, ¿de qué filtros lingüísticos, gestuales, partiría este registro?, ¿qué preguntas intentaría resolver, o mejor, qué tipo de “paisajes habitables” (mitos) buscaría crear y/o indagar?

Esta propuesta de un tipo de datificación crítica de la violencia contempla estas cuestiones para densificar conceptual y políticamente al registro y al análisis que se haga de los datos. Los registros digitales de violaciones a los derechos humanos intentan dar cuenta de aspectos de la vida que, en términos biográficos y como experiencias sociales traumáticas, no son medibles, pero sí han sido cuantificados. Esto ya ha sido abordado, por ejemplo, desde la etnografía a los indicadores de abusos a derechos humanos que realizó Merry, cuyo principio crítico es precisamente señalar los límites de medir lo inconmensurable (2016). El dato anclado o social que se intenta construir aquí busca desarrollar y responder preguntas del tipo: ¿Qué queda fuera de lo medible? ¿Se pueden hacer registros datificados de otro tipo, anclados a contextos y/o experiencias cualitativas, sin que esto suponga excluir a lo cuantitativo, sino utilizarlo como punto de entrada? ¿Es posible pensar en un tipo de registro datificado que pueda

ampliar los horizontes ontológicos de la datificación de la realidad hacia perspectivas que utilicen y registren la experiencia sin intentar agotarla en el propio lenguaje del registro, es decir, que contribuyan a dar cuenta de la realidad no únicamente midiendo las violencias, sino haciendo visible la inconmensurabilidad de la experiencia vivida por las víctimas de violaciones graves a los derechos humanos?

Si pensamos a partir de los aspectos intangibles del repertorio performático y de la inconmensurabilidad del registro de la violencia, cabe preguntar ¿es posible trabajar en análisis y visualizaciones de datos que desde una reorientación crítica del registro de la experiencia, puedan proponer una lectura o “montaje performático” de los datos, donde lo inconmensurable de la pérdida a nivel biográfico y social registrado en categorías cerradas muestre, por saturación del conteo de la tragedia, la insuficiencia de comprensión y explicación del fenómeno mediante la cuantificación del mismo?

Este tipo de insumos conceptuales para una datificación crítica de la violencia asumen la vigilancia epistemológica constante sobre los sesgos de lo registrado –y sobre las implicaciones de lo no registrado– en las bases de datos, así como un diálogo con otros análisis sobre experiencias similares –o sobre las mismas experiencias– para enriquecer los puntos de vista alrededor de lo que busca entender y analizar, y de sí mismo como recorte situado y falible de partes específicas de la realidad.

Asimismo, estas primeras aproximaciones para una datificación crítica de la violencia se suman al énfasis y al llamado a la exploración y generación de apropiaciones críticas de los datos y de las herramientas para trabajar con datos, a la “práctica inventiva con datos... para explorar cómo una reflexión crítica, histórica y sociológica acerca de las infraestructuras de datos puede convertirse en trabajo práctico con datos como parte de lo que podríamos llamar *práctica crítica de datos*” (Gray, Gerlitz, y Bounegru, 2018).

Los estudios críticos de datos apelan a “la producción de nuevos elementos del entorno digital, que hagan uso de los elementos existentes para producir nuevas verdades” (Ettlinger, 2018), así como a atender “una mirada de maneras no convencionales de imaginar/pensar/sentir los datos que emerge desde los bordes, aportando a la creación de prácticas alternativas de datos” (Milan y Treré, 2019). Esta es una respuesta a esos llamados, articulada desde registros digitales de la violencia en contextos latinoamericanos.

El dato dominante como infraestructura simbólica y el dato anclado como propuesta de ampliación de desacuerdos

Ahora bien, pensar en la datificación dominante y en sus cualidades ya delimitadas más arriba, también nos lleva a pensar en las infraestructuras simbólicas, que es como Susan Leigh Star piensa a los conceptos que sostienen la producción de conocimiento científico. Para la autora, todo tipo de infraestructura simbólica no puede ser considerada únicamente como el “detrás de cámaras” de la producción de saberes o como “las cosas sobre las que otras cosas pasan por encima” (Star, 2002), sino que funciona como punto de acceso para evidenciar y explorar mecanismos de generación de clasificaciones y ordenamientos legitimados, que son los que posibilitan y condicionan la producción de saberes científicos.

A partir de la desnaturalización de dichas infraestructuras del conocimiento –como lo propone la autora– es posible un cuestionamiento epistemológico que problematice las implicaciones y consecuencias de la clasificación validada por gran parte de la comunidad científica, al hablar de “datos” pensando únicamente en datos masivos y particularmente en sus aspectos numéricos en el contexto que nos atañe. Esta normalización es un efecto de una infraestructura simbólica poco o nada cuestionada por más de una comunidad que trabaja, probablemente, con distintos tipos de datos; pero que ha naturalizado el énfasis en lo numérico, como el aspecto fundamental a ser analizado en muestras masivas de éstos, o que incluso no considera posible o válido otro tipo de análisis de grandes volúmenes de datos. De este modo, la desestabilización de esta infraestructura simbólica supone la posibilidad de estar frente a bases de datos –particularmente las referidas a registros de violaciones a derechos humanos– pensadas para ser analizadas colocando sus aspectos cualitativos al centro, con el objetivo de ampliar los desacuerdos alrededor de las formas válidas de producción de conocimiento con ellos y al mismo tiempo abrir las posibilidades de generar “extrañamientos” o “bordes del lenguaje” (como los llama Star) mediante un nuevo tipo de análisis del fenómeno registrado, que precisamente lleve a formas de producción de conocimiento alternativo. Esta bifurcación no supone la descalificación de análisis numéricos, sino la ampliación de posibilidades analíticas de los datos con fines críticos.

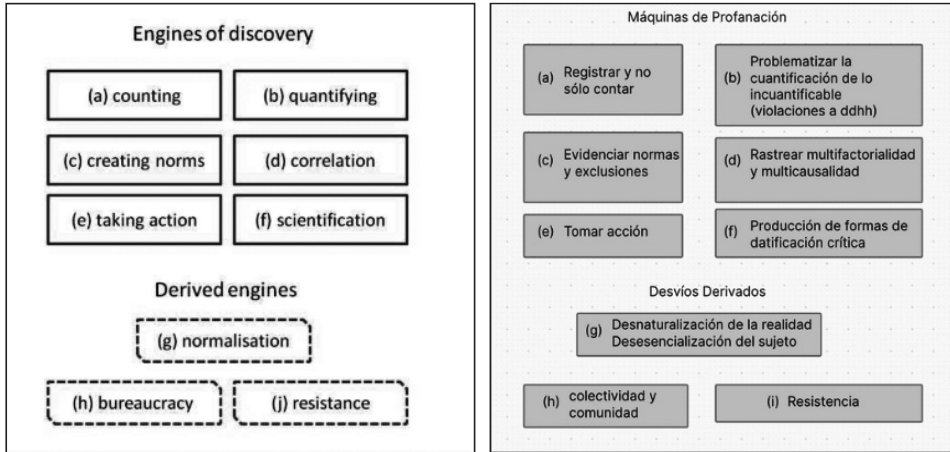
Máquinas de profanación y protocolos iconoclastas

En su propuesta de ensamblajes de datos, Kitchin y Lauriault (2014) señalan desde lo infraestructural a los mecanismos que operan como una suerte de bucle o camino entre instituciones, discursos y prácticas sociales con las que los procesos de datificación de la realidad se van concatenando para reproducirse, volverse referentes de sí mismos y consolidar la lectura de la realidad que datifican.

Los ensamblajes producen dos dinámicas que alimentan y son alimentadas por circuitos de producción-recolección de datos y generación de inteligencia digital: el “efecto loop” y los “motores de descubrimiento”. Estos últimos operan como un engranaje entre el conteo de información, que refuerza la legitimidad de la cuantificación como registro válido de la experiencia, la creación de normas y correlaciones “neutrales” a partir de esta cuantificación, la toma de acciones para “cientificar” todo el procedimiento y la reafirmación de la jerarquía de los hallazgos y de los métodos de datificación en cada ensamblaje. El primero remite a la reificación del proceso como forma de ordenamiento epistemológico y político sobre lo real (lo que aquí hemos señalado como la epistemología universalizante de la datificación dominante).

Este modelo resulta útil para estudios que buscan comprender cómo se construyen y entrelazan sistemas de regulación y normalización de la realidad a partir de sus respectivos ensamblajes de datos, pero también es útil para imaginar posibilidades de ensamblajes que partan de una epistemología alternativa de los datos, es decir, que no coloquen en el centro la mirada pragmática y, podríamos decir, “subjetivamente objetiva” de la datificación dominante como eje del discurso de verdad que buscan producir. En este sentido, cabe pensar en ensamblajes que no tengan como objetivo calcular o predecir comportamientos, sino evidenciar, por ejemplo, visiones de un aspecto de la realidad sobre el que otro tipo de ensamblajes amplían o consolidan discursos de exclusión o de perpetuación de un tipo de violencia. Se trata de una posible intervención, o profanación, de los efectos de los ensamblajes de datos y de sus motores que cuestione los alcances del registro, problematice los límites de la cuantificación de la experiencia, evidencie sus exclusiones, rastree multifactorialidades (en lugar de sólo buscar correlaciones) genere preguntas y parta de la desnaturalización de la realidad y la des-esencialización de los sujetos cuyas acciones han sido registradas.

FIGURA 3
Izquierda. Ensamblajes de datos en Kitchin y Lauriault (2014). Derecha. Máquinas de profanación.



Fuente: elaboración propia.

Ahora bien, desde la archivística, las bases de datos son archivos en el sentido tradicional del término, es decir, se trata de datos que se convierten en conocimiento, conocimiento que se convierte en información e información que se convierte –o puede hacerlo– en evidencia (Millar, 2017). Pero los archivos que se toman como neutrales u objetivos también pueden contribuir a una lectura que petrifique lo registrado, o lo encasille como un pasado acabado. Tanto la ontología numérica como la gubernamentalidad de vigilancia y control –dos de los aspectos mencionados como constitutivos de la datificación dominante– alimentan esta petrificación del significado del registro, tanto más si recordamos que el dato numérico suele ser pensado como “objetivo” o “neutral”.

A este respecto, Ariella Azoulay, desde una reflexión crítica del archivo fotográfico israelí acerca de la toma del territorio del pueblo palestino, señala una “actitud instrumental” en la fotografía, que opera cuando “iconizamos” la imagen, es decir, cuando la cosa o persona que aparecen en una fotografía sustituyen al acto del registro que supone la propia fotografía, de modo que se suprime el acto mismo del registro como *lugar y práctica de mediación de la realidad*, y al hacerlo, se convierte al registro en transparencia, en neutralidad, en objetividad; esto es, se dota al registro de una

suerte de soberanía sobre lo real, mediante la cual llega a sustituir lo que registra. La autora nos recuerda entonces que “la fotografía no documenta un concepto o demarca un evento, sino que es más bien un documento, el producto de un evento común para varios participantes” (Azoulay, 2014).

Dicho lo anterior, podemos afirmar que, al igual que una fotografía, el dato numérico tiende a ser instrumentalizado, y su “acto de registro” suprimido, cuando se piensa que el dato abarca al sujeto cuya práctica está documentando. Para contrarrestar esta actitud instrumental, Azoulay propone llevar a cabo un “protocolo iconoclasta” para “destruir” los íconos producidos por las imágenes y restituirles su cualidad de documentos, es decir, para situarlos y delimitarlos en tanto registros llevados a cabo en circunstancias históricas y políticas específicas.

Así, la datificación crítica de la violencia propone una mirada profana de los ensamblajes de datificación dominantes sobre violaciones a los derechos humanos, así como ejercer protocolos iconoclastas frente a la supuesta transparencia u objetividad del dato numérico. Estas conceptualizaciones contribuyen a la problematización tanto del dato como del proceso de datificación, buscando siempre cuestionar cuáles son los conceptos y supuestos fijados a los archivos.

Diseño de anclajes empíricos para una datificación crítica de la violencia

Como ha sido mencionado, los estudios críticos de datos han generado un importante repertorio de conceptos para posicionarse frente a distintos problemas provenientes de los procesos de datificación dominante. Sin embargo, la producción metodológica en el trabajo *con*, y no sólo *acerca de*, los datos no han ido a la par de la efervescencia conceptual en el campo. Para atender a los llamados de mayor exploración metodológica desde distintas voces, la datificación crítica de la violencia propone la generación de bases de datos propias, la transformación de textos regulares en bases de datos estructuradas y la segmentación de muestras pensadas desde valores categóricos. Al mismo tiempo, dialoga con posibilidades de construcción de anclajes empíricos desde una primera revisión de aspectos generales de los métodos digitales (Rogers, 2023) y de la lingüística de corpus (Hunston, 2022), pues ambas son propuestas que llevan a cabo análisis de bases

de datos masivos desde preguntas críticas, y si bien parten de recortes numéricos, su foco analítico no descuida valores categóricos.

La obtención y el manejo de datos masivos en los últimos años ha mantenido abierta la conversación acerca de las posibilidades de métodos mixtos y cualitativos como posibles marcos de aproximación a corpus de este tipo. Al respecto, Elisenda Ardévol (2017) ha identificado las coincidencias entre la ciencia de datos y los enfoques hermenéuticos, particularmente la etnografía, señalando que ambos están interesados en “hallar patrones para la investigación” que puedan ser tomados como insumos susceptibles de ser interpretados. Para la autora “modelizar el comportamiento humano” no es suficiente para llevar a cabo una interpretación de aspectos culturales a partir de patrones, sino que “sólo una descripción densa puede revelar el significado cultural, si tiene alguno, de estas conexiones”. En esta misma línea, Rodríguez Cano (2022) se ha pronunciado por evidenciar que “aunque sean inmensos, los datos se pueden interpretar cualitativamente”, siempre y cuando se mantenga una vigilancia epistemológica permanente frente a los datos, los métodos y los marcos interpretativos de los que se parta.

Los métodos digitales son un conjunto de técnicas útiles tanto para los estudios de medios (*media studies*) como para la investigación social (*social research*) (Rogers, 2023). La *Digital Methods Initiative* ha desarrollado un conjunto de herramientas para analizar lo “nativamente digital”, es decir, aquello que ha nacido dentro de internet. Este tipo de herramientas pueden ser utilizadas como puntos de acceso para la segmentación de bases de datos, con el fin de llevar a cabo análisis críticos más profundos centrados en aspectos cualitativos alrededor, pero no únicamente, de circuitos de producción, circulación y consumo de plataformas digitales.

Por su parte, la lingüística de corpus se refiere a una aproximación “al estudio del lenguaje que incluye recolectar grandes candidades de lenguaje ocurrido de manera natural y la utilización de *software* especializado que manipula ese lenguaje para obtener información acerca de frecuencias, co-ocurrencias y significado” (Hunston, 2022). Este campo ha desarrollado un amplio repertorio de conceptos y fórmulas para medir frecuencias, identificar y comparar estructuras sintácticas, así como para categorizar usos y orientaciones semánticas en corpus que pueden contener desde algunos cientos hasta millones de palabras.

El desarrollo de herramientas desde la lingüística de corpus ha ido de la mano de preguntas y reflexiones alrededor de las mutaciones en la

presentación del propio texto, en la medida en que tanto los soportes digitales como las herramientas para gestionar palabras han desbordado la linealidad de la escritura tradicional. Al respecto, Mayaffre y Poudat (2020) se refieren a formas de lectura “reticulares”, “tabulares” e “hipertextuales” mediante procesadores tabulares, desde hipervínculos de páginas web o incluso desde visualizaciones de tipo red, que obligan a lecturas con una jerarquía que no sólo es la del orden sintáctico de las oraciones.

Estas aproximaciones metodológicas nos permiten pensar en el diseño de anclajes empíricos para una datificación crítica de la violencia, mientras sean orientadas con los principios conceptuales planteados. Para ello, a continuación se mostrarán ejemplos de producción, segmentación y visualizaciones de bases de datos que permitan construir preguntas e intuiciones iniciales para el tipo de análisis al que nos referimos.

I

El volumen titulado *Cuando los pájaros no cantaban*, también llamado “volumen testimonial” del Informe final de la Comisión de la Verdad en Colombia tiene 515 páginas en total. Mediante el uso de AnConc, una herramienta de lingüística de corpus, es posible cargar el archivo completo como archivo “raw” para poder realizar búsquedas de enigramas (conjuntos de distinto número de palabras), colocaciones (identificación de palabras alrededor de una consulta específica dentro de todo el texto), palabras clave, entre otras segmentaciones del texto. La herramienta convierte en columnas y en filas los registros de palabras, es decir, convierte un texto regular en una base de datos estructurada. El engrama “a las” aparece 376 veces a lo largo del documento. Una mirada segmentada de conjuntos de tres palabras que inicien con ese “a las” nos muestra cientos de casos donde lo que sigue a dicha conjunción es la hora (a las 4, a las 7, a las 5, etcétera). ¿Qué significa, en un tomo compuesto de testimonios de víctimas de la violencia, que la hora a la que “algo terrible” ocurrió emerja de manera tan contundente? Si las experiencias traumáticas tienden a la suspensión de la conceptualización por parte del sujeto, ¿qué descripciones y análisis de estos fragmentos del relato y sus conexiones dentro del mismo pueden emerger con este tipo de hallazgos?

FIGURA 4
Capturas del proceso de conversión del volumen testimonial de la Comisión de la Verdad a base de datos y filtrado de clusters con la conjunción “a las” desde AntConc.

The screenshot displays the AntConc interface with the following components:

- Target Corpus (Left Pane):**
 - Corpus name: en_testimonial_CVC.db
 - Description:

Category	Description
full_name	Volumen_testimonial_CVC
short_name	Volumen_testimon
file_count	1
token_count	275809
type_count	20106
encoding	utf_8_sig
token_definiti	[\p(L)]+
ignore_header	False
ignore_items	False
number_repla	False
format	raw_files
indexer_type	type
indexer	simple_word_indexer
- Target Corpus (Top-Right Pane):**
 - Name: Volumen_testimon
 - Files: 1
 - Tokens: 275809
 - Search Query: a las
 - Cluster Types: 95, Cluster Tokens: 376, Page Size: 100 hits
 - Table of Clusters:

Cluster	Rank	Freq	Range
1 a las cinco	1	30	1
2 a las siete	2	28	1
3 a las seis	3	25	1
4 a las personas	4	21	1
5 a las ocho	5	20	1
6 a las dos	6	18	1
7 a las cuatro	7	17	1
8 a las tres	7	17	1
9 a las doce	9	14	1
10 a las once	9	14	1
11 a las diez	11	12	1
12 a las nueve	11	12	1
13 a las farc	13	11	1
14 a las mujeres	14	8	1
15 a las que	14	8	1
16 a las comunidades	16	6	1
- Target Corpus (Bottom Pane):**
 - Name: Volumen_testimon
 - Files: 1
 - Tokens: 275809
 - N-Gram Types: 1/116404, N-Gram Tokens: 376/275808, Page Size: 100 hits
 - Search Query: a las
 - Table of N-Grams:

Type	Rank	Freq	Range
1 a las	1	376	1

II

Uno de los lexicones publicados en el archivo digital de la Comisión de la Verdad en Colombia⁹ contiene una lista de palabras categorizadas por emociones (miedo, tristeza, repulsión, enojo, sorpresa y alegría). Con esta lista es posible generar un archivo de aristas donde cada palabra esté conectada con la emoción con la que la Comisión la categorizó. Esta base de datos, procesada con Gephi¹⁰ para generar una red dirigida compuesta de 5,568 nodos, muestra las categorías afectivas propuestas por la Comisión como nodos centrales en la red. “Sin categoría” es el nodo con mayor número

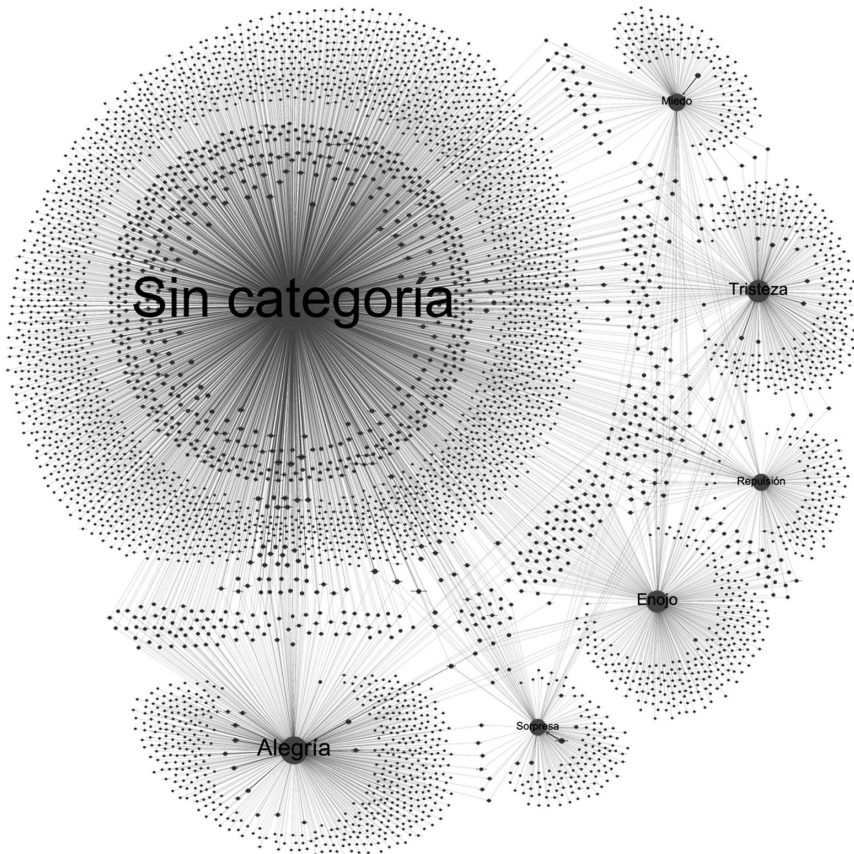
⁹ <https://www.comisiondelaverdad.co/analitica-de-datos-informacion-y-recursos#c2>

¹⁰ <https://gephi.org/>

de conexiones, ya que muchas de las palabras puestas en este lexicon no fueron relacionadas con ninguna de estas emociones. El resultado es una representación visual que toma lo numérico (las conexiones de cada emoción) de una base de datos como punto de acceso, no como objetivo, para dar cuenta de lo que queda fuera de la medición y abrir espacio a representaciones y análisis de lo que está en otra capa de estos mismos registros.

Figura 5

Red de interacciones entre palabras categorizadas en uno de los lexicones generados por la Comisión de la Verdad en Colombia. Los *hubs* de esta red son las afectividades seleccionadas por la Comisión.

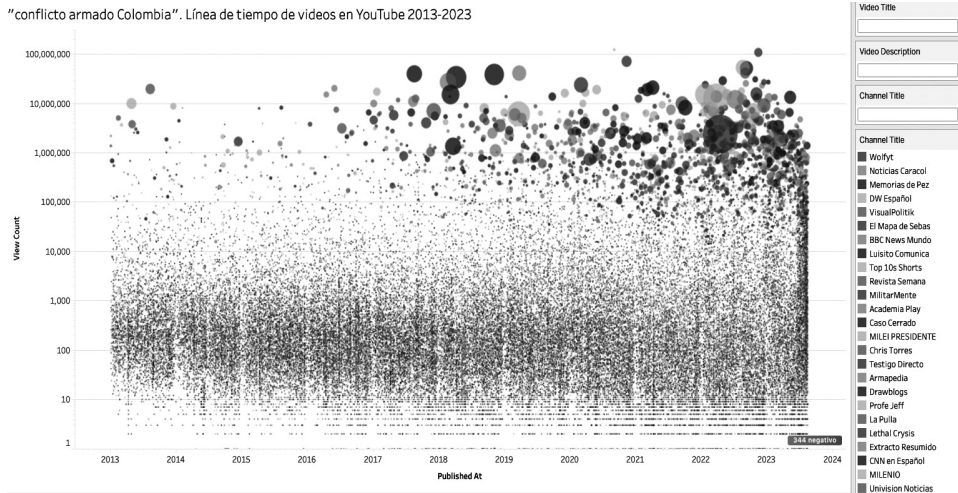


Fuente: elaboración propia.

III

Las plataformas digitales se han convertido en ágoras públicas donde se disputa el sentido de los acontecimientos sociales en la actualidad. La datificación de la violencia no sólo puede integrar bases de datos con registros de violaciones a los derechos humanos, sino discursos alrededor de estos acontecimientos. El portal de herramientas YouTube Data Tools, desarrollado por la Digital Methods Initiative, permite realizar consultas para descargar listas de videos acerca de un tema en específico en esta. Para este ejercicio se llevó a cabo una descarga de la consulta “conflicto armado Colombia”, de enero de 2013 a agosto de 2023. El resultado arrojó 66,987 videos, la base de datos contiene el título de cada video, el nombre del canal al que pertenece, la categoría a la que pertenece el canal (noticias, blogs, música, etc.), el número de vistas y comentarios hasta la fecha de descarga y la fecha exacta de publicación. Con estos datos, visualizados en una línea de tiempo interactiva y con búsqueda y filtro de palabras activados en Tableau, es posible identificar momentos en el tiempo, palabras clave en los títulos, procedencia geográfica de los canales o si se trata de contenidos de medios de comunicación, gobiernos, ONG's u otros actores políticos relevantes. Estas visualizaciones no intentan agotarse en los números de circulación de contenidos, sino explorar el contexto político en el que se pusieron en circulación modos de construcción de la mirada sobre la violencia, las víctimas o las posibilidades de paz en Colombia, es decir, permiten interrumpir los modos en que se puede confundir al registro de un acontecimiento con el acontecimiento en sí.

FIGURA 6
Visualización interactiva de videos de YouTube publicados de 2013 a 2023
con la consulta “Conflicto armado Colombia”.



Fuente: Elaboración propia. Disponible en: https://public.tableau.com/app/profile/vicabr/viz/YT_conflictoarmadocolombia_130101-230816_videolist/Historia1

Conclusiones

Este artículo ha señalado el panorama general y los énfasis conceptuales principales de los estudios críticos de datos. Como resultado de esto, ha propuesto una clasificación de sus énfasis en panópticos, decoloniales y neomarxistas, de apropiaciones críticas, de tensiones entre agencia y estructura e infraestructurales. Esta clasificación ha llevado a pensar en un tipo de datificación dominante, caracterizada por una ontología numérica, una racionalidad predictiva, una gubernamentalidad de vigilancia y control, una ética comercial y una epistemología universalizante de los datos.

Entre las áreas del conocimiento que se suman a la consolidación del campo de los estudios críticos de datos resaltan las ciencias sociales, la geografía y la comunicación. Esta última puede aportar desde su tradición en el estudio y apropiación de mediaciones técnicas de interacción social, y desde sus bases teóricas que desarticulan la supuesta objetividad del registro y del montaje de la realidad social desde mediaciones técnicas.

Al panorama de preocupación por la pérdida de garantías democráticas que emerge desde la datificación dominante, se ha llamado a atender procesos de datificación de la realidad insertos en contextos de fragilidad institucional, corrupción gubernamental y crisis de violaciones a los derechos humanos. En contextos latinoamericanos como Colombia y México, hay acercamientos a estos fenómenos realizados desde ONG's, medios alternativos, dependencias públicas y comisiones de la verdad.

Se ha presentado una serie de principios conceptuales para diseñar una datificación crítica de la violencia, constituida por una *racionalidad cualitativa* que parte de la reivindicación de lo inconmensurable registrado en datos masivos acerca de violaciones a los derechos humanos, una *política performática* que busca evidenciar los intersticios en las relaciones de poder imbuidas en la datificación, una *ética testimonial* que explora relatos invisibilizados o invalidados de las víctimas y una *epistemología crítica* que toma al dato como una infraestructura simbólica intervenible mediante la restitución de su condición de registro de la experiencia y no de sustituto de la experiencia.

Para trabajar en el diseño de anclajes empíricos a partir de estos principios conceptuales, se han colocado algunas herramientas y métodos provenientes de la lingüística de corpus y de los métodos digitales, enmarcados en la conversación actual acerca de las posibilidades de desarrollo de análisis cualitativos de bases de datos masivas. Los cruces en las bases de datos y las visualizaciones resultado de esta descentralización de lo numérico en los datos masivos tienen que ser entonces tomados no como “resultados” en sí, sino como espacios de inmersión crítica, explorables a partir de la generación de rutas de indagación, que busquen generar y hacer visibles relatos que den cuenta de la experiencia registrada fuera del marco epistemológico de la medición de dicha experiencia.

Por último, un aspecto relevante sobre el que habrá que seguir guiando la problematización del dato radica en entender que trabajar con datos es trabajar con distintos “mundos de datos” (Kitchin y Lauriault, 2014), mas no únicamente como muestras, sino como visiones o “cosmovisiones”, despliegues semióticos amplios producidos con datos, los cuales se activan al registrar y enmarcar la realidad con y desde los datos. Por lo tanto la problematización de los mundos datificados que aquí interesan –los de la violencia– puede profundizarse a partir de reflexiones acerca del dato y de la violencia como signos. Hacia allá habrá que hacer orientar nuevas exploraciones.

Bibliografía

- Ardévol, Elisenda. 2017. "Big Data y Descripción Densa." *Virtualis*, no. 7.
- Azoulay, Ariella. 2014. *Historia Potencial y Otros Ensayos*. México: CONACULTA.
- Barreneche, Carlos. 2019. ¿ "Data Corruption: The Institutional Cultures of Data Collection and the Case of a Crime-Mapping System in Latin America." *Canadian Journal of Communication* 44, no. 3: 343–50. <https://doi.org/10.22230/cjc.2019v44n3a3481>
- Barreneche, Carlos, Andrés Lombana Bermúdez, and Juan Ramos Martín. 2021. "Datificación en Contextos de Corrupción: Imaginarios Sociotécnicos y Prácticas de Resistencia Frente a Sistemas Antipobreza en Colombia." *Palabra Clave* 24, no. 3: 1–39. <https://doi.org/10.5294/pacla.2021.24.3.4>
- Dalton, Craig M., and Tim Stallmann. 2018. "Counter-Mapping Data Science." *Canadian Geographies / Géographies Canadiennes* 62, no. 1: 93–101. <https://doi.org/10.1111/cag.12398>
- Ettliger, Nancy. 2018. "Algorithmic Affordances for Productive Resistance." *Big Data & Society* 5, no. 1: 2053951718771399. <https://doi.org/10.1177/2053951718771399>
- Forensic Architecture. 2021. "Uraba Transaction Network." <https://despojo-uraba.forensic-architecture.org/>
- Gray, Jonathan, Carolin Gerlitz, and Liliana Bounegru. 2018. "Data Infrastructure Literacy." *Big Data & Society* 5, no. 2: 2053951718786316. <https://doi.org/10.1177/2053951718786316>
- Hepp, Andreas, Juliane Jarke, and Leif Kramp. 2022. *New Perspectives in Critical Data Studies: The Ambivalences of Data Power*. Palgrave Macmillan.
- "HRDAG - Human Rights Data Analysis Group." *HRDAG - Human Rights Data Analysis Group* (blog), 2023. <http://hrdag.org/jep-cev-colombia/>
- Hunston, Susan. 2022. *Corpora in Applied Linguistics*. 2nd ed. Cambridge University Press. <https://doi.org/10.1017/9781108616218>
- Kitchin, Rob. 2022. *The Data Revolution: A Critical Analysis of Big Data, Open Data & Data Infrastructures*. 2nd ed. Los Angeles, CA: Sage Publications.
- Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1, no. 1: 205395171452848. <https://doi.org/10.1177/2053951714528481>

- Kitchin, Rob, and Tracey P. Lauriault. 2014. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." *The Programmable City Working Paper 2*.
- Mayaffre, Damon, and Céline Poudat. 2020. "Quantitative Approaches to Political Discourse: Corpus Linguistics and Text Statistics." In *Speaking of Europe: Approaches to Complexity in European Political Discourse*, edited by Kjersti Flottum: John Benjamins Publishing Company.
- Mejias, Ulises A., and Nick Couldry. 2019. "Datafication." *Internet Policy Review* 8, no. 4. <https://doi.org/10.14763/2019.4.1428>
- Merry, Sally Engle. 2016. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago Series in Law and Society. Chicago: The University of Chicago Press.
- Millar, Laura. 2017. *Archives: Principles and Practices*. Second edition. Principles and Practice in Records Management and Archives. London: Facet Publishing.
- Orozco, Guillermo. 1987. *Televisión y producción de significados* (tres ensayos). México: Universidad de Guadalajara y AMIC.
- Pickren, Graham. 2018. "'The Global Assemblage of Digital Flow': Critical Data Studies and the Infrastructures of Computing." *Progress in Human Geography* 42, no. 2: 225–43. <https://doi.org/10.1177/0309132516673241>
- Ramos-Martín, Juan. 2022. "Capitalismo informacional desde el Sur. Fetichismo de la información y materialidades de la colonialidad de los datos." *Revista Latinoamericana Estudios de la Paz y el Conflicto* 3, no. 6: 58–68. <https://doi.org/10.5377/rlpc.v3i6.14051>
- Ricaurte, Paola. 2019. "Data Epistemologies, The Coloniality of Power, and Resistance." *Television & New Media* 20, no. 4: 35065. <https://doi.org/10.1177/1527476419831640>
- Rodríguez Cano, César Augusto. 2022. *Hypermethods. Repertories of Social Research in Digital Environments*. Primera edición. Universidad Autónoma Metropolitana. <https://doi.org/10.24275/9786072824812>
- Rogers, Richard. 2023. *Métodos digitales*: ITESO.
- Rouvroy, Antoinette, y Thomas Berns. 2018. "Gubernamentalidad algorítmica y perspectivas de emancipación: ¿lo dispar como condición de individuación mediante la relación?" <http://repositorio.flacsoandes.edu.ec/handle/10469/15424>
- Selwyn, Neil. 2020. "Re-Imagining 'Learning Analytics' ... a Case for Starting Again?" *The Internet and Higher Education* 46 (julio 2020): 100745. <https://doi.org/10.1016/j.iheduc.2020.100745>

- Star, Susan Leigh. 2002. "Infrastructure and Ethnographic Practice: Working on the Fringes." *Scandinavian Journal of Information Systems*: Vol. 14, Núm. 2, artículo 6.14.
- Taylor, Diana, y Marcela Fuentes, eds. 2011. *Estudios avanzados de performance*. México: Fondo de Cultura Económica. https://www.icmujeres.gob.mx/wp-content/uploads/2020/05/Performance_Butler.pdf
- Thatcher, Jim, David O'Sullivan, y Dillon Mahmoudi. 2016. "Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data." *Environment and Planning D: Society and Space* 34, no. 6: 990–1006. <https://doi.org/10.1177/02637758166633195>
- Van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12, no. 2.
- Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation." *Regulation & Governance* 12, no. 4: 505–23. <https://doi.org/10.1111/rego.12158>