

# *Ciencia de datos para la historia: datificar las fuentes para una historia (predictiva)*



## *Data Science for History: Datafication of Sources for a (Predictive) History*

LAURA MANRIQUE-GÓMEZ

Universidad de los Andes

Colombia

Correo: l.manriqueg@uniandes.edu.co

<https://orcid.org/0000-0003-0843-8157>

JAIME HUMBERTO BORJA GÓMEZ.

Departamento de Historia, Universidad de los Andes

Colombia

Correo: jborja@uniandes.edu.co

<https://orcid.org/0000-0002-9909-8890>

DOI: 10.48102/hyg.vi64.541

Artículo recibido: 22/03/2024

Artículo aceptado: 18/10/2024

### RESUMEN

La historiografía científica entrelazó tres pilares: documento, dato y hecho. Estos conceptos se revalorizaron con las digitalizaciones masivas a finales del siglo XX, que redefinió el concepto de *dato* para la historia. Una de las consecuencias fue la pérdida de control de los historiadores sobre los datos, por lo que estos asumieron un doble papel: usuarios de datos digitales o sus curadores. Este cambio supone dos retos: repensar el tratamiento de los datos análogos y digitales; y descubrir cómo aplicar técnicas de la ciencia de datos para enriquecer la práctica histórica. Este artículo examina los cambios producidos por las relaciones entre historia

y ciencia de datos, enfocándose en las bases de datos como nuevas narrativas historiográficas. La datificación del pasado es una de las principales consecuencias de la ciencia de datos que, con modelos de inteligencia artificial, convierte en el centro del debate el carácter predictivo de la historia.

*Palabras clave:* Historia digital, datificación, ciencia de datos, predicción, datos, inteligencia artificial

#### ABSTRACT

Historiography has traditionally relied on three pillars: document, data, and facts. These concepts gained new value in the late 20th-century with mass digitization efforts, leading to a redefinition of the role of *data* in history. A significant outcome of this shift was historians losing control over data and facts, resulting in a bifurcation of their roles into users and curators of digital data. This transformation brings forth two critical challenges: the need to re-evaluate the handling of both analog and digital data, and the exploration of how data science techniques can enhance historical research. This article explores the interplay between history and data science, focusing on databases as emergent historiographical narratives. We argue that the datafication of history is a fundamental result of data science's integration into the field and with the advent of artificial intelligence, the predictive dimension of history is becoming the central point of disciplinary debate.

*Keywords:* History, Digital History, Datafication, Data Science, Data

#### INTRODUCCIÓN

La segunda mitad del siglo XX estuvo especialmente marcada por un crecimiento de las tecnologías electrónicas y digitales, cuyos efectos se sintieron en todas las esferas del orden social. Este proceso tecnológico cobró una vital importancia en las décadas posteriores a la finalización de la Segunda Guerra Mundial en el contexto político y cultural de la Guerra Fría. La televisión, por ejemplo, aunque había tenido importantes desarrollos en la década de los años 30, despegó en la década de los años 50, ini-

ciando su proceso de expansión global. En el mismo contexto, se consolidó la computación (mainframes), tras los primeros desarrollos en tiempos de guerra. El colofón de este proceso fue la popularización tanto de la computadora personal en los años 80, como el internet y la web en la década siguiente. Las ciencias sociales se han preguntado, en los últimos veinte años, acerca de este impacto en el ordenamiento social y cultural. Incluso, uno de sus resultados es la formación de disciplinas o tendencias sociales y humanas vinculadas con la tecnología, como es el caso de las humanidades digitales y la historia digital.<sup>1</sup> Sin embargo, en la última década ha cobrado mayor relevancia la pregunta por el cómo afectan estas tecnologías la idea de pasado y futuro,<sup>2</sup> y cómo afectan las formas de hacer historia.

La historiografía ha generado diferentes maneras de narrar el pasado, la cual depende en buena medida de las tecnologías y las materialidades desde donde se piensa ese pasado. En la era de la cultura digital, nuevos usos públicos se prefiguran para la historia<sup>3</sup> y la ciencia de datos se ha revelado como un catalizador transformador que sobrepasa la mera acumulación de conocimientos, para afectar intrínsecamente los métodos y usos del presente, y potencialmente del futuro. Esta emergencia de tecnologías dentro del discurso histórico, revelan que *el dato* ya no es un singular aislado, dato(s) connota masividad de información, lo cual está impactando de manera directa el quehacer de la historia. Solo es

<sup>1</sup> El debate y la investigación sobre la formación de las humanidades computacionales, las humanidades digitales y su impacto en las disciplinas sociales y humanas es abundante. Para una revisión de sus problemáticas centrales se puede ver a Olivier Le Deuff, *Digital Humanities. History and Development* (Londres: Wiley, 2018).

<sup>2</sup> Un ejemplo es el actual proyecto de investigación colectiva, que explora la relación historia y futuro. En cada número de la revista *History and Theory* se publica una contribución en la cual se destaca frecuentemente esta relación con los impactos de la tecnología. El proyecto está dirigido por Simon Zoltán Boldizsár y Marek Tamm desde 2021. *History and Theory* 60, no. 1 (2021): 3-22.

<sup>3</sup> Jo Guldi y David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014).

hasta la década de los 2010 cuando se destaca la preocupación disciplinar por el *dato*: los historiadores empiezan a mirar la *Data Science* como herramienta para tratar la datificación del pasado, y con ello la vinculación de un nuevo lenguaje en función histórica: Big Data, *Machine Learning*, *Deep Learning*, *Inteligencia Artificial*, *lectura distante*, *minería de datos*, entre otras. Más recientemente, se integra la DS como marco conceptual para entender cómo construimos ese pasado.

En este contexto, el presente artículo trata los cambios del término “dato” tanto en la práctica histórica como en las reflexiones historiográficas, para entender cómo y desde dónde se ha comenzado a integrar el discurso histórico en una disciplina relativamente nueva, pero ya dominante en todos los campos de la actividad humana, la ciencia de datos. Se trata de responder a la pregunta ¿De qué manera los efectos de la cultura digital y la ciencia de datos afectan la idea de dato histórico y cómo se ha dado el proceso de desarrollo de la “datificación del pasado”? Para responder a la pregunta, se esbozan las relaciones entre historia y tecnología con el fin de comprender el lugar del dato, que más allá de una genealogía del concepto, se propone como punto de tensión entre la cultura análoga y las perspectivas de la cultura digital. Sobre esta perspectiva, se trata el surgimiento de las bases de datos como forma novedosa de relato y argumentación histórica, campo donde se puede entender la datificación del pasado y las metodologías de la ciencia de datos que permiten ver nuevos patrones históricos.

Muchos otros problemas se desprenden de esta perspectiva, que involucran desde el uso de metodologías puntuales como el Big Data, las implicaciones de la inteligencia artificial, el carácter predictivo de la historia, la historia contrafactual e incluso la transformación de la idea de pasado en estas “tecnosociedades”. Aunque no son objetivo principal de este texto, los datos nos abren esos panoramas y una problemática sobre la cual este artículo contiene una reflexión *latente*: ¿Tenemos la suficiente alfabetización

digital como historiadores, y más concretamente, contamos con la “alfabetización en datos” para hablar con propiedad del pasado, ante esta “sociedad de la información”? Incluso, ¿Estamos siendo testigos de la reinención de la historia como una disciplina cuantitativa, al modo de una ciencia exacta?

## HISTORIOGRAFÍA Y CULTURA DIGITAL

Resultado de un proceso tecnológico que se había iniciado a finales de la década de 1960, internet y la web se hicieron públicos y se popularizaron en los primeros años de la década de los 1990. La revolución que generó este “giro digital”, alimentado por la globalización política y económica de aquella década, es innegable, aunque el proceso estaba avanzando años atrás. El impacto de estas tecnologías favoreció que en poco tiempo se acuñara el concepto de “cultura digital”.<sup>4</sup> Esto ocurría a finales de la década de los 90, lo cual obedecía a fenómenos que no se reducen al consumo de nuevas formas de comunicación, sino a la conformación de una nueva *cultura de la información*. Esta se vinculaba estrictamente a la proliferación de una materialidad, el dispositivo digital, y a un conjunto de herramientas que afectan la vida humana en su conjunto global. Para historiadores digitales como Hannu Salmi, la cultura digital se entiende como el epítome del mundo moderno contemporáneo, en cuyo proceso de implementación se debe observar la tensión entre la cultura analógica y lo “nacido digitalmente”, en lo cual se insertan los nuevos productos culturales.<sup>5</sup>

Por supuesto, la cultura digital impactó las ciencias sociales y humanas aunque no de manera uniforme. Desde las Humanidades Digitales se ha interpretado el proceso de este impacto, bajo la pro-

<sup>4</sup> Pierre Levy, *Cibercultura. La cultura de la sociedad digital* (Barcelona: Anthropos, 2007).

<sup>5</sup> Hannu Salmi, *What is Digital History?* (Cambridge: Wiley, 2021), 21.

puesta de tres oleadas de consolidación, de las cuales se desprenden problemáticas distintas.<sup>6</sup> Estas tres oleadas se pueden extender a la historia digital. La primera, desde mediados de los años 80, cubre el proceso de digitalización masiva que llevó a la inmaterialización de millones de documentos análogos. Le siguió la fiebre de las herramientas digitales en la primera década de los 2000, que buscaban mejorar las condiciones de trabajo del historiador. En este contexto surgió la historia digital, paralelamente a las Humanidades Digitales, como una síntesis de las Humanidades computacionales y la Cliometría de la segunda mitad del siglo XX.<sup>7</sup>

Estos procesos llevaron a que muchas disciplinas sociales debatieran las relaciones entre la tecnología y la sociedad. El impacto del debate se incrementó historiográficamente en la medida en que el problema ya no era digitalizar y generar herramientas específicas para la investigación o la comunicación. Desde mediados de la década de los 2000, la complejidad del desarrollo tecnológico digital, la creciente importancia del dato, los nuevos dispositivos que surgían con la popularización de la pantalla táctil, y la aparición de las redes sociales, permitieron que se consolidaran nuevas disciplinas. La ciencia de datos fue una de estas nuevas disciplinas, frente a la cual, actualmente no hay espacio cultural que le sea indiferente. En este contexto, luego de la etapa inicial de digitalización de documentos y la incorporación de herramientas digitales en el trabajo histórico, emergió una tercera etapa que se ubica alrededor del 2011. En esta tercera etapa la reflexión historiográfica comenzó a preocuparse por una versión distinta de lo histórico, en la medida en que se incorporaron al

<sup>6</sup> David Berry, “Digital Humanities: First, Second and Third Wave”, *Stunlaw philosophy and critique for a digital age*, (2011). <http://stunlaw.blogspot.com/2011/01/digital-humanities-first-second-and.html>

<sup>7</sup> Sobre estos procesos de conformación de la historia digital, se pueden ver dos introducciones: Ian Milligan, *The Transformation of Historical Research in the Digital Age* (Cambridge: Cambridge University Press, 2022), especialmente el capítulo “Libraries and Data Base”. También, Hannu Salmi, *What is Digital History?*

diálogo histórico, elementos metodológicos y teóricos provenientes de la ciencia de datos. Una de las primeras incorporaciones fue el Big Data, cuyas primeras reflexiones históricas emergieron hacia 2013.<sup>8</sup> Pocos años después, se encuentran discusiones historiográficas que evidencian que esta no es solo una práctica metodológica, sino también una propuesta teórica.<sup>9</sup>

Evidentemente la cultura digital y la vinculación de la tecnología al debate historiográfico y a la práctica histórica ha permeado solo ciertos sectores disciplinares. En este contexto, se pueden reconocer dos tendencias: una gran mayoría de historiadores que son usuarios pasivos de la cultura digital, que consumen recursos de la web y emplean formas de comunicación digital. Un segundo grupo, que usa intencionalmente metodologías avanzadas de tecnologías computacionales y digitales en sus investigaciones. Algunos de estos asumen la historia digital como campo de trabajo, de investigación o reflexión historiográfica. Frente a esta situación, algunos historiadores han establecido la diferencia entre hacer historia desde lo digital y la historia digital.<sup>10</sup> Esta ambigüedad acerca de la incorporación de la tecnología en la práctica histórica, vuelve a reflejar el problema que se extiende a otros campos de la cultura contemporánea: la tensión que genera el tránsito de una sociedad de formas análogas, a otra sociedad de formas digitales. Es decir, formas digitalizadas y virtualizadas que aún conservan su esencia

<sup>8</sup> Por ejemplo, el libro de Manning que introduce el trabajo web “Collaborative for Historical Information and Analysis” (CHIA), liderado desde la Universidad de Pittsburgh: Patrick Manning, *Big Data in History* (Basingstoke: Palgrave MacMillan, 2013).

<sup>9</sup> Shawn Graham, Ian Milligan y Scott Weingart. *Exploring big historical data. The Historian's Microscope* (Londres: Imperial College Press, 2016).

<sup>10</sup> Un buen número de historiadores digitales hace la observación acerca de cómo la gran mayoría de los historiadores emplean elementos de la cultura digital para sus trabajos, ya sea consultas de archivos, fotografía digital, redes, publicaciones en la web, etcétera. Esto es hacer historia “desde lo digital”, a diferencia de quienes especializan el trabajo con herramientas digitales y en perspectiva de la historia digital. Respecto al uso de la fotografía digital en redes, véase el ejemplo en Milligan, *The Transformation of Historical Research*, 34-35.

en el mundo de lo análogo. El problema no es reciente, desde el surgimiento de la computación a mediados del siglo XX, algunos pocos humanistas recurrieron a estas nuevas tecnologías, como fue el caso de Frank Owsleys y Robert Busa en 1949.<sup>11</sup> Desde entonces, y en un proceso gradual hasta la década de los años 1990, algunos historiadores emplearon computadoras en sus investigaciones, especialmente relacionados con la cliometría. Estos usos espontáneos se transformaron en uso intensivo cuando las tecnologías digitales se convirtieron en el eje fundamental de los sistemas de información, así como espacio de reflexión.

En este contexto, vale la pena retomar la reflexión de Stefan Tanaka a partir de una cita de Elizabeth Eisenstein, quien afirma: “*Mi hipótesis de trabajo es que todas las visiones de la historia han sido moldeadas fundamentalmente por la forma en que se duplican los registros, se transmiten los conocimientos y se almacena y recupera la información*” [traducción de los autores].<sup>12</sup> La aparición de la imprenta en el siglo XV implicó nuevos modos de interpretación e información sobre la cultura de la oralidad, o la prensa del siglo XIX generó una nueva forma de recordar e informarse. Todas ellas afectaron la percepción del pasado y la función de la historia. Así se puede entender entonces la tensión entre lo analógico y lo digital en relación con la historia: un proceso mediante el cual la tecnología genera una materialidad comunicativa y unas formas diferentes de conocimiento, escritura y aprendizaje. Esta tensión comienza a dar lugar a nuevas formas de relación con la idea de pasado y el cómo se construye, se reconstruye o se interpreta ese pasado: la percepción del pasado cambia en relación con las tec-

<sup>11</sup> Adam Crymble, *Technology and the Historian. Transformation in the Digital Age* (Champaign: University of Illinois Press, 2021), 17. También, Steven E. Jones, Roberto Busa, S. J. and the Emergence of Humanities Computing. *The Priest and the Punched Cards* (Nueva York: Routledge, 2016).

<sup>12</sup> “*My working hypothesis is that all views of history have been fundamentally shaped by the way records are duplicated, knowledge transmitted, and information stored and retrieved*”. Stefan Tanaka, “The Old and New Of Digital History”, *History and Theory* 61, no. 4 (diciembre 2022), 10.



nologías de conocimiento. La actual penumbra, o el tránsito entre lo análogo y lo digital, es un proceso que está permeando el discurso histórico incluso desde lo que se recientemente se denomina “pensamiento computacional”.<sup>13</sup> La pregunta es entonces, ¿Qué tipo de pasado o formas históricas está provocando la cultura digital?

La pregunta nos vuelve a llevar al problema con el que comenzamos este apartado: el proceso tecnológico de los últimos 70 años marchó de manera paralela a las ciencias sociales y particularmente a la historia, pero solamente en los últimos 20 años han comenzado a cruzarse sus caminos. El resultado se ha dado en dos campos diferenciados: por un lado, el creciente uso de la tecnología en función de la investigación y la comunicación del pasado, es decir, en la práctica histórica; segundo, el lugar que ocupa la tecnología digital en la reflexión historiográfica. Por su-

<sup>13</sup> Ian Milligan propone que es necesario que los historiadores incorporen en su trabajo herramientas digitales y adquieran habilidades algorítmicas que le permitan comprender cómo opera el código, cómo son creados y almacenados los objetos digitales, y las dimensiones humanas detrás de los lenguajes de programación: “*This does not mean that all historians need to become programmers (...) But it does mean that they need to think algorithmically: to think about how code might operate, how digital objects are created and stored, and to realize the human dimensions behind programming languages*”. Ian Milligan, *History in the Age of Abundance How the Web Is Transforming Historical Research* (Montreal: McGill-Queen's University Press, 2019), 241. El pensamiento computacional, sin embargo, no es únicamente algorítmico. Otras destrezas como los procesos de pensamiento asociados con la abstracción, la recursión, la automatización, la simulación, el pensamiento paralelo y el pensamiento en función de los datos son parte del pensamiento computacional. Para el propósito de este artículo, abordamos el pensamiento computacional esencialmente en función de los datos o el *Data Thinking*. Véase Jeannette Wing, “Computational thinking and thinking about computing”, *Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences* 366 (2008): 3717–25; Peter Denning, J. y Matti Tedre. *Computational Thinking* (Cambridge: The Massachusetts Institute of Technology Press, 2019); Keith Kirkpatrick, “Parallel Computational Thinking”, *Communications of the ACM*, 60, no. 12 (2017): 17–19; Koby Mike et al., Computational Thinking in the Era of Data Science, *Communications of the ACM*, 65, no. 8 (2022): 33–35.

puesto, en este proceso de intercambio, hay elementos que son comunes en los dos campos pero con diferente percepción. Uno de estos elementos es la idea del *dato*, el cual se ha convertido en eje reflexivo que articula tecnología e historia. A pesar de que en los dos campos ha tenido una evolución particular, el *dato digital* está propiciando un enfoque historiográfico diferente. Entonces, se puede precisar la pregunta anterior de la siguiente manera: ¿En qué modo el concepto o idea de *dato* y las formas como se le tratan (apartado 2), está generando (o puede generar) una percepción o un tratamiento distinto del pasado (numeral 3 y 4)?

## EL DATO, ENTRE LA TECNOLOGÍA Y LA HISTORIA

En la segunda mitad del siglo XX, el desarrollo de los lenguajes algorítmicos y computacionales generó una transformación del concepto de *dato*, al convertirlo en una representación simbólica, numérica o alfabética de la naturaleza (y la realidad), pero que no es la naturaleza misma.<sup>14</sup> La era digital, particularmente a partir de los años noventa, acrecentó el lugar del dato en la sociedad, impactando directamente en la historiografía. La historia digital, aunque inicialmente enfocada en el desarrollo y uso de herramientas tecnológicas, ha evolucionado hacia una preocupación más profunda del dato en sí mismo. Pero el proceso para llegar a esta apropiación fue largo. En este apartado se busca problema-

<sup>14</sup> Una definición oficial la proporciona la Economic Commission for Europe of the United Nations (UNECE), “Terminology on Statistical Metadata”, en *Conference of European Statisticians Statistical Standards and Studies*, no. 53 (Ginebra, 2000), 6: “Data is the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means”. Véase también Organization for Economic Co-operation and Development (OECD), *OECD Glossary of Statistical Terms* (París: OECD Publishing, 2008). Sobre su impacto en las ciencias sociales, véase Lisa Gitelman. *Raw Data is an Oxymoron* (Cambridge: Massachusetts Institute of Technology Press, 2013).

tizar cómo en el *dato* se intersecta actualmente entre la historia y la tecnología. Una primera pista para esta comprensión la ofrece la pregunta qué es el *dato* en la historiografía “científica” y factual. Eduard Carr en su conocida obra introductoria a la historia, afirma: “Los datos, hayan sido encontrados en documentos o no, tienen que ser elaborados por el historiador antes de que él pueda hacer algún uso de ellos; y el uso que hace de ellos es precisamente un proceso de elaboración”.<sup>15</sup> Esta aseveración sirve para introducir el problema central: La historia científica que surgió en el siglo XIX, instauró una lógica de investigación disciplinar que entrelazaba tres elementos: documento, dato y hecho,<sup>16</sup> evidenciado en el texto de Carr.

Estos tres elementos marcan la estructura del hacer historia: el documento, que materializa el pasado, es la “mina” que proporciona *datos*, y a partir de estos, se construyen los *hechos*. Esto es en esencia la práctica histórica, y lo que media la relación documento-dato-hecho, es la aplicación de los métodos, que aportan el carácter “científico” a la disciplina. A partir de la década de 1970, los cambios de paradigma y la ruptura de fronteras que afectaron las distintas disciplinas de las ciencias sociales le aportaron a la historia transformaciones significativas, especialmente en lo temático y lo discursivo, pero mantuvieron prácticamente en la misma condición analítica la triada documento-dato-hecho. Sin embargo, se evidenciaron dos cambios importantes: se amplió el concepto de “documento histórico” a “fuente histórica”,<sup>17</sup> lo que implicó un cambio esencial en el concepto de archivo y la inclu-

<sup>15</sup> Eduard Carr, ¿Qué es la historia? (Barcelona: Ariel, 1984), 22.

<sup>16</sup> Jesse Torgerson, “Historical Practice in the Era of Digital History”, *History and Theory* 61, no. 4 (diciembre 2022): 38.

<sup>17</sup> Este cambio implica el uso de documentos en el sentido estricto del término tal como lo introdujeron los historicismos del siglo XIX, y de la extensión de la investigación hacia fuentes que hasta aquel momento eran “propiedad” de otras disciplinas sociales y humanas, como la imagen, la literatura o la oralidad. Esto es lo que se denomina “la revolución del documento”, a partir de Jacques Le Goff, *El orden de la memoria* (Barcelona: Paidós, [1977] 1991), 230-40.

sión de un sinnúmero de posibilidades en el uso de fuentes, lo que abrió la historia a nuevas metodologías y temas. El segundo cambio, es la ampliación del concepto “hecho” por “acontecimiento”, lo que implicó nuevas posibilidades interpretativas.

En otras palabras, *documento* y *hecho* se vieron alterados por los cambios de la historiografía, lo que implicó otras formas en la comprensión del pasado y la construcción del discurso histórico. Sin embargo, la idea de “dato” pasó relativamente desapercibida hasta la década de los años noventa cuando se desencadenaron las transformaciones de la cultura digital. Sin embargo, esta transformación no afectó inmediatamente la estructura metodológica del discurso histórico, pero sí evidenció varios fenómenos culturales cuyas repercusiones se reflejarían en el concepto de *dato* en las siguientes décadas. Entre ellos se encuentra la consolidación de la denominada “era de la información”; la digitalización de “fuentes primarias”, especialmente prensa, libros antiguos y documentos de archivos. La aparición de las redes sociales digitales y la expansión de la web 2.0 a mediados de la década de los 2000, complejizó el problema al incluir el *dato nacido digitalmente*<sup>18</sup> y el reto de cómo tratar los datos digitales. La respuesta de la disciplina histórica a esta nueva cuestión fue lenta. El surgimiento de la historia digital en aquella década no dio respuesta inmediata al problema del dato, sino que se preocupó más por el desarrollo, uso y tratamiento de las herramientas. Para finales de aquella década, se había consolidado una nueva disciplina que estudiaba los datos y su tratamiento, la *Data Science*. El concepto de dato permite intersecar dos campos aparentemente incompatibles, el

<sup>18</sup> El dato nacido digitalmente, básicamente compone la web. Solo a mediados de la década de 2010, se generó la preocupación por pensar la historia que allí se encuentra desde comienzos de la década de 1990. Desde entonces hay una creciente preocupación por esto que se está comenzando a denominar historiografía web; véase Niels Brügger e Ian Milligan, *The Sage Handbook of Web History* (Londres: Sage Publications, 2018). Sobre el dato nacido digital, véase la introducción de Milligan, *History in the Age of Abundance*, 3-28.

campo de las tecnologías digitales y el campo de la historia. Tal como se muestra en la Figura 1, el diagrama de Venn ilustra la conexión entre la tecnología y la historia, dos campos disciplinares diferentes que se nutren recíprocamente en lo referente a los datos. Desde la ciencia de datos, la historia apropia sus usos, y de manera recíproca, el ejercicio de la ciencia de datos apropia la reflexión histórica crítica en torno a los datos.

En síntesis, el dato se considera una representación simbólica de la realidad, ya sea numérica o alfabética, fundamental para las ciencias sociales y la historiografía. En la era digital, el dato ha cobrado una relevancia inusitada, propiciando un enfoque historiográfico distinto al permitir nuevas formas de interpretar y comprender el pasado. Así, el ‘dato’ en digital no solo sirve como evidencia de fenómenos observados, sino que en su tratamiento se encuentra el potencial para abstraer realidades complejas bajo nuevas perspectivas de análisis.

En los últimos diez años, los historiadores que se han aproximado a la ciencia de datos reconocen que el dato se ha transformado,

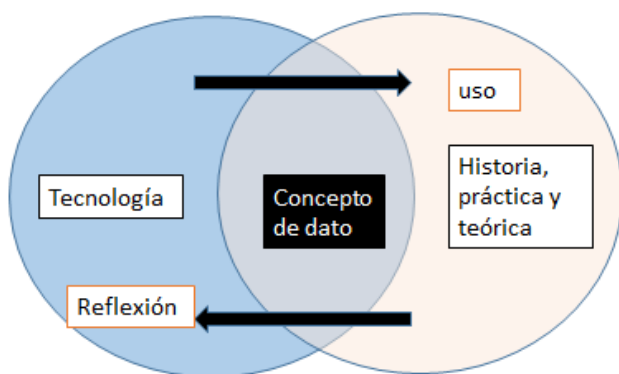


FIGURA 1  
El concepto del dato como intersección de campos.  
[Diseño original de los autores]

trayendo consigo nuevas percepciones y metodologías que accidentalmente chocan con los métodos tradicionales de las ciencias sociales. Historiográficamente esta situación presenta dos retos: la pérdida de control del historiador sobre la producción de datos; y la necesidad de adaptar la práctica histórica a las condiciones tecnológicas avanzadas donde el volumen y tratamiento de datos cambia radicalmente. El primer problema-reto, está relacionado con la crisis que introdujo en las ciencias sociales y la historia en particular, la recepción de la idea de dato que se formó en las tecnologías digitales. Desde el siglo XIX el historiador producía los datos para construir los hechos, y los historiadores recurrían a los datos de otros historiadores para la “construcción de conocimiento”.<sup>19</sup> El dato ya no lo produce el historiador, cualquier persona o institución con un dispositivo está produciendo datos digitales, así como hay datos en un documento colonial, los hay también en un meme, en cualquier *like* de redes sociales o en una transacción bancaria en línea. Todo esto es “fuente primaria” histórica.

Esta situación podría ser interpretada como una especie de banalización o pérdida del sentido de conocimiento que implica el *dato*, pero en realidad responden al segundo reto, la adaptación de la práctica histórica a una sociedad tecnologizada con un volumen y un tratamiento distinto del dato. Desde mediados de la década de los 2000, las humanidades y la historia digital, asumieron la tarea de poner en práctica los elementos que surgían de la teoría del dato computacional. En una primera instancia, adquirió importancia el uso del *metadato*<sup>20</sup> aplicado a la investigación social, y a comienzos de 2010, se implementaron metodologías para tratar los datos, como el Big Data, la lectura distante, la minería de datos,

<sup>19</sup> Torgerson. “Historical Practice in the Era of Digital History”, 42.

<sup>20</sup> Julia Flanders y Fotis Jannidis, *The Shape of Data in the Digital Humanities* (Nueva York: Routledge, 2019), 30-36. Véase también Johanna Drucker, *The Digital Humanities Coursebook. An Introduction to Digital Methods for Research and Scholarship* (Nueva York: Routledge, 2021), 52-56.

la minería de textos, por mencionar algunas.<sup>21</sup> Estas aplicaciones a la práctica histórica han crecido sistemáticamente en la última década, pero no tanto la reflexión historiográfica sobre el uso del dato.<sup>22</sup> La pregunta que se hacen Robertson y Mullen a propósito de la intervención de uno de los visionarios de la historia digital, es pertinente en este contexto: “Roy Rosenzweig reconoció hace casi veinte años que la era digital perturbaría la relación entre los historiadores y el trabajo de archivo y nos obligaría a preguntar: “¿Debería el trabajo de recopilar, organizar, editar y preservar fuentes primarias recibir el mismo tipo de reconocimiento y respeto que recibió inicialmente en la profesión?” [traducción de los autores].<sup>23</sup> En el centro de esta cuestión sobre la fuente, se encuentra la naturaleza del dato de la era digital.

La naturaleza del dato digital tiene dos niveles de comprensión: la primera es la condición técnica en la que se forma, el *dato nacido digitalmente* tiene un impacto muy distinto para lo histórico al compararlo con el *dato digitalizado*. Este último alberga un tipo de materialidad que lo respalda, mientras que el primero es básicamente inmaterial. A esta condición técnica se agrega el segundo nivel, el *dato renacido digitalmente*, es decir, el tratamiento técnico del dato sobre el dato, lo que pone en evidencia los problemas relacionados con su originalidad. En las situaciones tecnológicas actuales, es poco probable la verificación de la autenticidad del

<sup>21</sup> Sobre algunas metodologías digitales, véanse Shalin Hai-Jew (ed.), *Data Analytics in Digital Humanities* (Cham: Springer, 2017); (Waltham: Morgan Kaufmann, 1999); Drucker, *The Digital Humanities Coursebook*.

<sup>22</sup> Por ejemplo, Fred Gibbs y Trevor Owens, “The Hermeneutics of Data and Historical Writing”, en *Writing History in the Digital Age*, eds. Jack Dougherty y Kristen Nawrotzki (Ann Arbor, The University of Michigan Press, 2018), 159-70.

<sup>23</sup> “Roy Rosenzweig recognized almost twenty years ago that the digital era would unsettle the relationship between historians and archival work and force us to ask, “Should the work of collecting, organizing, editing and preserving of primary sources receive the same kind of recognition and respect that it did in the earlier days of the profession?” Stephen Robertson y Lincoln A. Mullen, “Arguing with Digital History: Patterns of Historical Interpretation”, *Journal of Social History* 54, no. 4 (verano 2021): 1008.

dato y más si estos son nacidos digitalmente. La pregunta latente es acerca del tipo de pasado que se construye con los diferentes tipos de datos que han sido procesados por algún medio digital.

De cualquier manera, los historiadores siguen empleando datos en su práctica histórica asimilando indistintamente el dato análogo y el digital. Este último ocasionalmente tratado con las metodologías propiamente digitales, y por lo general, de modos tradicionales. En otras palabras, conviven tres formas de relacionarse con el dato: como artefacto u objeto manufacturado, como una producción textual o como datos procesados por máquina.<sup>24</sup> Esta condición revela la ya mencionada tensión entre las formas de hacer historia análoga y la absorción de modelos y prácticas digitales. Las formas actuales de acercarse al dato tienen un elemento en común: ponen de presente la excesiva confianza que han depositado las sociedades tecnológicas en su inefabilidad, lo que representa un regreso a cierta ingenuidad “positiva” que considera que el dato cuenta por sí solo, lo que “está ahí”, una especie de materiales absolutos cuyo tratamiento algorítmico devela la realidad. Esta cuestión ha permitido poner en el debate de los humanistas e historiadores digitales la reflexión acerca de la condición humana del dato, es decir, la evidente consideración que son resultados culturales marcados por los espacios en los que son producidos, una interpretación de mundo.<sup>25</sup>

Estas reflexiones se produjeron a comienzos de la década de 2010, las cuales apuntaban a establecer esta naturaleza humanizada del dato. Principalmente tres artículos académicos sentaron directrices: Drucker en un artículo bastante citado, pretende mostrar cómo las visualizaciones no deben limitarse a representar datos, los cuales deben ser entendidos en su perspectiva etimológica: el dato como lo “dado” es diferente al dato como “capta”,

<sup>24</sup> Trevor Owens, “Defining Data for Humanists: Text, Artifact, Information or Evidence?”, *Journal of the Digital Humanities* 1, no. 1 (invierno 2011).

<sup>25</sup> Antonio Rojas, “Big data en las humanidades digitales. Nuevas conversaciones en el contexto académico global”, *Anuario acle de cultura digital* (2017), 72.



lo que implica entenderlos como una forma de conocimiento situado que sirve para organizar.<sup>26</sup> Por su parte, Owen establece un puente entre el dato informático y las humanidades a través de cuatro elementos de interpretación: texto, objeto, procesamiento y evidencia. Finalmente, Schöch revisa y clasifica forma de datos en relación con la complejidad, tamaño, limpieza y organización.<sup>27</sup> Estos tres ejemplos de conceptualización del dato digital para las humanidades, tienen un elemento determinante para la práctica histórica: el dato procesado, y el dato en cadenas masivas, contiene información “invisible” a los métodos tradicionales, lo cual implica un reto nuevo: la “alfabetización en datos”. Si dentro de la tradición historiográfica, como hemos visto, los datos son contruidos para elaborar hechos, en la cultura digital solo los datos estructurados se convierten en información.<sup>28</sup>

Aunque a menudo se emplean como sinónimos, “datos” e “información” poseen significados distintos. Los datos son referencias crudas y no interpretadas, mientras que la información es el resultado del procesamiento y análisis de esos datos, ofreciendo comprensión y contexto. En la historiografía, convertir datos en información implica interpretar cifras, fechas o eventos para reconstruir narrativas significativas del pasado, abriendo así nuevas dimensiones de análisis y enfoques historiográficos que han evolucionado con las tecnologías más recientes. En una tercera fase, después de que los datos son convertidos en información, el historiador debe asignarle un “significado” a esa información.

<sup>26</sup> Johanna Drucker “Humanities Approaches to Graphical Display” *Digital Humanities Quarterly* 5, no. 1 (2011), comentado por Jesse Torgerson, “Historical Practice in the Era of Digital History”, 47; Tanaka, “The Old And New Of Digital History”, 7; Susanna Alles Torrent, “Sobre la complejidad de los datos en Humanidades, o cómo traducir las ideas a datos”, *Revista de Humanidades Digitales* no. 4 (2019): 16.

<sup>27</sup> Christof Schöch, “Big? Smart? Clean? Messy? Data in the Humanities”, *Journal of Digital Humanities* 2, no., 3 (verano 2013).

<sup>28</sup> Sobre la importancia del dato estructurado en los métodos, véase Ian Foster et al., *Big Data and Social Science. Data Science Methods and Tools for Research and Practice* (Londres: CRC Press Taylor & Francis Group, 2021), 194-97.

La información se refiere a datos procesados y estructurados que ofrecen respuestas a preguntas específicas, mientras que el significado reside en el contexto y la interpretación que una cultura o un individuo atribuyen a esta información. En el ámbito historiográfico, asignar significado implica conectar eventos históricos con su impacto cultural, social o político, permitiendo una comprensión más profunda y enriquecedora del pasado que trasciende los simples registros históricos.

La importancia del dato digital, entonces, ha dado forma a lo que se conoce como la “era de la información”,<sup>29</sup> un concepto que define un giro dentro de la cultura digital. La intensificación de la digitalización de “fuentes” desde finales de la década de los 80 implicó una mayor circulación de datos y la exigencia de nuevas metodologías para tratarlos. Este proceso, acelerado con la popularización de internet pone de manifiesto para las siguientes décadas lo que Milligan ha llamado “la edad de la abundancia”<sup>30</sup> o también el “archivo infinito”. Entonces, el problema que se evidencia en este contexto es cómo se ha acercado la práctica histórica y la historiografía a la edad de la información en donde el reto ya no es hallar información, sino organizarla y procesarla. La alfabetización en datos, entendida como la formación de habilidades que permiten leer, comprender y comunicarlos, es un concepto relativamente reciente.<sup>31</sup> Este proceso no es solamente mecánico, tiene otras implicaciones no solo en la práctica histórica, sino también en la percepción del pasado. Esto sugiere una nueva pregunta: ¿Cómo afectan el tratamiento del pasado las perspectivas digitales del dato y las nuevas metodologías?

<sup>29</sup> Sobre la teoría de las sociedades de la información, véanse David Bell, *Cyberculture Theorists: Manuel Castells and Donna Haraway* (Nueva York: Routledge, 2006); Manuel Castells, *Change. The Impact of the Internet on Society: A Global Perspective* (Open Mind BBVA, 2014).

<sup>30</sup> Milligan, *History in the Age of Abundance*.

<sup>31</sup> Giovanni Schiuma y Daniela Carlucci, *Big Data in the Arts and Humanities: Theory and Practice* (Boca Ratón: CRC Press Taylor & Francis Group, 2018), 69-70.

Desde la década de los años 90 se ha teorizado sobre las bases de datos como lenguaje. “Después de que la novela y, más tarde, el cine, privilegiaran la narración como principal forma de expresión cultural de la era moderna, la era del ordenador introduce su correlato, que es la base de datos”,<sup>32</sup> dice Lev Manovich. Estas se incrementaron con la mencionada digitalización de fuentes en aquella época,<sup>33</sup> pero no creció al mismo al mismo ritmo la capacidad de leerlas con precisión.<sup>34</sup> Estas colecciones de archivos digitalizados abrieron nuevas posibilidades para proyectos de investigación como la iniciativa TEI y nuevas técnicas de comunicación.<sup>35</sup> Las bibliotecas y colecciones globales se ponían al alcance de cientos de interesados sin restricciones geográficas y temporales. En los años 2000, asentadas las bases de datos, fueron el punto de enlace entre los archivos análogos y su versión digital, pero con las ventajas que ofrecía la virtualización: podían reunir diversos archivos, permitían acelerar los procesos de búsqueda, generaban comodidades de visualización, etc. En este sentido, respondían a una nueva forma de entender el archivo, porque marcaban formas distintas de leer, comprender y contextualizar las fuentes. Las bases de datos reemplazaban simbólicamente al archivo como lugar físico. Pero no solo acumulaban datos, contenían un relato y formas de argumentación en la medida en que seleccionaban lo digitalizado, algunas voces estaban presentes, otras ausentes.

<sup>32</sup> Lev Manovich, *El lenguaje de los nuevos medios de comunicación. La imagen en la era digital* (Barcelona: Paidós, 2005), 283.

<sup>33</sup> David Berry y Fagerjord Anders. *Digital Humanities: Knowledge and Critique in a Digital Age* (Polity Press, 2017) 7; Dobson, James, “Interpretable Outputs: Criteria for Machine Learning in the Humanities”, *Digital Humanities Quarterly* 15, no. 2 (2021).

<sup>34</sup> Milligan, *The Transformation of Historical Research*, 27.

<sup>35</sup> Susan Schreibman, Ray Siemens y John Unsworth, eds., *A New Companion to Digital Humanities* (Chichester: Wiley Blackwell, 2016); Graham, Milligan y Weingart, *Exploring big historical data*.

Las bases de datos digitales se consolidaron como formas nuevas de conocimiento, como por ejemplo, la posibilidad de crear archivos digitales multimodales, es decir, aquellos acervos compuestos de textos e imágenes, lo cual generó gran expectativa. En 2007 se abrió la controversia cuando, en Estados Unidos, la Modern Language Association dedicó un número de su revista enfrentando los argumentos narrativos tradicionales y las estructuras de las bases de datos.<sup>36</sup> Estos debates generaron la posibilidad de entender cómo las bases de datos tenían un relato no lineal, una de las características de lo digital, de modo que tenían potencialmente muchas lecturas o enfoques.<sup>37</sup> Además, el uso de *datos estructurados* multiplicaba la información. Pese al debate, que incluyó posiciones más conservadoras, se le comenzó a considerar una forma de argumentación histórica, cuya cuestión parte del principio que “los datos son un artefacto o un texto que puede tener el mismo valor probatorio potencial que cualquier otro tipo de artefacto”.<sup>38</sup> Desde esta perspectiva, las bases de datos no solo son contenedores de información, hechos y cifras, sino que la interacción entre sus diversos componentes se comporta como argumentos. Incluso, el formato de la interfaz hace parte del relato argumental, que junto con las formas de visualización, conforman un conjunto narrativo.<sup>39</sup>

Historiográficamente, el que se considera el primer trabajo histórico que se hizo en perspectiva digital y que fue incorporado a la naciente web en la temprana fecha de 1993, *The Valley of the Shadow*<sup>40</sup> ya tenía consideraciones narrativas particulares

<sup>36</sup> PMLA: *Publications of the Language Association of America*, 2007 -5, Vol. 122; Drucker, *The Digital Humanities Coursebook*, 79-80.

<sup>37</sup> Véase, por ejemplo, el artículo de Katherine Hayles, “Narrative and Database: Natural Symbionts”, *PMLA: Publications of the Language Association of America, Special Topic: Remapping Genre* 122, no. 5, (octubre 2007): 1603-8.

<sup>38</sup> Stephen Robertson, “The Properties Of Digital History”, *History and Theory* 61, no. 4 (diciembre 2022): 92.

<sup>39</sup> Robertson y Mullen, “Arguing with Digital History”, 1011.

<sup>40</sup> El sitio original aún en servicio se encuentra en <https://valley.lib.virginia.edu/>,

como base de datos. La complejidad de las fuentes, entre las que se encontraban cartas, diarios, memorias, registros de censos, de la iglesia y del gobierno, así como informes de batalla, discursos y periódicos, hizo evidente para Thomas y Ayers, los historiadores líderes de este proyecto, que la tecnología de las computadoras podría ayudarles en su empeño. Acumular toda esta información y vincularla de manera básica, tenía el propósito de “*dar voz a cientos de personas individuales, para contar historias olvidadas de la era de la Guerra Civil*” tal como indica la introducción del portal web original. La versión web actual precisa el direccionamiento argumentativo de la colección digital, planteando que: “*Las dos comunidades, una en el Norte y otra en el Sur, experimentaron todos los desafíos nacionales, desde la secesión hasta la Reconstrucción*”. Este primer trabajo muestra cómo se avizoraba el futuro de las bases de datos como narración.

Otro ejemplo de este tipo de proyectos de bases de datos que se ha entendido como un trabajo historiográfico de amplio alcance y continuado a través de varias décadas, es el *The Proceedings of the Old Bailey*<sup>41</sup> que desde el año 2000 tiene en línea 197,000 juicios entre 1674 y 1913. El proyecto surgió como respuesta a los planes de digitalización de archivos históricos con financiación pública puestos en marcha en los años 90 en Gran Bretaña,

---

pero redirecciona al nuevo portal <https://valley.newamericanhistory.org/> 14 de octubre 2024. Este fue realizado por los historiadores William G. Thomas III y Edward L. Ayers, especialistas en el siglo XIX, quienes se interesaron en recoger experiencias de soldados confederados durante la guerra civil en dos condados cercanos a Nueva York a través de fuentes regionales. Ayers y Thomas pretendían hacer un trabajo de historia comparada entre el sur y el norte, mostrando detalles de la vida cotidiana para analizar el comportamiento de su gente frente al conflicto.

<sup>41</sup> Véase *The Proceedings of the Old Bailey Project* <https://www.oldbaileyonline.org/> consultada el 4 de octubre de 2024). Esta base de datos tiene a su vez un proyecto derivado y especializado, *London Lives*, que se concentra en sectores menos favorecidos del siglo XVIII. Reúne información de 8 archivos diferentes. Véase *London Lives Project. 1690 to 1800 Crime, Poverty and Social Policy in the Metropolis*. <https://www.londonlives.org/index.jsp> (consultada el 14 de octubre de 2024).

y buscaba asegurar que no se perdieran las fuentes de las clases trabajadoras que permitirían posteriormente hacer una historia social desde abajo. Tim Hitchcock y Robert Shoemaker, los historiadores que originaron el proyecto de *Old Bailey*, rápidamente fueron conscientes de que la colección debía estar guiada por un interés narrativo particular. Para Hitchcock, la cuestión era cómo obtener evidencias del comportamiento cotidiano que contenían las fuentes coleccionadas; mientras que para Shoemaker el interés radicaba en la posibilidad de crear un conjunto de datos sobre delitos, castigos y procedimientos judiciales. Ambas perspectivas han guiado el desarrollo de la colección durante las dos décadas posteriores, en particular, en el detalle de las decisiones sobre qué tipo de documentos digitalizar y qué temas o entidades particulares resaltar en cada documento mediante marcaciones XML-TEI, para búsquedas web posteriores.

Tanto el proyecto de *Old Bailey* como *Valley of the Shadow*, muestran la construcción de una base de datos en diferentes etapas, que reúne materiales de diferentes procedencias y contexturas. Este surgimiento de bases digitales históricas en el campo de la web, evidencia de qué manera se dio un lento proceso dentro de la historia como disciplina para tratar los datos en las nuevas condiciones tecnológicas. Algunas más visuales o mejor adaptadas al ámbito digital que otras,<sup>42</sup> lo cierto es que durante la década de los 2000 se consolidaba una historiografía digital, que se narraba desde las bases de datos.

<sup>42</sup> No es intención de este artículo evidenciar una historiografía de las bases de datos, pero se pueden mencionar algunos casos para evidenciar el proceso. El archivo censal medieval conocido como *Domesday*, produjo al menos tres formas de bases de datos diferentes entre sí. Uno de sus ejemplos se observa en PASE Project. The Prosopography of Anglo-Saxon England <https://pase.ac.uk/> (consultado el 14 de octubre 2024), y el análisis de este proceso en: Laura K. Morreale, "History As Antidote: The Argument For Documentation In Digital History", *History and Theory* 61, no. 4 (diciembre 2022): 65. Otros ejemplos se encuentran en The Portal to Texas History Project <https://texashistory.unt.edu/> (consultado el 14 de octubre de 2024).

En las siguientes dos décadas se refinaron tecnológica y temáticamente las bases de datos. Surgieron vinculadas a instituciones y resultado de los procesos de digitalización, como sucedió con las colecciones de museos, bibliotecas o archivos, instituciones que visibilizaban de mejor manera sus contenidos. Aunque muchos de estos proyectos no eran estrictamente históricos, aportaban fuentes digitalizadas para la investigación. Otras bases de datos recogían temáticamente lo dispuesto en muchos archivos, concentrando miles de documentos en un solo lugar y permitiendo el acceso con ciertas ventajas tecnológicas básicas: por ejemplo, buscadores y filtros. Pero más allá de intentar esbozar aquí el proceso de consolidación de nuevas formas de narrar la historia a través de datos o colecciones de datos, es importante marcar el cambio que comienza a gestarse a mediados de la década de los 2010.

El paso de la base de datos como “colección” interactiva, a un relato más complejo y con mayor ecosistema digital, se gestó precisamente en la década del 2010, donde las bases de datos intensificaron el uso de herramientas más complejas tecnológicamente, y algunas procedieron con dispositivos de ciencia de datos. Con estos avances, las bases de datos se consolidaron como formas de narrativa, a través de métodos de lecturas distantes y visualización de datos, habilitando la posibilidad de obtener información panorámica de las grandes colecciones.<sup>43</sup> Pero para llegar a este punto, fue necesario un proceso de asimilación del pensamiento computacional en función de los datos o el *data thinking*, para culminar en lo que se ha denominado “datificación”.<sup>44</sup> El dato

<sup>43</sup> Stéphan Sinclair y Geoffrey Rockwell, “Text Analysis and Visualization: Making Meaning Count”, en *A New Companion to Digital Humanities*, eds. Susan Schreibman, R. Siemens y J. Unsworth (Chichester: Wiley Blackwell, 2016): 280; Underwood, Ted, “How not to do things with words”, *Blog de Ted Underwood, Humanista digital* 25 de agosto de 2012. <https://tedunderwood.com/2012/08/25/how-not-to-do-things-with-words/>

<sup>44</sup> Viktor Mayer-Schönberger y Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston: Houghton Mifflin Harcourt, 2013), cap. 5, EPUB.

se refiere a la descripción de algo que puede ser “grabado, analizado y reordenado”,<sup>45</sup> de tal manera que la datificación implica un proceso de transformación de un fenómeno complejo, histórico en este caso, para representarlo de manera tabular. Esta representación permite preservar las características de relacionamiento definidas por quien diseña la tabla, para que los datos puedan ser sujetos de análisis. Finalmente, los datos deben ser transformados en representaciones binarias para garantizar su almacenamiento y procesamiento computacional.

La datificación en tanto que proceso, parte del principio de que los datos son una evidencia que respalda la observación de un fenómeno, pero son sujetos a reinterpretaciones posteriores en el marco de la argumentación académica. En ciencia de datos, para ser fuente de conocimiento, los datos deben ser el resultado de un proceso de revisión cuidadosa para garantizar la calidad de la información. Al igual que los datos históricos, que también son curados para establecer la conexión entre la evidencia y las afirmaciones. Es decir, tanto en ciencia de datos como en la práctica histórica, el investigador juega un papel activo en el proceso de abstracción de la realidad en datos. En el contexto de la historia, la datificación se entiende entonces como el proceso de conversión material de un documento histórico para que pueda habitar en un espacio digital y ser sujeto de una interpretación humana y computacional. Como tendencia tecnológica, la datificación descompone la información en nuevas formas de valor.<sup>46</sup>

Para dar sentido a estos elementos se pueden mencionar algunos ejemplos. Una de las características de las tendencias recientes de las bases de datos como proyectos históricos, es que estas ya no se cierran a una interacción pasiva de consulta entre la colección digital y el usuario. Ahora integran diferentes tipos de herramien-

<sup>45</sup> Mayer-Schonberger y Cukier, *Big Data*, cap. 5, EPUB..

<sup>46</sup> Margarita Shilova, “The Concept of Datafication: Definition and Examples”, *Data Science Central* 2 (junio 15, 2018). <https://www.datasciencecentral.com/the-concept-of-datafication-definition-amp-examples/>



tas para que el investigador pueda interactuar ya no solo con la base de datos, sino con los datos en el sentido complejo del término. Esto tiene dos implicaciones: una complejización del relato histórico, y una política *open source*, que permite la reutilización de los datos, adecuándolos no solo a las intenciones de quien construye el proyecto, sino ampliando el margen para que el investigador proponga nuevos modelos y haga otras preguntas con el mismo material datificado. Esto por supuesto implica una mayor demanda de “alfabetización en datos”, y en algunos casos, conocimientos en programación. Old Bailey vuelve a ser un ejemplo. Las últimas versiones ofrecen una serie de recursos para trabajar directamente sobre los datos, incluida un API (*Application Programming Interfaces*), cuya introducción dice así: “Nos comprometemos a hacer que nuestros datos estén disponibles para usos educativos, académicos, creativos e innovadores. Esta página describe algunas de las formas en que puede acceder a diferentes versiones de los datos para su análisis y reutilización” [traducción de los autores].<sup>47</sup> De este modo, el investigador puede adecuar los datos y su tratamiento a problemas aún más específicos frente a los que ofrece la base de datos original.

Existen otras estrategias. En la producción historiográfica latinoamericana se encuentra el proyecto *ARCA* (Arte colonial americano),<sup>48</sup> un portal que integra 25 mil pinturas coloniales americanas producidas entre 1550 y 1830. Las pinturas han sido datificadas a través de un análisis de cada una de las imágenes en

<sup>47</sup> “We are committed to making our data available for educational, scholarly, creative and innovative uses. This page outlines some of the ways in which you can access different versions of the data for analysis and reuse”. <https://www.oldbaileyonline.org/about/data> (consultada el 14 de octubre de 2024).

<sup>48</sup> Proyecto del historiador colombiano Jaime Humberto Borja de la Universidad de los Andes. Véase <https://arca.uniandes.edu.co/> (consultado el 14 de octubre de 2024). Existen dos versiones del proyecto; la primera se publicó en 2015 y contó con visualización de datos con tecnología Tableau y diagramas de fuerzas. La segunda versión, publicada en 2023, es relacional, lo que implica un modelado de datos distinto y una propuesta visual más compleja.

40 campos, lo que arroja cerca de 875,000 elementos etiquetados, lo cual genera un conjunto de metadatos significativo. A diferencia de las bases de datos de imágenes tradicionales de museos e instituciones, esta se construyó bajo el modelo relacional, que permite hacer colecciones particulares de casi cualquier elemento, y además cuenta con un sistema de visualizaciones en tiempo real que permite construir patrones, ver modelos y responder nuevas preguntas. Adicionalmente, es Open Source, tiene un API pública y da acceso a los datos del desarrollo. Este tipo de propuestas implican la acomodación para la historia como disciplina, de métodos nacidos en la cultura digital, y que sirven para el tratamiento de datos. En este caso se trata de las lecturas distantes y la minería de datos, los que a su vez están implicados en las perspectivas de la visualización computacional y otras técnicas de Big Data y ciencia de datos.<sup>49</sup>

Como estos dos ejemplos, se han producido otras bases de datos como *Digging Early Colonial History*, que datifica relaciones geográficas coloniales especialmente novohispanas, con un análisis computacional a gran escala.<sup>50</sup> *Fragmentarium*, por su parte, es un laboratorio en el cual los interesados pueden compartir fragmentos de manuscritos medievales y procesarlos con otros, mediante técnicas digitales que permiten catalogar, describir, transcribir, ensamblar y reutilizar los manuscritos.<sup>51</sup> Este sitio puede proveer bajo ciertas condiciones, acceso a un CMS (Content Management System) para que el investigador pueda “personalizar” su investigación. Como se podrá observar, muchas de estas bases de datos parten de la digitalización de documentos físicos que han

<sup>49</sup> En los últimos años se ha desarrollado un material importante sobre metodologías digitales aplicadas a las ciencias sociales. Véase Franco Moretti, *Lectura distante* (México: Fondo de Cultura Económica, 2013).

<sup>50</sup> Véase The Digging into Early Colonial Mexico Project, <https://www.lancaster.ac.uk/digging-ecm/> (consultado el 14 de octubre de 2024).

<sup>51</sup> Sobre el CMS, véase <https://fragmentarium.ms/pages/about/digital-laboratory> (consultado el 14 de octubre de 2024).

sido datificados y que se convierten en la etapa inicial de insumo para los proyectos de ciencia de datos.

La datificación implica para la historia, no únicamente un asunto metodológico, sino que invoca nuevas formas de producción historiográfica. La base de datos es un género emergente de hacer historia. A esto nos referimos con la afirmación que es una forma de relato, no solo una “colección” de artefactos. Finalmente, este es uno de los objetivos generales de la datificación, exponer datos al análisis de patrones, es decir, los datos masivos permiten identificar procesos que no son visibles y que al convertirlos en datos, pueden ser monitoreados, rastreados, analizados y optimizados.<sup>52</sup> En el caso de la historia, se trata de convertir un patrón en una narrativa. Entonces, ¿cómo pueden los historiadores enriquecer la lectura de los datos digitales? Los datos requieren de otras preguntas y disposiciones metodológicas y tecnológicas. En este lugar la ciencia de datos ofrece alternativas.

#### LA CIENCIA DE DATOS COMO HERRAMIENTA PARA LA PRÁCTICA HISTÓRICA

La ciencia de datos nacida originalmente de la estadística adquirió su nombre en 1974 pero solo fue hasta comienzos de los 2000 que fue reconocida como disciplina independiente. Desde entonces ha permeado todos los campos de las sociedades contemporáneas, y desde mediados de la década de 2010 se ha considerado una herramienta de apoyo para los historiadores. La ciencia de datos combina la capacidad de procesamiento informático con las herramientas de analítica estadística, lo cual permite a los investigadores abordar problemas complejos y multifacéticos. En el contexto de las humanidades, representa una posibilidad de reevaluar la evidencia histórica y acercarla al mundo digital

<sup>52</sup> Shilova, “The Concept of Datafication”; Manning, *Big Data in History*, 19.

contemporáneo, atendiendo a las tendencias investigativas que armonizan lo cuantitativo y lo cualitativo.

En línea con lo expuesto anteriormente, los datos son inferidos y no inherentes y por lo tanto, corresponden a un tipo de modelamiento específico producido en una investigación. Uno de los aspectos que ha permitido que ciertas prácticas históricas converjan en la ciencia de datos, es que ambas parten del principio de que los datos son dinámicos, productos de procesos analíticos iterativos y no meros hechos dados. Los datos se sitúan dentro de sus contextos, lo cual permite descubrir narrativas más complejas. Este reconocimiento explícito de los fundamentos epistemológicos del dato reafirma el valor de los métodos históricos, y permite marcar el paso de lo factual a una mayor comprensión de cómo reconstruimos y comprendemos el pasado.<sup>53</sup> Los datos en contexto, denominados “datos situados”<sup>54</sup> evitan las trampas de que los datos induzcan una excesiva simplificación de una realidad compleja y abre camino al diálogo interdisciplinario. La convergencia de la ciencia de datos en el contexto epistemológico de la historia ha mostrado dos transformaciones significativas, que pueden impactar ciertas formas de la evidencia y el conocimiento histórico: una nueva ampliación del concepto de archivo y una resignificación de la relación dato-fuente.

En cuanto a la primera transformación, el impacto de la ciencia de datos más evidente en la práctica histórica ha sido sobre el archivo histórico. El crecimiento exponencial de la digitalización de fuentes, así como la masiva integración de datos a la web en la década de los 2000, sembraba dudas de cómo los historiadores se podían acercar al “archivo infinito”.<sup>55</sup> La consolidación de la ciencia de datos ha aportado luces para resolver este problema,

<sup>53</sup> Drucker, “Humanities Approaches to Graphical Display”.

<sup>54</sup> Matthew Lavin, “Why Digital Humanists Should Emphasize Situated Data over Capta”, *Digital Humanities Quarterly* 15, no. 2 (2021).

<sup>55</sup> Jairo Antonio Melo Flórez, “Historia Digital: La Memoria en el Archivo Infinito”, *Historia Crítica* 1, no. 43 (2011): 82-103.

ya que ha permitido la expansión masiva del almacenamiento, la capacidad de procesar Big Data y la diversidad modal del archivo datificado que puede componerse de textos, imágenes, videos, información geográfica, páginas web, música, planos arquitectónicos o de ingeniería, código de programación, entre otros. Ese archivo multimodal puede ser procesado con modelos de inteligencia artificial, para convertir el archivo histórico en un reino de infinitas posibilidades. Algunas de las técnicas que se han integrado son la visualización, las pruebas estadísticas y los modelos entrenados con aprendizaje de máquina que permiten hacer lecturas distantes de los archivos datificados, y descifrar tendencias y patrones previamente invisibles en grandes conjuntos de datos.

La segunda transformación, es la resignificación de la relación dato-fuente que inicialmente se evidencia en la posibilidad de procesamiento de “fuentes no tradicionales” como las mencionadas anteriormente, lo cual abre la puerta para pensar múltiples y nuevos tipos de fuentes históricas adicionales a los documentos escritos. Adicionalmente, la resignificación de la relación dato-fuente se da por el potencial en el procesamiento de corpus compilados y curados por investigadores anteriores, una práctica comúnmente conocida como *Reutilización de Datos* en el campo de las Humanidades Digitales. La reutilización de datos puede ser descrita como un proceso iterativo que incluye actividades de exploración, recolección y resignificación.<sup>56</sup> Esta *reutilización de datos*, organizados y relacionados en bases de datos, es la actividad que permite a los investigadores aplicar conjuntos de datos existentes a nuevas preguntas de investigación e indagaciones historiográficas.

Uno de los mejores ejemplos es el reuso de datos en el tratamiento de la prensa del siglo XIX presente en proyectos como

<sup>56</sup> Xiaoguang Wang, Qingyu Duan y Mengli Liang, “Understanding the process of data reuse: An extensive review”, *Journal of the Association for Information Science and Technology* 72, no. 9 (2021): 1161–82.

el de *News Eye*, *Viral Texts* o el de *Oceanic Exchanges Project*.<sup>57</sup> La investigación de *Oceanic Exchanges* por ejemplo, es un trabajo colaborativo entre universidades de Estados Unidos, México, Alemania, Finlandia, Reino Unido y Holanda, para constituir una base de datos respondiendo a preguntas sobre la circulación de noticias transatlánticas y el intercambio de información transnacional y multilingüe de 1840 a 1914. Este proyecto ha dado lugar posteriormente a casi 50 productos académicos diferentes entre artículos, conferencias, visualizaciones y subconjuntos temáticos de dataset de prensa. La prensa, a pesar de ser una de las fuentes tradicionales para hacer historia, adquiere nuevas posibilidades interpretativas al haber sido estructurada en una base de datos curada, porque permite a los investigadores abordar su contenido y significado con nuevas preguntas y técnicas. Desde ese punto de vista, la evidencia histórica es dinámica al igual que los datos.

Otro caso interesante de reutilización de datos a partir de proyectos de datificación y marcado de textos, es el de procesamiento de escritos antiguos. Ejemplos como el proyecto *Historical Ink*<sup>58</sup> de prensa latinoamericana del siglo XIX que pretende la aplicación de técnicas de *layout* y el entrenamiento de modelos Deep Learning para el perfeccionamiento de la datificación de documentos antiguos mediante uso de herramientas OCR (Optical Character Recognition) para transcripción de textos, que posteriormente pueden ser procesados con modelos de lenguaje para “predecir” posibles cambios semánticos en las palabras utilizadas

<sup>57</sup> Véanse News Eye Project <https://www.newseye.eu/about/>; The Viral Texts Project. Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines <https://viraltexts.org> y Oceanic Exchanges Project. Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914. <https://oceanicexchanges.org/outcomes/> (consultadas el 14 de octubre de 2024).

<sup>58</sup> Véase Laura Manrique-Gómez, et al., “Historical Ink: 19th Century Latin American Spanish Newspaper Corpus with LLM OCR Correction”, en: *Proceedings of the 4<sup>th</sup> International Conference on natural Language for Digital Humanities NLP4DH* (Miami: Association for Computational Linguistics, 2024).

en el siglo XIX.<sup>59</sup> Por su parte, el proyecto *KITAB*, de textos árabigos que busca entender la circulación de ideas e información; el proyecto *EMLO*, enfocado en colecciones correspondencia para hacer análisis de redes de personas, y el trabajo *D-Scribes*, para mejorar la legibilidad de papiros escritos en griego antiguo;<sup>60</sup> han logrado exitosamente proveer a los historiadores, fuentes tradicionales resignificadas y con ampliadas posibilidades de investigación para descubrir patrones de circulación de textos, ubicar autorías y mejorar la comprensión de sus contenidos usando ciencia de datos y modelación de algoritmos de aprendizaje de máquina.

No obstante, la aplicación de la ciencia de datos a la investigación histórica también pone de relieve desafíos y barreras de acceso. A la sombra de la ciencia de datos, se han construido repositorios de fuentes digitales que dependen de una infraestructura adecuada para su persistencia en el tiempo<sup>61</sup> y que idealmente deberían estar organizados siguiendo los principios de acceso gratuito y atendiendo los lineamientos FAIR: *Findable* (Encontrable), *Accesible* (Accesible), *Interoperable* (Interoperable), *Reusable* (Reutilizable), lo cual es un estándar en muchos campos para la gestión de datos científicos.<sup>62</sup> Sin embargo, los grandes conjuntos de datos suelen requerir infraestructuras de cómputo costosas y el conocimiento técnico especializado para acceder a ellos, cuestión

<sup>59</sup> Tony Montes, Laura Manrique Gómez y Rubén Manrique, *Historical Ink: Semantic Shift Detection for 19th Century Spanish*, en *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*, eds. Nina Tahmasebi, Syrielle Montariol, Andrey Kutuzov, David Alfter, Francesco Periti, Pierluigi Cassotti y Netta Huebscher (Bangkok: Association for Computational Linguistics, 2024) 29-41.

<sup>60</sup> Véanse *KITAB* Project <https://kitab-project.org/about/>; *EMLO* Project <http://emlo.bodleian.ox.ac.uk/>; *D-Scribes* Project <https://d-scribes.philhist.unibas.ch/en/> (consultados el 14 de octubre 2024).

<sup>61</sup> Jo Guldi, "Scholarly Infrastructure as Critical Argument: Nine principles in a preliminary survey of the bibliographic and critical values expressed by scholarly web-portals for visualizing data", *Digital Humanities Quarterly* 14, no. 3 (2020).

<sup>62</sup> Mark D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data* 3, no. 1 (marzo 2016), 160018.

que ha generado desventajas, en particular para los historiadores del sur global.

La infraestructura puede ser entendida en un sentido amplio como aquellos requisitos tecnológicos materiales que hacen factible un proyecto de fuentes datificadas, es decir, desde el hardware y el software, hasta los grupos interdisciplinarios de investigadores capacitados para su manejo, ha sido identificado por diversos humanistas digitales<sup>63</sup> como un factor crucial para reflexionar sobre la brecha existente en la posibilidad de generar proyectos digitales de los humanistas e historiadores digitales. Esta diferenciación es abismal entre el norte y el sur global, en donde persisten profundas inequidades de acceso a la educación en ciencia y tecnología, fuentes de financiación, políticas institucionales para permitir el trabajo interdisciplinario, e incluso acceso a infraestructura mínima como la conectividad a internet.<sup>64</sup> Esta brecha de infraestructura unida a la falta de iniciativas regionales para centralizar trabajos de historia digital, explica el limitado acceso que los autores, al momento de escribir este artículo, tuvieron para consultar aquellos proyectos en el sur global que hubiesen trascendido la fase inicial de digitalización para pasar a una fase de datificación.

Afortunadamente el futuro inmediato es prometedor para disminuir tales brechas de infraestructuras. La comunidad de ciencia abierta ha generado desarrollos importantes en los últimos años, y existen numerosos repositorios que permiten indexar productos académicos atendiendo los estándares requeridos y que adquieren sustancial importancia en la publicación de resultados de inves-

<sup>63</sup> Véanse Lorella Viola, *The Humanities in the Digital: Beyond Critical Digital Humanities* (Cham: Springer, 2023); Gerben Zaagsma, Digital History and the Politics of Digitization, *Digital Scholarship in the Humanities* 38 (2, junio 2023): 830–51.

<sup>64</sup> Véase Adam Crymble y María José Afanador-Llach, “Digital History: The Globally Unequal Promise of Digital Tools for History: UK and Colombia’s Case Study” en *Teaching History for the Contemporary World*, eds. Adele Nye y Jennifer Clark (Singapur: Springer, 2021), 85–98.



tigación de científicos sociales.<sup>65</sup> Ejemplos de estos repositorios abiertos para reutilización de datos son *Figshare*, *Dryad*, *Data-verse*, *OSF*, *MendeleyData* y *Zenodo*, y modelos de código abierto como *Github*, *Hugging Face* o *Roboflow Universe*.

De la misma manera, numerosas herramientas abiertas pueden hacer que los experimentos con ciencia de datos sean cada vez más accesibles. Los conjuntos de datos curados y las colecciones históricas tienen la particularidad de ser densas en significados, aunque no necesariamente demasiado grandes en tamaño. A esta característica de los datos propios de las humanidades y ciencias sociales, se le refiere como *Deep Data*.<sup>66</sup> El concepto de *Deep Data* enfatiza que los conjuntos de datos pueden no tener el tamaño de millones de observaciones que se manejan en Big Data, pero son semánticamente muy ricos porque son los resultados de proyectos de investigación de largo plazo, con una estricta curaduría digital. Por ejemplo, el laboratorio de historia digital de la Universidad de Luxemburgo<sup>67</sup> tiene un foco particular en proyectos con esta particularidad del deep data, que contemplan investigaciones tanto de grandes colecciones de datos, así como también colecciones más reducidas en extensión, pero con mucho valor cultural para la explicación de fenómenos sociales.

Esta mirada permite que coexistan con igual valía proyectos de investigación sobre fuentes datificadas de cientos de miles de textos impresos,<sup>68</sup> con trabajos basados en algunas decenas de archivos personales que contienen una riqueza transmedia de cartas, fotografías, correspondencia, objetos, videos, documentos o di-

<sup>65</sup> Maryl Maciej et al., “A Case Study Protocol for Meta-Research into Digital Practices in the Humanities”. *Digital Humanities Quarterly* 14, no. 3 (2020).

<sup>66</sup> Benjamin Štular y Mateja Belak, “Deep Data Example: Zbiva, Early Medieval Data Set for the Eastern Alps”, *Research Data Journal for the Humanities and Social Sciences* 7, no.1 (2022): 1–13.

<sup>67</sup> Véase C<sup>2</sup>DH. Laboratorio de Historia Digital de la Universidad de Luxemburgo <https://www.c2dh.uni.lu/projects> (consultada el 14 de octubre de 2024).

<sup>68</sup> Véase Impresso Project. Media monitoring the past. <https://impresso-project.ch/> (consultada el 14 de octubre de 2024).

bujos.<sup>69</sup> La complejidad y las posibilidades de análisis de los datos profundos abren muchas posibilidades a los humanistas digitales interesados en reutilizar datos, porque no deben necesariamente depender de infraestructuras de cómputo costosas para procesar dataset de datos profundos, y además, pueden acceder a realizar experimentos de ciencia de datos en plataformas de versión gratuita como las de los *Jupyter Notebooks*. Estas posibilidades, reducen las disparidades de acceso a las nuevas tecnologías que la ciencia de datos ofrece para el avance de las ciencias humanas y en especial para los investigadores del sur global tal como lo evidencia el proyecto colaborativo *Zbiva-Ariadne*. El proyecto condensó información de 1,106 sitios arqueológicos de los Alpes eslavos, del año 500 al año 1000, con 21 metadatos diferentes. Este trabajo que se extendió durante más de 30 años puede ser leído en una hoja de cálculo tradicional, pero la riqueza de su información permite mostrar, por ejemplo, detalles científicos de 3,482 tumbas y más de 15,000 artefactos cuidadosamente datados, con la ayuda de algoritmos de visualización con la información geoespacial.<sup>70</sup>

Otro de los desafíos que trae la ciencia de datos para los humanistas, es abordar críticamente la excesiva confianza sobre la “objetividad” de los datos. A diferencia de otros campos de aplicación de la ciencia de datos en donde hay “hechos” solo con analizar y organizar los datos, en las ciencias sociales se requiere un trabajo adecuado de traducción, interpretación y representación de los datos.<sup>71</sup> Por esta razón, en esta tarea de datificación, es importante establecer parámetros de cómo estandarizar, etiquetar y equilibrar un conjunto de datos para evitar modelos compu-

<sup>69</sup> Véase Radio Luxembourg. The station that changed our world. <https://radio.lu/en/home/> (consultada el 14 de octubre de 2024).

<sup>70</sup> Véase Zbiva-Ariadne Project. <http://zbiva.zrc-sazu.si/> (consultada el 14 de octubre de 2024).

<sup>71</sup> Amanda Wasielewski, *Computational Formalism. Art History and Machine Learning* (Cambridge: The Massachusetts Institute of Technology Press, 2023), 22-24.

tacionales que posteriormente sean entrenados con sesgos no reconocidos y por lo tanto, se llegue a conclusiones poco claras. Wasielewski trae una cita de Andrew Ferguson, que resume el problema: “Los sistemas basados en datos incorporan los prejuicios de sus constructores y de la sociedad en general. Los datos no son ciegos. Los datos somos nosotros, simplemente reducidos a código binario” [traducción de los autores].<sup>72</sup> A esta situación se suma una problemática que hace parte de los debates de la historiografía desde el cambio del paradigma discursivo vinculado a la objetividad del discurso.

Diversos investigadores han llamado la atención sobre este problema, porque una cualidad que tienen los datos es que, en el momento de ser procesados por medios digitales para ser transformados, en este caso, en evidencia histórica, se convierten en abstracciones con una cualidad artificial.<sup>73</sup> El problema es central porque la pregunta que sigue cuestionando la elaboración histórica es cómo se estructura la relación con los datos y cómo se construyen argumentos con ellos. El procesamiento por máquina no invalida el carácter social que poseen los datos, al tiempo que es el entorno cultural el que determina la captura y los tipos de interpretaciones que determinan los datos disponibles.<sup>74</sup> Por esta razón los datos no solo cuentan lo que está ahí, contienen un relato. La datificación sin embargo, presenta una herramienta nueva para la práctica histórica y su reflexión sobre la objetividad del discurso construido a partir de datos. El proceso de datificar al constituir una base de datos, permite dejar explícita la documentación relativa a las consideraciones metodológicas y materiales en las que los investigadores basaron sus decisiones interpretativas,

<sup>72</sup> “Data-based systems import the biases of their builders and the larger society. Data is not blind. Data is us, just reduced to binary code.” Wasielewski, *Computational Formalism*, 25.

<sup>73</sup> Owens, “Defining Data for Humanists”.

<sup>74</sup> Sun-Ha Hong, “Predictions Without Futures”, *History and Theory* 61, no. 3 (2022): 375.

las ausencias o silencios que permanecen, así como los posibles sesgos en la selección de las fuentes a datificar.

La documentación del proyecto *I-Cem* es representativa de esta posibilidad epistemológica para el debate historiográfico, y ejemplifica el reconocimiento explícito de los elementos “subjetivos” o de interpretación que median el proceso de datificación. El *I-CemProject*, uno de los recursos de datos históricos más grandes del mundo que centraliza la información de censos de Gran Bretaña de 1851 a 1911, concentra alrededor de 185 millones de entradas de censos por persona y está a disposición pública de investigadores en temas de historia económica, social y demográfica. En la guía metodológica del proyecto, se documentó todo el proceso de datificación, que incluye fotos de las fuentes físicas originales de los censos, muchas de ellas en manuscrito y que requirieron técnicas de paleografía para ser leídas, transcritas, y luego modeladas en una base de datos para vaciar allí la información que contenían.

Tal como lo narran en detalle en su guía metodológica, los académicos vinculados al proyecto *I-Cem*, realizaron una investigación histórica detallada para darle contexto a cada uno de los censos datificados, previo a definir qué significado tendrían las variables seleccionadas<sup>75</sup> para su transcripción en los registros digitales. También generaron estrategias para determinar equivalencias de magnitudes en el tiempo y así posibilitar estudios comparativos; determinaron qué hacer con los datos faltantes, repetidos, confusos o ilegibles, y definieron los metadatos que

<sup>75</sup> En promedio, son 40 variables estandarizadas por cada censo, que recogen condiciones relativas al lugar de nacimiento, sexo, edad, estado marital, posición en la familia, lenguaje hablado, género y edad de la cabeza de familia, ocupación, condición de discapacidad, número de habitaciones en la casa, tamaño de la casa, número de hijos, de familiares vivos, de sirvientes, de mujeres solteras y hombres solteros en el hogar y edades, entre otros. En el sitio web hay disponible documentación de diccionarios que dan cuenta de los temas generales de estas variables.: *I-CEM Project*. <https://www.essex.ac.uk/research-projects/integrated-census-microdata#documents> (consultada el 14 de octubre de 2024).

consideraron más importantes y que reunían la citación bibliográfica de la fuente, la unidad geográfica cubierta, si se usaron medidas de equivalencia en algunos datos, entre otros. La información puede ser consultada parcialmente en línea a través de un visualizador de muestra que permite conjugar variables para visualizar su comportamiento en el tiempo, de tal manera que el investigador puede tener una experiencia diferente con las fuentes. Esta experiencia de consultar las fuentes datificadas permite probar muchas hipótesis de relaciones y patrones, y hacer nuevas preguntas históricas a las fuentes tradicionales. Con la base de datos curada del *I-Cem* se puede por ejemplo abordar un estudio de la condición de discapacidad en el tiempo en Gran Bretaña, las ocupaciones de las mujeres en Irlanda o la variación del uso de dialectos en Escocia durante la revolución industrial. El proyecto es una muestra de una producción historiográfica no tradicional, que da cuenta de una integración del pensamiento computacional con el pensamiento histórico, y que permite comprender que la labor de datificación de las fuentes, no implica desnaturalizar las fuentes, sino por el contrario, preservarlas, dotarlas de un sentido contextual y permitir su accesibilidad para el avance de la investigación histórica.

La ciencia de datos está desempeñando un papel crucial en la expansión de las posibilidades de la práctica historiográfica moderna, en la medida en que los datos y la evidencia histórica deben considerarse como una abstracción analítica producto de un proceso y no meramente “datos”. Los datos y la evidencia histórica están situados en un contexto, y sirven de apoyo a argumentos interpretativos abiertos a relecturas y nuevas comprensiones. Frente a datos digitales y algoritmos avanzados, los historiadores no solo tienen el reto de adaptarse a los nuevos métodos de análisis, sino también afrontar el reto de generar una exposición detallada del aparato epistemológico empleado en el uso de estas tecnologías. Un compromiso disciplinar que involucra garantizar el tratamiento adecuado de los datos, desde su recopilación y

preparación hasta su interpretación y presentación en contextos académicos y públicos.

Sin embargo, aquí se encuentra un último reto. En el ecosistema de la ciencia de datos, la colaboración se convierte en una llave fundamental para develar las particularidades del pasado, y redefinir el papel del historiador como parte de un equipo científico diverso. Como bien lo menciona Wasielewski,<sup>76</sup> el problema no es que no se pueda producir una investigación relevante a partir de un conjunto cuantitativo de datos, sino que a menudo los investigadores no logran interrogar las capas de representación que surgen allí. Y esa, justamente, debe ser la labor de los historiadores e historiadoras: comprender y usar críticamente las posibilidades que brinda la ciencia de datos para datificar el pasado, lo cual implica el trabajo con desarrolladores y programadores.

#### REFLEXIONES FINALES: HISTORIA PREDICTIVA Y FUTURISMO EN LA HISTORIOGRAFÍA

La ciencia de datos está impactando el campo de la historia y ha posibilitado el trabajo colaborativo a partir de las narraciones históricas de las bases de datos, que son posteriormente resignificadas por parte de nuevos investigadores. Sin dudas, el pasado no puede ser reducido únicamente en datos, pero son una forma de significación de este. Los datos son abstracciones de realidades más complejas y de ahí que las bases de datos sean una producción y una narrativa historiográfica. Además, estas fuentes datificadas generan un potencial impacto social al servir de insumo para la modelación de algoritmos de inteligencia artificial, ya que pueden reproducir patrones de información sesgada si las fuentes datificadas no han sido debidamente tratadas. Esta situación justifica la participación transversal de historiadores y científicos sociales

<sup>76</sup> Wasielewski, *Computational Formalism*, 23.

en los proyectos de la ciencia de datos, para asegurar la reflexión juiciosa y la datificación crítica de las fuentes. Es decir, la historia se está nutriendo de la forma de hacer de la ciencia de datos, pero recíprocamente, la ciencia de datos puede nutrirse de las formas propias de hacer historia, en especial del pensamiento crítico sobre la curaduría de fuentes. Pero hay otro aspecto fundamental en el que se debe seguir reflexionando en próximas investigaciones para comprender cómo está cambiando la ciencia de datos la caracterización de las temporalidades que aborda la historia.

La llegada de la cultura digital ha permitido que la historia influya no solo en el pasado y presente, sino también en múltiples futuros posibles, y aquí es donde la ciencia de datos actúa como catalizador transformador. Algunos historiadores sugieren que la historia puede contribuir a la comprensión de grandes problemas contemporáneos como el cambio climático y la gobernanza internacional, mediante técnicas de big data y minería de texto.<sup>77</sup> También se ha explorado la historia contrafactual, que surge como una práctica valiosa en el diseño de videojuegos y la novela histórica, permitiendo la evaluación de futuros técnicamente posibles y la reflexión sobre eventos pasados.<sup>78</sup> Vinale<sup>79</sup> detalla por ejemplo, cómo la novela histórica italiana ha utilizado este recurso literario para ofrecer enfoques críticos y educativos sobre la memoria colectiva y la historia alternativa. Por su parte, Tobias Blanke<sup>80</sup> defiende el uso de la analítica predictiva como un método para entender relaciones pasadas y enriquecer inter-

<sup>77</sup> Jo Guldi y David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014).

<sup>78</sup> Ylva Grufstedt, *Shaping the Past: Counterfactual History and Game Design Practice in Digital Strategy Games* (Boston: De Gruyter, 2022).

<sup>79</sup> Adriano Vinale, "From History-writing to (Hi)story-telling: Historical Novel, Alternate/Counterfactual History and Implicit Uchrony" *History* 108, no. 382 (2023a): 407–20. Adriano Vinale, "(Hi)story-Telling: An Introduction to Italian Alternate and Counterfactual History". *History* 108, no. 382 (2023b): 355–64.

<sup>80</sup> Tobias Blanke "Predicting the Past", *Digital Humanities Quarterly* 12, no. 2 (2018).

pretaciones del pasado. La historia predictiva, por ende, combina métodos científicos y el pensamiento crítico para proporcionar una comprensión más amplia del pasado y guiar el uso de la historia para el presente y futuros alternativos. La historia se ubica en la cúspide de una nueva era, donde la experiencia del pasado informa sobre las probabilidades del futuro.

Sin embargo, estas posibilidades analíticas en la historiografía todavía son tratadas con recelo por la tradición historiográfica. Incluso la historiadora digital Jo Guldi<sup>81</sup> en su reciente trabajo sobre el arte de minería de textos para la historia, se abstiene explícitamente de utilizar el término “predicción” en historia para que no se entienda como sinónimo de las “fantasías de la profecía”, contrario a la “factualidad que se busca en los datos”. Guldi advierte que convertir el análisis de datos en una herramienta de predicción sin tener en cuenta el contexto social y cultural puede llevar a errores significativos. La predicción de los modelos siempre debería estar acompañada de una comprensión profunda de las realidades que construyen esos datos. Por tanto, los historiadores deberían colaborar con científicos de datos para garantizar que el enfoque crítico de las ciencias humanas enriquezca la comprensión de los datos y el horizonte de impacto social de las predicciones de los modelos.

Ahora bien, es importante aclarar que la “predicción” en historia difiere fundamentalmente de la predicción estadística tradicional. La estadística se basa en la cuantificación y patrones del pasado para prever la ocurrencia de futuros eventos con cierta probabilidad. Por otro lado, la historia “predice” explicaciones de eventos que ya han ocurrido, mediante narrativas construidas sobre hipótesis plausibles derivadas de un estudio riguroso de los datos. Es decir, del uso de modelos como el *machine learning* se derivan explicaciones probables estadísticamente, que requieren la

<sup>81</sup> Jo Guldi, *The Dangerous Art of Text Mining* (Cambridge: Cambridge University Press, 2023).



aplicación de pensamiento crítico para evitar conclusiones erróneas o especulaciones infundadas.

En la práctica histórica ya están sucediendo esas “predicciones”. La Unión Europea por ejemplo, viene materializando una iniciativa muy ambiciosa en el marco del proyecto supranacional *Time Machine*.<sup>82</sup> El proyecto tiene como objetivo unir el pasado de Europa con potentes infraestructuras y tecnologías digitales, creando un sistema de información digital colectivo que mapee la evolución económica, social, cultural y geográfica europea a través de los tiempos. En el enfoque propuesto, la digitalización es solo el primer paso de una larga serie de procesos de extracción, incluida la segmentación de documentos y la comprensión mejorada por aplicaciones de Realidad Aumentada/Virtual (AR/VR), que conducen a simulaciones de hipotéticas reconstrucciones espacio-temporales en 4D. Tal como indica el portal web del proyecto, la reflexión crítica del conocimiento histórico impacta las decisiones del presente y la proyección futura de la sociedad: “Dichos modelos computacionales son recursos clave para desarrollar nuevas reflexiones críticas sobre nuestro pasado y futuro, permitiendo nuevos conocimientos para historiadores, científicos sociales, profesionales de las artes creativas, formuladores de políticas y para el público en general, con un importante denominador común: contribuir a la toma de decisiones informadas. Desde la vida cotidiana hasta lo académico, profesional y político.”

Las series de datos históricas, obtenidas de fuentes datificadas se han convertido en el insumo de un modelo de aprendizaje de máquina que permite predecir los comportamientos o explicaciones más probables de nuevos datos. Por ejemplo, en el proyecto *Arca* mencionado anteriormente, se identificaron cerca de 200 gestos quirológicos<sup>83</sup> (posiciones de las manos) con sus respectivos


<sup>82</sup> Véase The Time Machine Project. <https://www.timemachine.eu/about-us/> (consultada el 14 de octubre de 2024).

<sup>83</sup> Jaime Humberto Borja Gómez, *ARCA* (Arte Colonial Americano) (Principal) [Data set] Zenodo. <https://doi.org/10.5281/zenodo.10864546>

significados, mediante técnicas mixtas de datos semi-estructurados y modelos de aprendizaje de máquina. Los metadatos de estos gestos permiten “predecir” o identificar otros gestos y explicar la cultura gestual colonial americana. Esas predicciones, que son el corazón de la inteligencia artificial, también deberían estar sujetas a una cuidadosa reflexión interdisciplinaria. Los datos, como observaciones del pasado, tienen un poder explicativo pero no implican una causalidad lineal. Por ese motivo, los resultados, y no solo los datos de entrada también deben ser interpretados por los historiadores. Esta situación se puede comprender mejor en un ejemplo de la práctica histórica con ciencia de datos que fue mencionado anteriormente. El proyecto *D-Scribes* reúne una colección de papiros griegos antiguos que originalmente se encuentran muy deteriorados, pero fueron datificados y con modelos de inteligencia artificial predictiva se logran identificar rasgos paleográficos comunes. El resultado de estos modelos computacionales permite completar el texto de los papiros así como argumentos de datación para su ubicación histórica. En este proyecto, la “predicción” es el texto completo del papiro descifrado. Por ende, la labor del historiador también implica curar el relato contrafactual más probable contenido en documentos poco legibles.

El proyecto *KITAB* mencionado anteriormente, también muestra un ejemplo de historiadores siendo curadores de predicciones, se predice sobre el pasado, sobre hipótesis explicativas plausibles del pasado. El proyecto *KITAB* usa modelos con fuentes antiguas documentales árabes datificadas, e identifica continuidades gramaticales y sintácticas en diferentes textos, lo cual permite “predecir” autorías y también identificar hipótesis plausibles de circulación y reuso de textos antiguos. Sin embargo, cada una de estas posibilidades, es revisada cuidadosamente por los historiadores y lingüistas expertos. La prevención de la historia tradicional frente al uso de categorías como “predicción” o “contrafactual”, se deriva de que su significado se asocia comúnmente a la especulación propia de contextos no académicos, o recurso estético de

artistas y literatos. No obstante, la historia predictiva y las narraciones contrafactuales pueden ser nuevas alternativas académicas para los historiadores que se sirven de modelos de aprendizaje de máquina.

Con la masificación de los modelos de inteligencia artificial, presentes ahora en la cultura digital, es posible que los historiadores e historiadoras comencemos a construir conocimiento mediante la experimentación con ciencia de datos. Hay un potencial muy importante en la práctica de reutilización de las observaciones datificadas del pasado, con el fin de usarlas para predecir explicaciones factibles. Es imperativo sin embargo, que el pensamiento crítico de la historia sea aplicado rigurosamente, para advertir que los resultados de algoritmos no implican una causalidad lineal en relación con un hecho del presente o una posibilidad para especular irresponsablemente con acontecimientos del futuro: “El papel de la predicción no se limita a técnicas analíticas específicas que, por ejemplo, sintetizan nuevos resultados sobre la base del modelo y los datos de entrenamiento o que intentan específicamente identificar futuros acontecimientos como la reincidencia”.<sup>84</sup> En este contexto, la predicción histórica produce “nuevos problemas”, abarca futuros estadísticamente probables y explicaciones plausibles, y se convierte en una categoría epistemológica en el trabajo historiográfico. 

## BIBLIOGRAFÍA

### *Fuentes Documentales*

El sitio web de las fuentes documentales datificadas referidas a continuación, y que sirvieron de base para este artículo, fueron revisadas por última vez el 14 de octubre de 2024.

ARCA Project, Arte colonial Americano (ARTCA): Cultura visual colonial. <https://arca.uniandes.edu.co/> <https://losingeniosdelpincel.uniandes.edu.co/#!/intro/1>

<sup>84</sup> Véase la interesante reflexión sobre el problema del impacto de la predicción en la historia, en Sun-Ha Hong, “Predictions Without Futures”, 381.

C<sup>2</sup>DH. Laboratorio de Historia Digital de la Universidad de Luxemburgo.  
<https://www.c2dh.uni.lu/projects/>  
 D-Scribes Project. <https://d-scribes.philhist.unibas.ch/en/>  
 Early Modern Letters Online Project (EMLO). <http://emlo.bodleian.ox.ac.uk/>  
 Fragmentarium Project. Laboratory for Medieval Manuscript Fragments.  
<https://fragmentarium.ms/> <https://fragmentarium.ms/pages/about/digital-laboratory>  
 I-CEM Project, <https://www.essex.ac.uk/research-projects/integrated-census-microdata#>  
 Impresso Project. Media monitoring the past. <https://impresso-project.ch/>  
 KITAB Project. <https://kitab-project.org/about/>  
 London Lives Project. 1690 to 1800. Crime, Poverty and Social Policy in the Metropolis, <https://www.londonlives.org/index.jsp>  
 News Eye Project. <https://www.newseye.eu/about/>  
 Oceanic Exchanges Project, Tracing Global Information Networks In Historical Newspaper Repositories, 1840-1914. <https://oceanicexchanges.org/>  
 PASE Project. The Prosopography of Anglo-Saxon England. <https://pase.ac.uk/index.html>  
 Radio Luxembourg. The station that changed our world. <https://radio.lu/en/home/>  
 The Proceedings of the Old Bailey Project, <https://www.oldbaileyonline.org/>  
 The Portal to Texas History Project, <https://texashistory.unt.edu/>  
 The Viral Texts Project. Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines. <https://viraltxts.org>  
 The Valley of the Shadow Project. Two Communities in the American Civil War, <https://valley.newamericanhistory.org/>  
 Time Machine Project. <https://www.timemachine.eu/about-us/>  
 Zbiva-Ariadne Project. <http://zbiva.zrc-sazu.si/>

### *Obras Publicadas*

Allés Torrent, Susanna. “Sobre la complejidad de los datos en Humanidades o cómo traducir las ideas a datos”. *Revista de Humanidades Digitales*, no. 4 (2019): 1-28.  
 Bell, David. *Cyberculture Theorists: Manuel Castells and Donna Haraway*. Nueva York: Routledge, 2006.

- Berry, David. "Digital Humanities: First, Second And Third Wave". *Stunlaw philosophy and critique for a digital age*, (2011). <http://stunlaw.blogspot.com/2011/01/digital-humanities-first-second-and.html>
- Berry, David M. y Fagerjord Anders. *Digital Humanities: Knowledge and Critique in a Digital Age*. Cambridge: Polity Press, 2017.
- Blanke, Tobias. "Predicting the Past". *Digital Humanities Quarterly* 12, no. 2 (2018).
- Brügger, Niels e Ian Milligan. *The Sage Handbook of Web History*. Londres: Sage Publications, 2018.
- Carr, Eduard. ¿Qué es la historia? Barcelona: Ariel, 1984.
- Castells, Manuel. *Change. The Impact of the Internet on Society: A Global Perspective.*, Open Mind BBVA, 2014.
- Crymble, Adam. *Technology and the Historian: Transformations in the Digital Age*. Champaign: University of Illinois Press, 2021.
- Crymble, Adam y María José Afanador-Llach. Digital History: The Globally Unequal Promise of Digital Tools for History: UK and Colombia's Case Study. En *Teaching History for the Contemporary World*, eds. Adele Nye y Jennifer Clark, 85-98. Singapur: Springer. (2021). [https://doi.org/10.1007/978-981-16-0247-4\\_7](https://doi.org/10.1007/978-981-16-0247-4_7)
- Denning, Peter J. y Matti Tedre. *Computational Thinking*. Cambridge: The Massachusetts Institute of Technology Press, 2019.
- Dobson, James. "Interpretable Outputs: Criteria for Machine Learning in the Humanities". *Digital Humanities Quarterly* 15, no. 2 (2021). En: <https://www.digitalhumanities.org/dhq/vol/15/2/000555/000555.html>.
- Drucker, Johanna. "Humanities Approaches to Graphical Display". *Digital Humanities Quarterly* 5, no. 1 (2011). En: <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Drucker, Johanna. *The Digital Humanities Coursebook. An Introduction to Digital Methods for Research and Scholarship*. Nueva York: Routledge, 2021.
- Economic Commission for Europe of the United Nations (UNECE) Terminology on Statistical Metadata. *Conference of European Statisticians Statistical Standards and Studies* no. 53. Ginebra, 2000.
- Flanders, Julia y Fotis Jannidis. *The Shape of Data in the Digital Humanities*. Nueva York: Routledge, 2019.
- Foster, Ian, Rayid Ghani, Ron Jarmin, Frauke Kreuter y Julia Lane. *Big Data and Social Science Data Science Methods and Tools for Research and Practice*. Londres: CRC Press Taylor & Francis Group, 2021.

- Gibbs, Fred y Trevor Owens. "The Hermeneutics of Data and Historical Writing". En *Writing History in the Digital Age*, eds. Jack Dougherty y Kristen Nawrotzki, 159-70. Ann Arbor: The University of Michigan Press, 2018.
- Gitelman, Lisa. *Raw Data is an Oxymoron*. Cambridge: Massachusetts Institute of Technology Press, 2013.
- Graham, Shawn, Ian Milligan y Scott Weingart. *Exploring big historical data. The Historian's Microscope*. Londres: Imperial College Press, 2016.
- Grufstedt, Ylva. *Shaping the Past: Counterfactual History and Game Design Practice in Digital Strategy Games*. Boston: De Gruyter Oldenbourg, 2022. En: <https://doi.org/10.1515/9783110692488>
- Guldi, Jo. "Scholarly Infrastructure as Critical Argument: Nine principles in a preliminary survey of the bibliographic and critical values expressed by scholarly web-portals for visualizing data". *Digital Humanities Quarterly* 14, no. 3. (2020) <http://digitalhumanities.org:8081/dhq/vol/14/3/000463/000463.html>
- Guldi, Jo. *The Dangerous Art of Text Mining*. Cambridge: Cambridge University Press, 2023. En: <https://doi.org/10.1017/9781009263016>
- Guldi, Jo y David Armitage. *The History Manifesto*. Cambridge: Cambridge University Press, 2014.
- Hai-Jew, Shalin, ed. *Data Analytics in Digital Humanities*. Cham: Springer, 2017.
- Hayles, Katherine. "Narrative and Database: Natural Symbionts", *PMLA: Publications of the Modern Language Association of America, Special Topic: Remapping Genre* 122, no. 5 (octubre., 2007): 1603-8.
- Hong, Sun-Ha. "Predictions Without Futures", *History and Theory* 61, no.3 (2022): 371-90.
- Jiawei, Han, Micheline Kamber y Jian Pei, *Data Mining Concepts and Techniques*. Waltham: Morgan Kaufmann, 1999.
- Jones, Steven E. *Roberto Busa, S. J. and the Emergence of Humanities Computing. The Priest and The Punched Cards*. Nueva York: Routledge, 2016.
- Kirkpatrick, Keith. "Parallel Computational Thinking", *Communications of the ACM* 60, no. 12 (2017): 17-19.
- Koby, Mike, Noa Ragonis, Rinat B. Rosenberg-Kima y Orit Hazzan. "Computational Thinking in the Era of Data Science". *Communications of the ACM* 65, no. 8 (2022): 33-35.
- Lavin, Matthew. "Why Digital Humanists Should Emphasize Situated Data over Capta". *Digital Humanities Quarterly* 15, no. 2 (2021). En: <https://www.digitalhumanities.org/dhq/vol/15/2/000556/000556.html>

- Le Deuff, Olivier. *Digital Humanities. History and Development*. Londres: Wiley, 2018.
- Le Goff, Jacques. *El orden de la memoria*. [1977], Barcelona: Paidós, 1991.
- Levy, Pierre. *Cibercultura. La cultura de la sociedad digital*. Barcelona: Anthropos, 2007.
- Maciej, Maryl, Costis Dallas, Jennifer Edmond, Ingrida Kelpšienė, Michelle Doran, Marta Kołodziejska y Klaudia Grabowska. "A Case Study Protocol for Meta-Research into Digital Practices in the Humanities". *Digital Humanities Quarterly* 14, no. 3 (2020). En: <http://digitalhumanities.org:8081/dhq/vol/14/3/000477/000477.html>
- Manning, Patrick. *Big Data in History*. Hampshire: Palgrave MacMillan, 2013.
- Manovich, Lev. *El lenguaje de los nuevos medios de comunicación. La imagen en la era digital*. Barcelona: Paidós, 2005.
- Manrique-Gómez, Laura, Tony Montes, Arturo Rodríguez-Herrera y Rubén Manrique. 2024. Historical Ink: 19th Century Latin American Spanish Newspaper Corpus with LLM OCR Correction. En *Proceedings of the 4th International Conference on natural Language for Digital Humanities NLP4DH*. Miami: Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.2407.12838> [En prensa].
- Mayer-Schonberger, Viktor y Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.
- Melo Flórez, Jairo Antonio. "Historia Digital: La Memoria en el Archivo Infinito". *Historia Crítica* 1, no. 43 (2011): 82-103. En: <https://doi.org/10.7440/histcrit43.2011.06>
- Milligan, Ian. *History in the Age of Abundance How the Web Is Transforming Historical Research*. Montreal: McGill-Queen's University Press, 2019.
- Milligan, Ian. *The Transformation of Historical Research in the Digital Age*. Cambridge: Cambridge University Press. En: <https://doi.org/10.1017/9781009026055>
- Montes, Tony, Laura Manrique-Gómez y Rubén Manrique. Historical Ink: Semantic Shift Detection for 19th Century Spanish. En: *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*. eds. Nina Tahmasebi, Syrielle Montariol, Andrey Kutuzov, David Alfter, Francesco Periti, Pierluigi Cassotti y Netta Huebscher, 29-41. Bangkok: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.lchange-1.4>

- Moretti, Franco *Lectura distante*. México: Fondo de Cultura Económica, 2013.
- Morreale, Laura K. 2022. History As Antidote: The Argument For Documentation In Digital History. *History and Theory* 61, no. 4 (diciembre), 74-66.
- Organisation for Economic Co-operation and Development (OECD). *OECD Glossary of Statistical Terms*. París: OECD Publishing, 2008.
- Owens, Trevor. “Defining Data for Humanists: Text, Artifact, Information or Evidence?”, *Journal of the Digital Humanities* 1, no. 1 (invierno) 2011. *PMLA. Publications of the Language Association of America*. 5, Vol. 122. 2007.
- Robertson, Stephen. “The Properties Of Digital History”, *History and Theory*, 61, no. 4 (diciembre 2022): 86-106.
- Robertson, Stephen y Lincoln A. Mullen. Arguing with Digital History: Patterns of Historical Interpretation. *Journal of Social History* 54, no. 4 (verano 2021): 1005-22.
- Rojas, Antonio. “Big data en las humanidades digitales. Nuevas conversaciones en el contexto académico global”. *Anuario acle de cultura digital* (2017):68-78.
- Salmi, Hannu. *What is Digital History?* Cambridge: Wiley, 2021.
- Shilova, Margarita. The Concept of Datafication: Definition and Examples. *Data Science Central* 2 (junio 2018). <https://www.datasciencecentral.com/the-concept-of-datafication-definition-amp-examples/>
- Schiuma, Giovanni y Daniela Carlucci. *Big Data in the Arts and Humanities: Theory and Practice*. Boca Ratón: CRC Press Taylor & Francis Group, 2018.
- Schöch, Christof. “Big? Smart? Clean? Messy? Data in the Humanities”, *Journal of Digital Humanities* 2, no. 3 (Verano 2013).
- Schreibman, Susan, Ray Siemens y John Unsworth, eds. *A New Companion to Digital Humanities*. Chichester: Wiley Blackwell, 2016.
- Sinclair, Stéfan y Geoffrey Rockwell. “Text Analysis and Visualization: Making Meaning Count”. En: *A New Companion to Digital Humanities*, editado por S. Schreibman, R. Siemens, y J. Unsworth, 274-90. Wiley Blackwell, 2016.
- Štular, Benjamin y Mateja Belak. “Deep Data Example: Zbiva, Early Medieval Data Set for the Eastern Alps”. *Research Data Journal for the Humanities and Social Sciences* 7, no.1 (2022), 1-13. En: <https://doi.org/10.1163/24523666-bja10024>
- Tanaka, Stefan. “The Old and New of Digital History”, *History and Theory* 61, no. 4 (diciembre 2022): 3-18.



- Torgerson, Jesse W. "Historical Practice in the Era of Digital History". *History and Theory* 61, no. 4 (2022): 37–63. En: <https://doi.org/10.1111/hith.12276>
- Torrent, Susanna Alles. "Sobre la complejidad de los datos en Humanidades, o cómo traducir las ideas a datos", *Revista de Humanidades Digitales*, No. 4 (2019).
- Underwood, Ted. "How not to do things with words." *Blog de Ted Underwood, Humanista Digital*. (25 de agosto de 2012). <https://tedunderwood.com/2012/08/25/how-not-to-do-things-with-words/>
- Vinale, Adriano. "From History-writing to (Hi)story-telling: Historical Novel, Alternate/Counterfactual History and Implicit Uchrony". *History* 108, no. 382 (2023a): 407–20. En: <https://doi.org/10.1111/1468-229X.13354>
- Vinale, Adriano. "(Hi)story-Telling: An Introduction to Italian Alternate and Counterfactual History". *History* 108, no. 382 (2023b): 355–64. En: <https://doi.org/10.1111/1468-229X.13355>
- Viola, Lorella. *The Humanities in the Digital: Beyond Critical Digital Humanities*. Cham: Springer, 2023. <https://doi.org/10.1007/978-3-031-16950-2>
- Wang, Xiaoguang, Qingyu Duan y Mengli Liang. "Understanding the process of data reuse: An extensive review". *Journal of the Association for Information Science and Technology* 72, no. 9 (2021): 1161–82. En: <https://doi.org/10.1002/asi.24483>
- Wasielewski, Amanda. *Computational Formalism. Art History and Machine Learning*. Cambridge: The Massachusetts Institute of Technology Press, 2023.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific Data* 3, no. 1 (marzo 2016), 160018. En: <https://doi.org/10.1038/sdata.2016.18>
- Wing, Jeannette. "Computational thinking and thinking about computing". *Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences* 366: (2008), 3717–25.
- Zaagsma, Gerben. "Digital History and the Politics of Digitization". *Digital Scholarship in the Humanities* 38, 2 (junio 2023): 830–51. <https://doi.org/10.1093/llc/fqac050>
- Zoltán, Simon Boldizsár y Marek Tamm. *History and Theory* 60, no. 1 (2021): 3–22.