Revista de Humanidades Digitales

Sobre la complejidad de los datos en Humanidades, o cómo traducir las ideas a datos

On the Complexity of Data in the Humanities, or How to Translate Ideas into Data

Susanna ALLÉS TORRENT
University of Miami
susanna_alles@miami.edu

ABSTRACT

This article explores the concept of digital data in the Humanities through three different perspectives: first, I analyze the meaning of the term *data* from a historical point of view, and I examine it within the context of the digital humanities highlighting its technical meaning. Next, I discuss the nature and the typologies of data, and I raise some debates from recent years, such as its apparent objectivity. Finally, and dealing with production, management and analysis of data, I delve into the phenomenon of big data and the challenges of its application in the humanities. The goal of this paper is to stress the centrality of data in the humanistic work, and how, for today's humanist, there are inescapable digital competencies.

KEYWORDS

Data, Humanities, Digital Humanities, Big Data, Culture of Data.

RESUMEN

Este artículo explora el concepto de datos informáticos aplicados a las Humanidades a través de tres perspectivas diferentes: en primer lugar, se analiza el significado del término desde un punto de vista histórico y se sitúa en el contexto de las Humanidades Digitales prestando atención a su acepción más técnica. A continuación, se discute sobre la naturaleza, las tipologías y se abordan algunos debates de los últimos años, como es el de su aparente objetividad. Por último, y en conexión con la producción, gestión y análisis de los datos, se toma en consideración el fenómeno big data y los desafíos de su aplicación en el terreno de las Humanidades. Se pretende, en definitiva, poner de relieve el papel central de los datos en el quehacer humanístico y cómo, para el humanista actual, existen competencias digitales básicas ineludibles.

PALABRAS CLAVE

Datos, Humanidades, Humanidades Digitales, big data, cultura de datos.

RHD 4 (2019)

ISSN
2531-1786

Recibido 14/07/2019 – Aceptado 20/07/2019

1. HUMANIDADES DIGITALES Y LA CENTRALIDAD DE LOS DATOS

El término *dato* no es un concepto ajeno a la historia de las Humanidades o las Ciencias Sociales, pero sí lo ha sido, en cambio, la adopción de este término bajo la acepción de dato informático[1]. Reconducir las investigaciones humanísticas a la idea de dato y a su contexto material y digital no es siempre una tarea fácil, ni un ejercicio al que el humanista esté acostumbrado. La adopción de metodologías digitales, tanto en su uso instrumental como en el hermenéutico, es un hecho consumado en muchas disciplinas científicas, y paulatinamente se ha visto también abanderado por lo que se bautizó como Humanidades Digitales (HD). Esta disciplina surge como una consecuencia inevitable del cambio tecnológico hacia un paradigma digital que ha abrazado todas las esferas de la sociedad a nivel global, incluyendo la cultura o las letras. Alrededor de esta centralidad de los datos, han surgido hoy en día un sinfín de profesiones dedicadas exclusivamente a su gestión, su creación, recolección, estructuración, almacenaje, mantenimiento, e incluso –en el sector privado– su comercialización. Por todo ello, entender la materialidad de los datos informáticos, su recolección y su gestión se convierte ahora en una competencia ineludible para cualquier investigación.

En el presente artículo exploro el concepto de datos informáticos aplicados a las Humanidades insistiendo en algunos puntos clave: en primer lugar, analizo el significado del término desde un punto de vista histórico y trato de situarlo en el contexto de las HD poniendo especial énfasis en su sentido más técnico. A continuación, me adentro en la naturaleza, las tipologías y en algunos debates de los últimos años, como es el de la aparente objetividad de los datos. En fin, y en conexión con la producción, gestión y análisis de los datos, tomo en consideración el fenómeno big data y los desafíos de su aplicación en el terreno de las Humanidades.

1.1. Del *datum* al dato informático

Históricamente, los datos –y me refiero siempre a este concepto desde el punto de vista de las Ciencias de la Computación, como una representación simbólica y codificada de un hecho

[1] La bibliografía, en este sentido, es muy amplia, basta de momento recordar la observación de Lisa Gitelman que defiende que los humanistas deben interesarse por los datos: "In some sense, data are precisely *not* the domain of humanistic inquiry. Yet we propose that students and scholars in the humanities do worry about data, broadly speaking, to the extent that they worry about how their objects of study have been assumed as well as discerned" (Gitelman, 2013, pp. 3-4). Y, por otro lado, las afirmaciones ofrecidas por Christine Borgman (2007) según la cual los datos se encuentran en la base de cualquier actividad científica: "Data [...] are the foundation of scholarship" (p. 115), aunque su definición no sea siempre unánime en las diferentes disciplinas: "Data rarely are *things* at all. [...] Rather, data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship. Those representations vary by scholar, circumstances, and over time. Across the sciences, social sciences, and the humanities, scholars create, use, analyze, and interpret data, often without agreeing on what those data are" (Borgman, 2015, p. xviii).

o valor empírico– no han sido parte ni del discurso humanístico ni –en la mayoría de los casos– de su metodología investigadora. Aún así, como se ha venido señalando (Gitelman, 2013, pp. 3-4) y como pretendo poner de relieve en este artículo, me parece una cuestión esencial entender el papel y la relevancia que los datos pueden jugar en nuestro quehacer humanístico, pues es preciso que comprendamos de qué manera nuestros objetos de estudio son interpretados informáticamente y se convierten en algo potencialmente procesable por una computadora.

Para entender la evolución de cualquier término es esencial reconstruir su historia. Por más que se haya ya repetido hasta la saciedad la etimología de la palabra *dato*, tanto su origen como sus primeros usos ayudan a captar mejor su sentido actual. La palabra procede del latín *datum*, neutro de *datus* (dado), participio pasivo de *dare* (dar). Su significado está claro: dado, algo dado, regalo, lo que se da. Las definiciones proporcionadas en diferentes diccionarios históricos se relacionan con dos conceptos básicos: el de antecedente y el de documento. Gómez de Silva (1988) define el término *dato* como "antecedente, hecho dado o medido que se usa para llegar a una conclusión o tomar una decisión" (p. 209); Sandoval de la Maza (1995) insiste en la misma idea: "antecedente necesario para llegar al conocimiento exacto de una cosa o para deducir las consecuencias legítimas de un hecho" (p. 246). En su diccionario etimológico y bajo la entrada *dar*, Joan Coromines (1954) ofrece para el término *dato* únicamente la acepción de *informe*, *testimonio* (p. 109). Unas líneas más arriba, aparece otro término: data, en el sentido de fecha, usado por primera vez en la *Historia general de España* de Juan de Mariana (1601). Ambos términos están estrechamente relacionados, de manera especial si pensamos en otras lenguas como el inglés, donde tenemos la diferencia *date-data*. Coromines señala que *data* procede del bajo latín, y va "referido a *charta* 'documento', participio de *dare* 'dar', en el sentido de 'extendido, otorgado', palabra que en las escrituras latinas precede inmediatamente a la indicación del lugar y fecha" (p. 109). Vemos, pues, que, por un lado, Coromines aclara el aspecto más material y textual del término (en cuanto documento y testimonio), y, por el otro, otras definiciones aducen el sentido de antecedente y hecho concreto. Además, el término aparece conectado a otros que apuntan hacia la misma dirección de la concretización, como el caso de data, es decir, hacia conceptos como fecha, número o lugar.

Una consulta al *Corpus Diacrónico del Español* (CORDE) arroja referencias interesantes en cuanto a las primeras apariciones de la palabra *dato*[2]. Así, vemos que antes del 1700 los ejemplos tanto en su forma singular como en plural son realmente escasos. En singular por lo general las ocurrencias corresponden a su forma en latín o italiano[3], aunque la primera parece

[2] *Corpus Diacrónico del Español* (CORDE), accesible desde: http://corpus.rae.es/cordenet.html.

[3] Otras ocurrencias son en realidad errores de transcripción que no han sido corregidos: en singular, aparece, por ejemplo, en lugar de la forma *daño*; y en plural, a veces corresponde a la forma *daros*

encontrarse en la *Varia fortuna del soldado Píndaro* de Gonzalo de Céspedes y Meneses escrita en el año 1626. El fragmento se sitúa en un contexto judicial y hace referencia –siguiendo la definición de Coromines– a los documentos o pruebas materiales a manos de un letrado: "Mas como se trocaron los *datos* con la venida del juez, y éste procedía aora con tantas extorsiones, mudó consejo"[4]. Pocas o ninguna otra aparición con este mismo significado se encuentra hasta el 1763, en plena Ilustración: "Estos son los *datos* de cuyo conocimiento depende la solución del gran problema que nos ocupa" (*Carta a Rodríguez Campomanes* de Benito Bails. Ed. M. Avilés, 1983. CORDE). Es decir, *datos* en el sentido de pruebas o evidencias. En la misma línea se sitúa otro fragmento de Pedro Rodríguez Campomanes que incluso los cuantifica:

> La segunda diligencia es leer la obra anónima por dentro y ver los autores o hechos que cita y de ahí se infiere que el autor escribió después de ellos y necesariamente ha de ser posterior a lo menos contemporáneo; y ya se tienen dos datos dentro de los cuales se ha de buscar la época cierta del escritor (Carta a José Ruete. Ed. M. Avilés, 1983. CORDE).

Ya a finales de siglo aparecen otras obras, como el *Diario de viaje* (1789-1794) de Francisco Xavier de Viana, en que utiliza abundantemente la palabra *datos* para referirse a coordenadas geográficas[5]. Otra acepción de estos mismos años hace referencia a los precios de diversos materiales[6]. En el siglo siguiente, el significado se irá ampliado a cualquier tipo de información, y avanza hacia naturalezas menos concretas y más abstractas. Cabe decir también que muchas de las ocurrencias van acompañadas de adjetivos tales como *datos seguros*, o *datos ciertos*, insistiendo en el carácter objetivo del término.

La palabra *dato*, aplicada a las Ciencias de la Computación, es obviamente un producto del siglo XX, pero la idea subyacente y el uso es mucho más antiguo. Entre las primeras apariciones del término encontramos referencias al documento físico, en tanto que prueba, y a las coordenadas geográficas o precios de mercado. En cierto sentido, remontándonos al origen de la palabra, volvemos en realidad al mismo significado que tiene hoy en día en el campo de las HD. El valor concreto del término se observa en estos usos tempranos hasta al menos finales del siglo XVIII, y va aumentando su uso y ampliando su significado a lo largo de la centuria siguiente.

(infinitivo acompañado del pronombre de segunda plural), a una fruta tropical (Gonzalo Fernández de Oviedo, *Historia general y natural de las Indias*, 1535-1557, II, 331. CORDE), o bien a *datu* o *dato* como título de jefe o monarca de Filipinas. También se da el caso que el término sea parte del texto de la introducción moderna del texto propiamente dicho.

[4] El CORDE ofrece la edición de Arsenio Pacheco, Espasa-Calpe (Madrid), 1975; en la edición del 1733 (p. 28) aparece la forma *dados* (*sic*).

[5] Entre los múltiples ejemplos, considérese: "Al mediodía aunque hubiese aún bastante niebla pudimos sin embargo observar la latitud de 35°45' y longitud de 48°04'00", con cuyos *datos* situados en el plano de Taforo, nos demoraba aquella isla al N. 59° O. distancia 29 millas" [Francisco Xavier de Viana, *Diario de viaje* (1789-1794), I. Ed. Sofía Corchs Quintela, Biblioteca Artigas (Montevideo), 1958, I, p. 31].

[6] Ejemplos de ello se encuentran en el mismo CORDE: Luis Proust, *Anales del Real Laboratorio de Química de Segovia* (1791), Ed. Imprenta Antonio Espinosa, Segovia, p. 59, 88, 124, 126.

En el caso de la lengua inglesa, Daniel Rosenberg (2013) traza una historia del término *data* para entender de qué manera se fue transformando. Al igual que el español, aunque de manera más precoz, se empezó a utilizar en el siglo XVII y se naturalizó en el siglo siguiente. Según este autor, sus primeros usos están en clara conexión con el desarrollo de conceptos modernos de conocimiento y de argumentación. Su primera aparición corresponde al 1646 en el *Oxford English Dictionary*, y sus primeros usos tratan especialmente hechos bíblicos, teológicos, o simples hechos cronológicos (p. 18).

En cuanto a la evolución del término en español, algunas herramientas ofrecen una panorámica interesante, como el buscador en línea Google NGram Viewer que genera un gráfico de frecuencia contabilizando las fuentes digitalizadas por Google[7]. En las dos siguientes imágenes podemos ver, por un lado, la evolución del término *datos* desde 1800 hasta el 2000, por el otro, la combinación de diferentes tipos de datos (históricos, biográficos, científicos, informáticos).

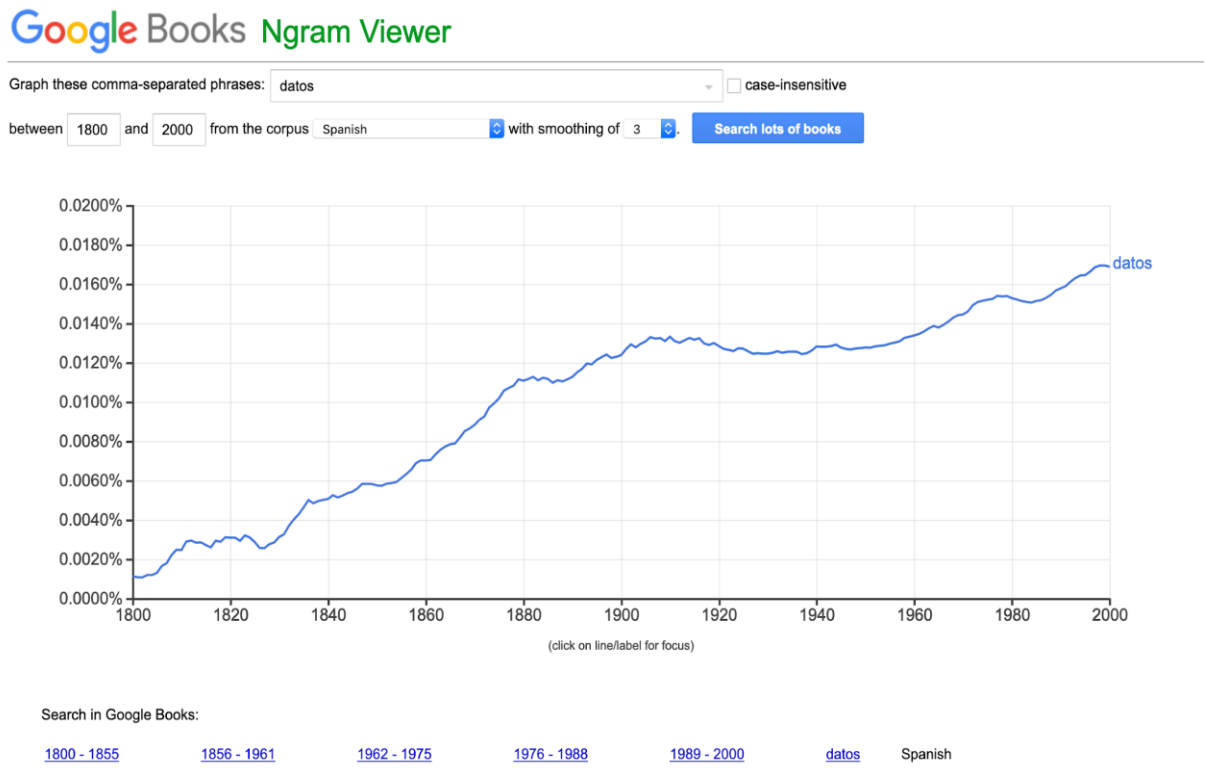

Figura 1. Gráfico de frecuencias de la palabra datos. Fuente: NGram Viewer.

[7] Accesible desde: https://books.google.com/ngrams. El problema de esta herramienta, como se ha señalado, consiste en el no saber exactamente qué hay en ese corpus y con qué criterios se ha construido; los resultados pueden no corresponderse a la realidad, especialmente teniendo en cuenta que la digitalización de Google del patrimonio escrito sigue siendo solo parcial. Sin embargo, actualmente ofrece la posibilidad de descargar la lista de obras del corpus textual, hecho que lo hace más transparente.

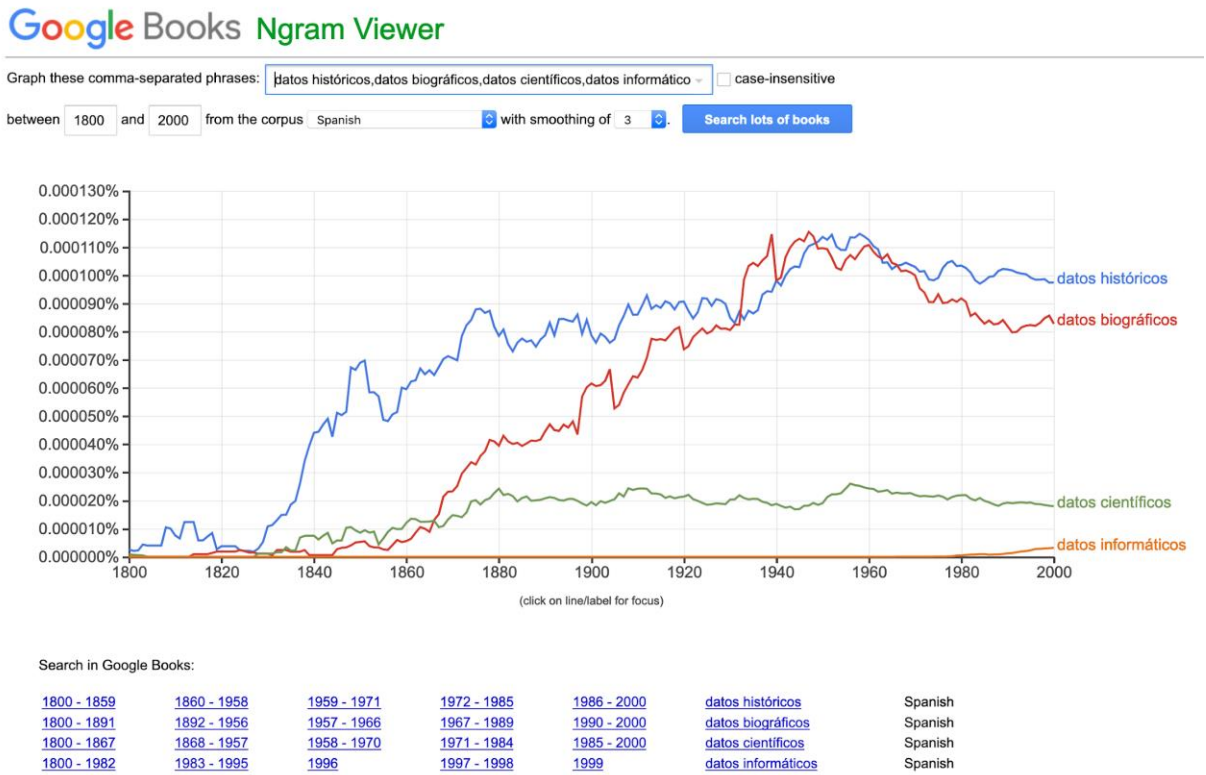

Figura 2. Gráfico de frecuencias de los términos *datos históricos*, *datos biográficos*, *datos científicos*, datos *informáticos*. Fuente: NGram Viewer.

El término, como se ha señalado, empieza a tener un uso creciente durante el siglo XIX y despunta especialmente a partir de la década de los noventa. Además, es interesante ver cómo, mientras que *datos científicos* se mantiene constante desde 1880 hasta el siglo XXI, la combinación *datos históricos* y *datos biográficos* tienen un mayor uso en la década de los cincuenta. *Datos informáticos*, en cambio, aparece sólo, y por obvias razones, a partir de los últimos años de los ochenta y durante toda la década siguiente.

1.2. Las Humanidades Digitales y la materialidad de la información

Los datos, en tanto que expresión informática, y en general el componente computacional, juegan un papel central en las HD, y en cualquier investigación que abrace metodologías digitales. De entre las miles de definiciones que ha habido sobre esta disciplina[8], una de las más genéricas y acertadas, creo, sigue siendo la propuesta por Melissa Terras. Esta concibe las HD como la intersección entre las tecnologías digitales y las Humanidades, cuyo objetivo es la de producir y usar aplicaciones y modelos que hagan posible nuevos modos de enseñanza e investigación, incluyendo tanto las Humanidades como las Ciencias de la Computación. Además, –y esta tendencia ha sido especialmente más relevante en zonas de Latinoamérica y Caribe (del Rio Riande, 2015; 2016; 2018)– las HD se han centrado también

[8] El ya conocido sitio web https://whatisdigitalhumanities.com recoge actualmente 817 diferentes definiciones entre 2009 y 2014.

en estudiar el impacto de estas tecnologías en el campo del patrimonio cultural, las instituciones, las bibliotecas, los archivos y la cultura digital (Terras, 2012)[9].

Especialmente desde mediados de los años 2000, las HD emergieron como un campo conscientemente innovador respecto a otras disciplinas tradicionales, no sólo desde el punto de vista metodológico, sino también en tanto que comunidad académica y científica. Así, al cabo de una década, Lisa Spiro (2012) se aventuró a proponer una serie de valores genéricos que son aplicables en ámbito académico y profesional, insistiendo especialmente en la idea de comunidad. Estos valores eran: acceso abierto, colaboración, compañerismo, conectividad, diversidad y experimentación[10]. Además, Spiro (2011) había señalado que otro de los objetivos de las HD debía ser la mejora de la enseñanza y el aprendizaje, así como el compromiso social (*public engagement*) que en años recientes ha tenido su plasmación en lo que se conoce como *public humanities*, en la línea de iniciativas como la ciencia abierta y ciudadana. En fin –y reconduciéndonos al tema que nos ocupa– las HD eran y son –según Spiro (2011)– las encargadas de hacer posible un mayor acceso a la información cultural a través de la recolección, la gestión, la manipulación, la modelización y la explotación de los datos.

Quisiera insistir en esta última misión y subrayar el papel central que los datos han tenido desde los inicios de lo que venimos conociendo como HD. Originariamente, la adopción por parte de los humanistas de metodologías propias de las Ciencias de la Computación respondió a una cuestión de necesidad y efectividad, a diferencia de lo que sucede hoy en día en que a menudo se entrevera con una cuestión ideológica como las que señalaba Spiro. El gran desafío, en un primer momento, consistió en transformar el tema de estudio en datos informáticos procesables por las computadoras. Pongamos por caso la famosa historia de Roberto Busa[11], ¿qué es lo que buscaba este cura erudito que algunos han considerado uno de los fundadores de las HD? Pues básicamente poder llevar a cabo búsquedas concretas al interno de un corpus ingente –en su caso– de Tomás de Aquino. De ahí salió a principios de 1950 el *Index Thomisticum*[12], que no era más que una modernización de lo que se había hecho desde la Edad Media, las famosas concordancias. IBM (International Business Machines Corporation) ofreció ayuda a Busa, no tanto por su interés en el Aquinate, sino porque comprendió que la elaboración de un software de concordancias abría nuevas posibilidades para explorar las fuentes textuales. A partir de ahí y hasta los años setenta, además de la aparición de otros

[9] La definición propuesta en inglés es la siguiente: "Digital Humanities research and teaching takes place at the intersection of digital technologies and humanities. DH aims to produce and use applications and models that make possible new kinds of teaching and research, both in the humanities and in computer science (and its allied technologies). DH also studies the impact of these techniques on cultural heritage, memory institutions, libraries, archives and digital culture" (Terras, 2012).

[10] Los términos utilizados en inglés son: *openness, collaboration, collegiality and connectedness, diversity, experimentation* (Spiro, 2012).

[11] Para una historia de las HD, véase Hockey (2008), Vanhoutte (2013), y para la figura de Roberto Busa, la monografía de Steven Jones (2016).

[12] Proyecto hoy todavía en vigor, accesible desde: http://www.corpusthomisticum.org/it/index.age.

programas de concordancias, como COCOA (acrónimo de COunt and COncordance Generation), emergieron nuevos métodos de análisis cuantitativos sobre el estilo de los autores (campo que hoy en día llamamos estilometría), y las primeras iniciativas de traducción automática (*Translation Machine*) (Vanhoutte, 2013, pp. 122-123). Todas estas iniciativas pueden ser reconducidas a la necesidad de transformar los textos analógicos en datos discretos y poder procesarlos de una manera más ágil con la computadora. La primera ingesta de datos se centró en los textos, y con ello disciplinas que dirigían su atención hacia este soporte, como la lingüística de corpus o la filología, fueron las pioneras del campo. Es más, se popularizó especialmente entre aquellas disciplinas históricas acostumbradas a trabajar con el pasado a través de lo único tangible que les quedaba: los textos.

La década de los setenta, con la aparición de asociaciones internacionales, la celebración de conferencias, y la publicación de revistas científicas, como *Literary and Linguistic Computing*, el énfasis se trasladó al archivado y la preservación electrónica de los textos, donde surgieron proyectos insignia como *Oxford Text Archive* (OTA, 1976), o el *Thesaurus Linguae Graecae* (TKG). La década siguiente vivió la revolución del ordenador personal y la interfaz gráfica (Apple vende su primera computadora en 1976) que cambia por completo la experiencia de usuario. Las listas de concordancias son reemplazadas por las búsquedas gracias al sistema operativo DOS. En 1987 se celebra el primer congreso celebrado en Vasar College (Pughkeepsie, Nueva York) sobre la Text Encoding Initiative (TEI), un sistema de codificación de textos que planteaba un estándar de intercambio de textos, en ese entonces basado en SGML (Standard Generalized Markup Language) (Ide y Sperberg-McQueen, 1995)[13].

A lo largo de los años noventa, llegó internet y los exploradores como Mosaic (1993). Los intereses humanísticos empezaron a abrazar la variedad de objetos de estudio con los que venían trabajando desde hacía décadas: los historiadores del arte empezaron a manipular imágenes digitales; la musicología y los estudios de cine, objetos multimedia; los historiadores y geógrafos los GIS (Geographic Information System; ArcGIS aparece en 1999); los filólogos empiezan a hacer sus primeros pinitos con las ediciones digitales. El punto crucial en la historia de las HD se produjo, como es sabido, en el año 2004 con la publicación de *A Companion to Digital Humanities* (Schreibman et al., 2004). El nombre que se había utilizado en el mundo anglosajón hasta ese momento para designar la disciplina en ciernes había sido *Humanities Computing*, que, al no parecer muy atractivo, se rebautizó como *Digital Humanities*. Esta nueva etiqueta ponía el énfasis en *Humanities* y no en *Computing* y la hacía seguramente más amigable para los humanistas más tradicionales (Kirschenbaum, 2010, pp. 55-56).

[13] El texto puede ser todavía consultado en línea: "Design Principles for the Text Encoding Guidelines", TEI ED P1, 14 Diciembre 1988. Accesible desde: https://tei-c.org/Vault/ED/edp01.htm.

La misma historia oficial de las HD, pues, pone en evidencia el papel central de los datos desde los albores de la aplicación de las Ciencias de la Computación en las Humanidades. Hoy en día se maneja una gran diversidad de métodos cuantitativos que no podrían ser concebidos sin un enfoque basado en los datos. Y aunque el desafío de transformar nuestros objetos de estudio se haya ya materializado en métodos y proyectos de valor indiscutible, falta a nivel global una alfabetización, especialmente entre los humanistas, del cómo plasmar las investigaciones en algo procesable por nuestros ordenadores que enriquezca y que no sea un mero complemento prescindible.

2. LA NATURALEZA DE LOS DATOS

2.1. La quimera de las clasificaciones

Antes de adentrarme en cuestiones más prosaicas sobre qué son y qué tipos de datos predominan en las HD, me gustaría abordar el tema desde una perspectiva más bien literaria y filosófica. Concebir nuestras investigaciones humanísticas en términos de datos no es siempre tarea fácil y requiere, ante todo, un esfuerzo de estructuración y clasificación. A este propósito, se ha traído a colación en alguna ocasión (Sperberg-McQueen, 2004) un famoso relato de Jorge Luis Borges, titulado "El idioma analítico de John Wilkins" (1952), donde el protagonista acometía la empresa de categorizar el universo:

> También había propuesto la formación de un idioma análogo, general, que organizara y abarcara todos los pensamientos humanos. [...]
> Dividió el universo en cuarenta categorías o géneros, subdivisibles luego en diferencias, subdivisibles a su vez en especies. Asignó a cada género un monosílabo de dos letras; a cada diferencia, una consonante; a cada especie, una vocal. Por ejemplo: *de*, quiere decir elemento; *deb*, el primero de los elementos, el fuego; *deba*, una porción del elemento del fuego, una llama (pp. 122-123).

Así, Borges explica las diferentes categorías que Wilkins establece para el universo en general: cómo clasifica las piedras, los metales, o la belleza. El autor, cómicamente, evidencia lo ambiguo, redundante y deficiente que este sistema puede parecer, y a modo de comparación recuerda las objeciones que un tal Franz Kuhn hizo a propósito de una enciclopedia china llamada *Emporio celestial de conocimientos benévolos*, donde se establecía una clasificación particular de los animales:

> En sus remotas páginas está escrito que los animales se dividen en (a) pertenecientes al Emperador, (b) embalsamados, (c) amaestrados, (d) lechones, (e) sirenas, (f) fabulosos, (g) perros sueltos, (h) incluidos en esta clasificación, (i) que se agitan como locos, (j) innumerables, (k) dibujados con un pincel finísimo de pelo de camello, (l) etcétera, (m) que acaban de romper el jarrón, (n) que de lejos parecen moscas (Borges, 1952, pp. 123-124).

Añadía además que toda clasificación es arbitraria pero necesaria e inalienable del ser humano, o como dice Borges:

> notoriamente no hay clasificación del universo que no sea arbitraria y conjetural. La razón es muy simple: no sabemos qué cosa es el universo.
> [...] La imposibilidad de penetrar el esquema divino del universo, no puede, sin embargo, disuadirnos de planear esquemas humanos, aunque nos conste que estos son provisorios (p. 124).

Lo que hoy nos parece una genialidad artificiosa más del argentino es, a su vez, un buen punto de partida para dar a entender algunas ideas básicas entorno a nuestra actividad digital. En primer lugar, la arbitrariedad del símbolo y el hecho de que no toda equivalencia responde necesariamente a una causa natural. Se trata de un proceso por el que decidimos que un objeto, un atributo o un valor será representado por un símbolo fruto de la abstracción, así por ejemplo, el mismo código binario –con el que toda computadora opera en su nivel más profundo aunque es invisible a nivel de usuario–, a través del cual nuestros números y letras pueden ser entendidos por la computadora: y así, la letra *a* corresponde a 01100001, o la palabra *hola* a 01101000 01101111 01101100 01100001[14]. En segundo lugar, que el género humano –aunque consciente de la complejidad– tiende instintivamente a organizar sus conocimientos en categorías, a crear sistemas de clasificación que son, en el fondo, modelos de conocimiento. Establecer categorías, equivalencias, jerarquías y estructuras entre los conceptos es también una de las dinámicas intrínsecas de las Ciencias de la Computación. La estructuración de los datos, de hecho, es indispensable para el procesamiento informático. Piénsese en cómo se organizan los archivos en una página web, en los lenguajes de programación, en los lenguajes de metadatos, en la creación de ontologías y taxonomías para la web semántica, o en la creación de esquemas para el modelado de datos, entre un sinfín más de casos. Todos requieren un esfuerzo de clasificación. El reto del humanista, en este sentido, es convertir esa clasificación en términos informáticos, ya sea creando una representación individual (como en la concepción de una base de datos relacional) o siguiendo las guías de algún estándar (por ejemplo, al codificar un texto según las guías directrices de la Text Encoding Initiative[15]).

Este texto borgiano sirvió de inspiración a Michel Foucault para llevar a cabo *Les mots et les choses* (1966). El prefacio a su obra inicia con una reflexión sobre la taxonomía propuesta por la mencionada enciclopedia china, y el encanto exótico de un pensamiento que es diferente del nuestro, y que demuestra simplemente la imposibilidad de concebirla. Pero Foucault puntualiza que no es solo el hecho de enumerar esos elementos aparentemente arbitrarios lo que atrae nuestra atención, sino la relación que se establece entre los mismos (p. 8). Por lo general, según Foucault, toda taxonomía viene determinada por la solidez y la evidencia que

[14] Equivalencias extraídas de: https://www.qbit.it/lab/bintext.php.
[15] Accesibles desde: https://tei-c.org/guidelines/.

garantizan la posibilidad de una yuxtaposición (p. 8). Esta yuxtaposición no es siempre clara y viene determinada por una compleja conexión de orden que se establece entre las cosas. Por ejemplo, Foucault utiliza el símil de la mesa de operaciones médicas donde se dejan todos los instrumentos: puede ser que esos utensilios no tengan a primera vista ninguna conexión, pero lo que los une es su utilidad para llevar a cabo la operación. En toda cultura –señala Foucault– entre el uso de lo que uno podría llamar código ordenador y las reflexiones sobre el orden, se encuentra la experiencia pura del orden y de las maneras de ser. De ello se deduce que ningún sistema de clasificación es neutro, objetivo o evidente de por sí, y todos dependen de un determinado contexto cultural o ideológico. A Foucault le atraía el saber cómo y por qué, en el marco del espacio epistemológico de una cultura, ciertas configuraciones han dado lugar a formas diferentes del conocimiento empírico, el "socle positif des connaissances" (1966, p. 13); por ello hablaba no tanto de una historia del conocimiento, sino de una arqueología de las Ciencias Humanas, porque se remontaba al proceso previo de la plasmación de los saberes. El objetivo era, en definitiva, conocer qué mecanismos han llevado a ciertas culturas a establecer un orden determinado entre las cosas.

Esta reflexión puede ser trasladada al acto de imaginación de los datos que es siempre un acto de clasificación, de agrupamiento, división, y jerarquización, a través de unos determinados principios subyacentes que a veces pueden ser difíciles de recuperar (Gitelman, 2013, pp. 8-9). Un ejemplo muy claro lo podemos ver en el caso de los metadatos, como aquellos utilizados por las bibliotecas. ¿A qué responde la clasificación de un determinado objeto? Todo bibliotecario sabe cuáles son las características que debe recoger a la hora de catalogar un documento (autor, título, año, etc.). Toda clasificación consiste en la identificación de las propiedades del objeto que se está clasificando, y por tanto cuanto mejor se conozca ese objeto, más precisa será la clasificación. De ahí que deba ser el humanista que deba afrontar este reto, y no otra persona. Desafortunadamente, no hay ningún esquema de clasificación que sea completamente exhaustivo, ni una sola manera correcta de clasificar un objeto (Sperberg-McQueen, 2004), pero eso, como decía Borges, no puede disuadirnos de proponer una clasificación.

2.2. Definiciones, tipos, estructuras, organización y naturaleza

Pasemos ahora a establecer a qué nos referimos cuando hablamos de datos. Christine Borgman (2007, pp. 119-120) recoge en un capítulo titulado "Data: Input and Output of Scholarship" diferentes definiciones que nos pueden ser de ayuda. La primera corresponde al manual de referencia del *Open Archival Information System* (OAIS), un estándar que permite el intercambio de informaciones entre archivos de acceso abierto, donde el dato es definido como una representación de la información susceptible de ser reinterpretada por la computadora a través de un sistema formalizado para poder ser comunicada, interpretada y procesada:

> A reinterpretable representation of information in a formalized manner suitable for communication, interpretation or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking or a moon rock specimen (CCSDS, 2012, pp. 1-10).

El dato es pues en primer lugar la expresión mínima digital y la representación simbólica de un objeto, atributo o valor empírico que se plasma de manera digital a través de una anotación binaria de 0 y 1, en datos discretos y no ya analógicos.

En términos muy genéricos, es útil también una segunda definición proporcionada en 1999 por el National Research Council (NRC) que divide los datos en hechos, números, letras o símbolos que representan un objeto, una idea, una condición, una situación o algún otro factor:

> *Data* are facts, numbers, letters, and symbols that describe an object, idea, condition, situation or other factors. A data element is the smallest unit of information to which reference is made. [...] For purposes of this report the terms *data* and *facts* are treated interchangeably, as is the case in legal contexts (p. 15).

Aquí el problema es saber exactamente qué entendemos por hecho (*fact*), y la naturaleza potencialmente ambigua del término[16].

Los datos –según el NRC (1999, p. 15)– pueden ser textuales (tal sería el caso del texto plano o el texto que podría aparecer en bases de datos como bibliografías, directorios, diccionarios), numéricos (ej. propiedades concretas, estadísticas, valores), imagen, video (imágenes fijas o con movimiento, películas), sonido o audio. Cada uno de estos datos requiere unos procesos y programas específicos.

Los datos además pueden tener estructuras diferentes. Como señala Christof Schöch (2013), en su artículo sobre la naturaleza de los datos en Humanidades, los casos más frecuentes son las estructuras lineales, como pueden ser las hojas de cálculo de Google Sheet[17]; las estructuras jerárquicas, como el caso del documento en XML; o bien pueden tener una estructura multi-relacional, como en el caso de los nodos de una red, donde se utilizan softwares como Gephi[18].

Los datos también pueden organizarse de maneras concretas. Por un lado, podemos tener, por ejemplo, simples textos planos en cuyo caso se trataría de datos no estructurados; por el otro, se pueden mencionar los documentos XML, en que la información ha sido codificada con una cierta estructura, anidada y jerárquica, y los datos aparecen semi-estructurados. Por último, y seguramente el más difundido, son los datos estructurados en bases de datos, entre las que se

[16] Rosenberg establece una distinción interesante entre la naturaleza de los términos *data*, *fact* y *evidence*. Los hechos son ontológicos, porque son o no son, existen o han existido; las evidencias son epistemológicas porque construyen un conocimiento, mientras que los datos son retóricos por naturaleza, porque ayudan a elaborar un argumento: "there are important distinctions here: facts are ontological, evidence is epistemological, data is rhetorical" (2013, p. 17).
[17] Accesible desde: https://www.google.com/sheets/about.
[18] Accesible desde: https://gephi.org.

encuentran el tipo SQL o bases de datos relacionales[19]. Debe señalarse que estas estructuras no son excluyentes, por ejemplo, un corpus textual constituido por textos planos y en archivos separados, puede luego formar parte de una base de datos, particularmente a partir de sus metadatos.

En fin, es importante tener en cuenta la naturaleza, y establecer las diferencias entre datos y metadatos. En cualquier caso, se trata siempre de datos, pero unos constituyen la representación digital de algún aspecto del objeto de estudio, mientras que los metadatos son informaciones añadidas que lo clasifican, en inglés *data about data*. En general, con este término nos referimos siempre a los datos informáticos que recogen información sobre el contenido, la estructura o el contexto de una fuente, ya sea documento, proyecto o archivo (Baca, 2008). Los metadatos también son concebidos para poder gestionar los recursos digitales, hacer posible la interoperabilidad entre objetos similares, y, sobretodo, poder asegurar su archivado y preservación. Algunos tipos de metadatos son usados sólo en ciertas comunidades, y no pertenecen sólo a una institución. Los estándares web más conocidos son: Dublin Core Metadata Initiative; MAchine REadable Cataloging (MARC), desarrollado por y para la catalogación en las bibliotecas; Metadata Object Description Schema (MOS); Encoded Archival Description International Standard (EAD) para los archivos; Text Encoding Initiative (TEI); Resource Description Framework (RDF) para la web semántica; Metadata Encoding and Transcription Standard (METS); o el ya mencionado Open Archival Information System (OAIS).

2.3. El ecosistema de los datos y algunos debates

La dinámica que se establece al trabajar con datos es particular y el desafío del humanista consiste en convertir esos objetos en datos discretos, no analógicos, y poder procesarlos informáticamente. Schöch propone una suscinta definición sobre los datos en humanidades:

> Data in the humanities could be considered a digital, selectively constructed, machine-actionable abstraction representing some aspects of a given object of humanistic inquiry (Schöch, 2013).

Como vemos en la figura 3, en toda investigación digital los datos representan o valen por el objeto de estudio; en la mayoría de los casos, son una representación parcial, pero al mismo tiempo constituyen el único punto de acceso en términos informáticos. A esta relación desigual, se le añade, desde el punto de vista del humanista, una complejidad más, esto es, la del uso y comprensión de herramientas digitales para el procesamiento de los datos sobre nuestro objeto de estudio. La curva de aprendizaje, en este sentido, es en muchas ocasiones imponente y desafiadora.

[19] Accesible desde: https://es.wikipedia.org/wiki/MySQL.

Figura 3. Esquema del trabajo con datos en Humanidades.

De la misma manera que los científicos consiguieron traducir el lenguaje (en tanto que conjunto de símbolos) y otros procesos en código binario (Laue, 2004), los humanistas deben transformar sus objetos de investigación en algo tangible y concreto. Pero a diferencia de los datos en ciencias, en Humanidades los datos son más complejos, imbuidos con significados múltiples, cargados de identidades superpuestas y abiertos a interpretación (Levi, 2013, p. 36). Un buen ejercicio para entender esto es la comparación de una hoja de cálculo de un estudio en biología, por poner un ejemplo, donde aparecen en su mayoría solo cifras, y una de un estudio en Humanidades, donde especialmente en los primeros estadios de preparación vemos más texto e incluso algunos campos con interrogantes, opciones múltiples, paréntesis o notas varias.

Para profundizar en la naturaleza de los datos en Humanidades, es interesante la distinción propuesta por Owens (2011) según la cual pueden ser concebidos desde tres perspectivas diferentes: primero, la del artefacto o objeto manufacturado (ej. hoja de cálculo); segundo, la del texto interpretable, en tanto que objeto creado por un humano y para una audiencia, es decir, el humanista puede interpretar los datos como si fueran un trabajo de autor cuyas intenciones pueden ser dignas de consideración y exploración; tercero, la perspectiva de los datos como información procesable por computadoras para ser visualizada o manipulada a través de metodologías diferentes. De hecho, este último tipo de manipulación, puede generar nuevos objetos, artefactos o textos que pueden ser a su vez interpretados y explorados. Por ejemplo, un equipo de investigación puede poner a disposición un corpus textual que puede ser consultado y descargado en acceso abierto; a continuación, otro investigador puede utilizar

todo ese corpus o una parte (según género o autor, por ejemplo) y llevar a cabo estudios de estilometría, cuyos resultados tanto estadísticos como de visualización constituirán otro producto digital, que a su vez habrán generado otro conjunto de datos. Por lo tanto, Owens indica que los datos pueden también contener valor de evidencia (*evidentiary value*), es decir, funcionar como evidencia para apoyar una idea o un argumento.

Desde el punto de vista del artefacto u objeto manufacturado, se ha planteado el problema de la objetividad de los datos. De la misma manera que Foucault se preguntaba por los mecanismos que establecían el orden entre las cosas, la simple recopilación de los datos plantea la legitimidad de los mismos. El hecho de que empecemos nuestras investigaciones con la recolección de datos entraña a su vez algunos riesgos. Como indica Gitelman hay que proceder con cautela pues a veces damos por supuesto que al ser datos son transparentes y objetivos:

> This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we're not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity (2013, pp. 2-3).

En Humanidades a veces transformar las ideas en datos procesables significa reducir o simplificar, de ahí el escepticismo de algunos estudiosos de la literatura que sienten que se pierde el valor contextual o los matices culturales que el objeto o ese valor pueda tener. Existe ciertamente la imposibilidad de ser objetivos, cuando la objetividad es considerada relativa. Como indica Schöch, los objetos de estudio en Humanidades, como pueden ser una lengua, un texto, una pintura o una composición musical, son sistemas semióticos cuyas dimensiones van más allá de lo físicamente medible, unas dimensiones que dependen de la semántica y la pragmática, es decir, del significado de un contexto concreto. Una tal concepción parece conducirnos a un callejón sin salida, pues toda expresión digital será siempre parcial. Pero lo cierto es que toda abstracción requiere una expresión material, empezando con el lápiz y el papel, aunque en ese proceso se pierda algo. Los datos necesitan ser imaginados como tal y concebidos digitalmente, y ahí es donde el humanista agradece una mínima alfabetización computacional.

Un ejercicio no siempre fácil es el de cómo estructurar nuestra investigación en términos de una base de datos relacional, ¿qué elementos constituirían las tablas? ¿qué campos tendrá cada tabla? ¿qué tipo de valores se incluirán en cada campo? ¿cuál es la relación existente entre los elementos y las tablas? Es este un acto de creatividad y de interpretación, y lo más desafiante de todo es que no existe una sola y correcta manera de plasmarla. A otro nivel más básico y para ejemplificar las diferentes posibilidades de anotación de los datos: imaginemos, por ejemplo, las diferentes maneras de señalar una misma fecha: ¿Qué orden establecemos para indicar el año, el mes, el día?, ¿utilizamos sólo números?, ¿adoptamos el guión o la barra

inclinada?, ¿cómo manejamos fechas inciertas? o ¿cómo indicamos que un cierto autor nació, floreció o murió en la segunda mitad del siglo XIX?

Entra en juego también, en el seno de la discusión de la objetividad, una cuestión de método. La mera recolección de datos implica una serie de principios metodológicos que pueden estar conectados a una determinada teoría, a una serie de suposiciones o conjeturas. Por ejemplo, imaginemos el caso de una edición digital donde la recolección de las diferentes variantes textuales de un texto se atiene a una escuela de Crítica Textual: ¿Se recogerán todos y cada uno de los movimientos del autor para generar una edición genética?, ¿se anotarán los diferentes manuscritos y se señalarán todas las variantes para llevar a cabo una edición crítica?, o ¿se limitará a proporcionar una fiel transcripción diplomática del texto? La objetividad no es más a veces que una quimera, y los datos –citando una vez más a Gitelman (2013, p. 6)– requieren de nuestra participación, los datos nos necesitan, y en última instancia dependen de nuestra interpretación. Se ha señalado (Owens, 2011) que la producción de los datos requiere decisiones sobre el qué y el cómo recolectar y cómo codificar y anotar esa información; cada una de esas decisiones no son neutras y ofrecen una perspectiva de análisis diferente.

El informe ya mencionado del NRC (1999, p. 15) estableció una distinción entre datos brutos (*raw data*), procesados (*processed*) y verificados (*verified*). En Humanidades y en años recientes, especialmente el término *raw data* ha suscitado polémicas. Lisa Gitelman, en su monografía *Raw Data is an Oxymoron*, retomaba una metáfora culinaria, señalando que los datos en Ciencias Sociales y Humanidades se sirven siempre cocinados, y nunca enteramente crudos[20]. Los datos no nos son casi nunca dados, sino que son generados, incluso manufacturados, normalmente por personas. Lev Manovich insistía en esta idea: "However, the passive/active distinction is not quite accurate since data does not just exist –it has to be generated" (2001, p. 224). Por ello, Johanna Drucker propuso el término *capta*, lo que se ha recogido, proporcionado (Drucker, 2011).

Al adentrarnos en las problemáticas que rodean los datos en las Humanidades y su anotación material, no es solo el de la objetividad de la información, sino también otras cuestiones que merecen mencionarse. En primer lugar, ciertos conceptos humanísticos son más complejos de plasmar que en las disciplinas de ciencias. Algunas veces la naturaleza de ciertos conceptos es más vaga y los objetos con los que trabaja el humanista (recursos textuales, visuales, audio) no siempre se prestan a abstracciones geométricas o matemáticas (Presner y Shepard, 2016). Además, los datos en HD tienen una dimensión interpretativa que es inherente a nuestra disciplina y que plantea retos a la hora de la anotación. De hecho, Drucker señala que los documentos o objetos en humanidades no son ni continen datos de por sí, "they have to be

[20] "Data are always already 'cooked' and never entirely 'raw'" (Gitelman, 2013, p. 2). Traducción propia.

remediated to become 'data' –quantified and discrete information units" (2015, p. 245), y ahí reside el desafío. Por ejemplo, ¿cómo se cuantifica o se materializa en datos el estudio de la comicidad de una obra?, ¿cómo anotamos y detectamos la ironía de un texto? Como decíamos, en algunos casos, la reducción de ciertos fenómenos a datos empíricos puede obviar otras variables que el investigador ha decidido no recoger. Como señala Schöch (2013), esto ha llevado a un cierto recelo ante el aparente empirismo de la investigación basada en datos, pues al fin y al cabo obvia el contexto del objeto y su interpretación. Así pues algunos investigadores en estudios literarios y culturales se resisten a hablar de datos cuando conciben sus temas de estudio, porque sienten que reducir nuestras investigaciones a datos informáticos es una simplificación, una banalización de nuestra disciplina, y que se pierde la dimensión interpretativa. El contraargumento a esta crítica consiste en que en el terreno de las HD, cada uno de los datos que recogemos pone a su vez dilemas de interpretación diferentes, y por tanto la dimensión interpretativa nunca está ausente. En cuanto a la pérdida de información y a la simplificación, yo diría que no existe investigación completa y que todo trabajo es siempre fruto de una selección, de una delimitación de nuestro campo de estudio. En alguna ocasión he oído también que las HD parecen esconder un cierto fetichismo por lo material, y que en realidad encubren una falta de base teórica[21]. A eso respondería que no se trata de fetichismo, sino más bien de la necesidad de responder a la lógica informática, es decir, del transformar el objeto de estudio en un formato procesable por la computadora siguiendo un determinado lenguaje de programación, de codificación o una concreta estructura de datos.

Otros problemas que afrontamos es que los datos en Humanidades son a menudo escasos, parciales o inexistentes, y deben crearse, hecho que requiere mucho tiempo y esfuerzo. Especialmente la información que tenemos sobre el pasado, como indica Marche (2012), es en muchas ocasiones fragmentaria, y hay todavía grandes cantidades de datos que, simplemente faltan o que no podemos identificar. Sólo por poner dos ejemplo, ¿cuántos autores no aparecen aún en los catálogos de autoridades de las principales bibliotecas o plataformas dedicadas a ello?, ¿cuántos textos no están todavía digitalizados ni disponibles en línea? A medida que aparecen más datos para la investigación en las Humanidades, se facilita nuestra investigación, pero el camino por recorrer es todavía largo.

A ello hay que añadir también el mito de la separación entre lo académico y lo técnico, que según mi experiencia tiene dos consecuencias graves. La primera es que algunos proyectos en Humanidades incluyen algún aspecto digital presionados por las convocatorias nacionales e internacionales, pues consideran que las posibilidades de ser financiados serán mayores si prometen un resultado digital. Este oportunismo digital disfraza, bajo forma de nuevas

[21] Sobre la crítica de la ausencia de base teórica, recomiendo la sección "In defense of theory" en Rojas Castro (2017, pp. 66-67).

metodologías, un desinterés considerable hacia las HD y lo que estas representan. El resultado, en muchos casos, consiste en meras páginas web elaboradas por empresas externas que conciben una base de datos sin ninguna perspectiva de interoperabilidad con los estándares utilizados por la comunidad HD, ni de continuidad e independencia por parte del equipo investigador. La segunda es que el trabajo digital todavía es concebido como una *techné*, como un trabajo técnico y mecánico, que puede hacer cualquier persona. Cada vez más, se valora el investigador en Humanidades con una formación informática, o a la inversa, pero la realidad es que no se aprecia lo suficiente la labor digital como para ser considerada una actividad científica.

Personalmente, creo que los conjuntos de datos generados por los investigadores deberían formar parte de la actividad académica y reconocerse como tal. No solo las publicaciones tradicionales deben tener el reconocimiento de la comunidad científica, sino también la producción de datos, que no es ni mucho menos una tarea mecánica o libre de implicaciones intelectuales y metodológicas. Como señala Borgman (2007), la creación y publicación de los datos debería reconocerse como un resultado más de nuestra actividad investigadora y formar parte de un ecosistema académico digital sostenible "data are outputs of research, inputs to scholarly publications, and inputs to subsequent research and learning. Thus they are the foundation of scholarship" (p. 115).

3. BIG DATA

Con la aparición del fenómeno del big data hacia 2011 se han abierto múltiples debates que han tocado sobretodo industrias como la médica, la científica, las aseguradoras, la investigación o la economía. Muchas de las preguntas que se plantearon al inicio siguen todavía abiertas, y muchos hablan ya de la era del big data. En múltiples ocasiones, big data ha sido considerado una revolución tecnológica y como tal también un fenómeno cultural con una dimensión social (Barlow, 2013, p. 1; Levi, 2013, p. 35): no solo se están generando nuevas empresas, nuevos perfiles de trabajo, sino también una ingente cantidad de datos sobre las personas, que plantean problemas de privacidad, nuevas maneras de gestión y de control social y, en consecuencia, de su impacto sobre la sociedad. Como no podía ser de otra manera, el término también se ha trasladado al terreno de las Humanidades y ha generado el debate de si podemos o no hablar de big data en este campo. Veamos pues, en este último apartado, qué se entiende por este término y algunas de las diferentes propuestas para su adopción.

A principios de los 2000, Meta Group publicó un breve informe (Laney, 2001) que, aunque completamente ajeno a las disciplinas humanísticas y a los datos que en ellas se producen, establecía que el comercio electrónico había planteado desafíos en la gestión de los datos (lo que él llamó *3D Data Management*) en tres dimensiones diferentes (3v): volumen, velocidad y variedad. A medida que avanzamos y producimos contenidos digitales, el volumen

de los datos que deben manejarse está creciendo exponencialmente, y en consecuencia se necesitan softwares con mucha más capacidad de procesamiento y especialmente de almacenamiento; a más volumen, se necesita también más velocidad de procesamiento (en la respuesta de las páginas web, en el análisis de la disponibilidad del inventario, el tiempo de ejecución de las compras, etc.); por lo cual, señala que no hay una barrera mayor para una gestión de los datos efectiva que la variedad de formatos incompatibles, de estructuras no alineadas, y semánticas inconsistentes. Posteriormente a la aparición del término big data, IBM (2013) añadió una cuarta dimensión: *veracity*, en tanto que los datos deben ser fiables y ciertos.

De entre las muchas definiciones que se han dado sobre big data[22], algunas insisten en el carácter desestructurado de los datos (en contraposición a las bases de datos relacionales) que pueden ser extraídos, por ejemplo, de las redes sociales, blogs, redes de sensores, imágenes, etc. Otras emfatizan la mayor capacidad de procesamiento del hardware y lo asocian a nuevas tendencias como el *machine learning* y la inteligencia artificial. Obviamente, a medida que van creándose mayores sistemas de datos se requiere que las máquinas aprendan y puedan predecir. Ward y Barker conectan el término big data con el de *data analytics* y al descubrimiento de significado a través de los datos, y además lo asocian a ciertas tecnologías (NoSQL, Apache Hadoop). También se vincula inevitablemente a ciertas empresas que generan colosales cantidades de información a través de sus usuarios, empresas tales como Google, Oracle, IBM, Microsoft, Facebook o Amazon.

Ward y Barker señalan tres factores críticos a la hora de saber cuándo estamos trabajando con big data: tamaño (*size*), complejidad de las estructura de los datos (*complexity*), y el uso de ciertas tecnologías (*technologies*), herramientas y técnicas que son útiles para procesar un conjunto de datos de grandes y complejas características. Por ello, proponen la siguiente definición: "big data is a term describing the storage and analysis of large and/or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning" (Ward y Barker, 2013).

El objetivo del big data consiste en aglutinar, almacenar, procesar la mayor cantidad de información posible; esta información es analizada de manera cuantitativa y visual con el fin de encontrar patrones, establecer normas o detectar dinámicas, y así poder predecir conductas sociales, tendencias de mercado o otros fenómenos a escala masiva. Mayer-Schönberger y Cukier, en su monografía dedicada al big data, subrayan esta dimensión predictiva de consecuencias globales:

> Big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more (Mayer-Schönberger y Cukier, 2013, p. 6).

[22] Véase para un panorama de las diferentes definiciones Ward y Barker (2013).

En cuanto al tamaño no hay mucha unanimidad, Intel ofrecía un número aproximativo de lo que puede considerarse big data, y consideraba grande aquellas organizaciones que generaban una media de 300 terabytes de datos a la semana (Ward y Barker, 2013). Y en Humanidades ¿a partir de cuándo es considerado big data? El equipo de historiadores de Shawn Graham (2015) considera big data cualquier cantidad de información que un individuo no pueda leer en un lapso razonable de tiempo y que requiera una intervención computacional para poder ser entendida[23]. En el caso de un corpus textual, por poner un ejemplo, uno pequeño no comprende más de 5 millones de tokens; mientras que un corpus mediano contiene entre cinco y quinientos millones; más allá de estas cifras se trabaja con grandes corpus. Un simple documento en Word (*.doc), de una página y unas cuatrocientas palabras, pesa 144 kilobytes (kB), ese mismo documento en texto plano pesa 2 kB. Un corpus pues de quinientos millones de tokens puede no sobrepasar los 2,38 gigabytes. Estos números dan a entender cuál es el desafío de un corpus textual para considerarse big data, especialmente si empieza a considerarse grande a partir de cientos de terabytes (1 terabyte equivale a 1.024 gigabytes).

Desde hace algunos años se ha venido discutiendo sobre el papel del big data en las Humanidades (Schöch, 2013; Levi, 2013; Kaplan, 2015; Weingart 2015, Manovich, 2016; Rojas Castro, 2017, O'Carroll 2018). Proliferan a nivel internacional convocatorias que incentivan el uso de grandes cantidades de datos[24]. Algunos investigadores señalan que en Humanidades siempre ha habido big data (Levi, 2013), mientras que otros son más pesimistas, como Rojas Castro (2017), especialmente si como parámetros se toman las 3v (volumen, velocidad y variedad) antes mencionadas. Aún así, Rojas sugiere que las HD se están aproximando a las técnicas utilizadas por el big data como puede ser minería de textos, estilometría o procesamiento natural de textos.

Uno de los grandes desafíos del big data es el hecho de que este busca encontrar tendencias generales que ayuden a encontrar nuevos patrones, que puedan ser útiles y rentables para ciertas industrias. La investigación humanística, en cambio, se ha centrado siempre en buscar la exclusividad, el individualismo, lo que sobresale. El concepto de canon es, a fin de cuentas, la reducción de ciertas obras que son consideradas mejores y más representativas. Lev Manovich (2016) también apuntaba que disciplinas como la Sociología han estado asimismo más centradas en encontrar patrones sociales generales de conducta humana, y no de aquellas individuales que se diferenciaban del flujo global; de manera análoga, lo que ahora se conoce como *cultural analytics* también se dedica a encontrar patrones a nivel global utilizando ingentes cantidades de datos. Pero señalaba que el fin último de algunas de estas

[23] "For us, big data is simply more data that you could conceivably read yourself in a reasonable amount of time –or, even more inclusively– information that requires or can be read with computational intervention to make new sense of it" (Graham, Milligan y Weingart, 2015, p. 17).

[24] Una convocatoria de las más conocidas es el *Digging into Data*, accesible desde: https://diggingintodata.org.

nuevas tendencias basadas en el big data podía también consistir en rastrear aquellos casos particulares que se diferencian de la tendencia global.

Una distinción que me parece muy pertinente en el campo de las Humanidades ha sido la propuesta por el ya mencionado Schöch (2013), que establece una distinción entre big data y smart data, abogando por la segunda categoría para las disciplinas humanísticas o, idealmente, una mezcla de ambos, pues, como dice él: "only smart big data enables intelligent quantitative methods" (Scöch, 2013). He aquí el resumen de su propuesta:

SMART DATA	*BIG DATA*
Datos (semi-)estructurados	Datos no estructurados
Datos explícitos y enriquecidos (marcado, anotaciones, metadatos)	Datos sin formato (*raw data*)
Datos limpios (*clean data*)	Datos desorganizados (*messy data*)
Pequeños (*small data*) (requiere la intervención humana y mucho tiempo)	Grandes (*big data*) (generados de manera automática)
Importancia del modelado de datos (esquemas, base de datos relacionales)	Datos NoSQL (formatos diversos)

Tabla 1. Síntesis de las diferencias propuestas en Scöch (2013).

Un ejemplo de smart data puede ser un texto en XML-TEI para una edición digital, o un corpus lingüístico. El mayor handicap de los smart data es que no son escalables, no pueden ser automatizados, porque en la mayoría de los casos debe llevarse a cabo la anotación manual por parte de los investigadores, y eso toma mucho tiempo. Un ejemplo de big data podría ser la manipulación de miles de tweets para analizar un tema en particular, o el análisis masivo de imágenes descargadas de otra red social.

Algunas de las preguntas que plantean los big data en las Humanidades seguirán abiertas mientras no aparezcan más recursos y más resultados. Trabajos recientes parecen apuntar hacia dos direcciones, por un lado, aunque no podamos hablar de big data en cuanto al volumen, sí se han adoptado técnicas propias de este fenómeno, especialmente en lo que concierne al *machine learning* y minería de textos. Por el otro, big data no debe solo mirarse desde el punto de vista más técnico sino que también debe estudiarse desde la perspectiva de la experiencia del usuario y de lo que supone en el marco de la cultura digital. Creo además que los avances hacia el uso más efectivo de los datos en Humanidades pasa por, al menos, tres variables adicionales: la actualización de los recursos disponibles, la apuesta por las tecnologías

de la web semántica, y el uso de funcionalidades que benefician la independencia del investigador/usuario, como el uso de las APIs[25].

El mayor problema al hablar de big data en Humanidades –y especialmente desde aquellos que trabajamos con materiales históricos– es justamente que todavía no existen conjuntos de datos concebidos para la aplicación de técnicas del big data. Pondré un ejemplo, en el caso de los corpus textuales, las lenguas peninsulares gozan de muchas iniciativas tanto desde el punto de vista cronológico (como la increíble labor del Hispanic Seminary of Medieval Studies[26]) o de géneros literarios (Colección de textos caballerescos hispánicos[27]). El problema es que algunas herramientas que sí podrían funcionar como big data no están preparadas todavía para ser analizadas desde una tal perspectiva. Un caso paradigmático es el de CORDE, que aunque sigue siendo la herramienta más valiosa para llevar a cabo búsquedas sobre la historia de la lengua, presenta algunos desfases desde la experiencia de usuario y desde el conjunto de datos ofrecidos: su interfaz se encuentra ya algo anticuada en cuanto a la presentación de los resultados que no pueden ser exportados en ningún tipo de formato procesable, lo que limita enormemente una consulta más personalizada.

Otras grandes iniciativas, como podría ser la Biblioteca Miguel de Cervantes, podrían constituir –como lo han hecho hasta ahora– una fuente valiosísima para descarga de datos semi-estructurados que podrían ser tratados con técnicas mencionadas de big data. En algunos de estos casos nos encontramos con problemas de derechos de autor, pues los textos, aunque son antiguos, corresponden a ediciones modernas. También hay algunos corpus que empiezan a publicarse en acceso abierto y en formatos simples para ser fácilmente procesables y analizados por técnicas y programas diferentes. Entre ellos podemos mencionar: CLiGS textbox[28] que contiene diversos corpus en español tanto en .txt como en XML-TEI; IMPACT-es diachronic corpus[29] con 8 millones de palabras, un lexicon complementario de diez mil lemas y publicado bajo una licencia Creative Commons; o el Corpus de Sonetos del Siglo de Oro[30] con anotación métrica albergado y publicado en GitHub.

En el terreno de las HD, uno de los campos más en boga desde hace ya algunos años es el de las tecnologías de la web semántica (RDF, URIs, HTTP) y los recursos ofrecidos por los

[25] Una API, correspondiente a sus siglas en inglés Application Programming Interface o, en español, Interfaz de Programación de Aplicaciones, es un conjunto de instrucciones escritas a través de un lenguaje de programación que establece el proceso o protocolo de comunicación entre diferentes sistemas. Las APIs están normalmente bien documentadas porqué definen la manera en la que una cierta aplicación (por ejemplo, Facebook, Twitter, un catálogo de biblioteca, etc.) puede ser consultada y utilizada por otras.

[26] Accesible desde: http://www.hispanicseminary.org/index-en.htm.

[27] Accesible desde: https://textred.spanport.lss.wisc.edu/chivalric/texts%20bilingue.html.

[28] Accesible desde: https://github.com/cligs/textbox.

[29] Accesible desde: https://www.digitisation.eu/tools-resources/language-resources/impact-es.

[30] Accesible desde: http://adso.gplsi.es/index.php/en/adso-project.

datos enlazados en abierto (Linked Open Data)[31]. De entre los proyectos más conocidos están los propuestos por la Fundación Wikimedia, especialmente Wikidata[32], una especie de gran base de datos enfocada en los elementos al que otorgan un identificador único, o DBpedia[33] que extrae las informaciones presentes en los artículos de Wikipedia. Actualmente algunas bibliotecas, archivos y museos han empezado su transformación, como la Biblioteca Virtual Miguel de Cervantes[34], la Biblioteca Nacional de España (BNE)[35] o el Museo del Prado[36]. Los humanistas –como propone Levi (2013, p. 36)– debemos aprender a focalizar nuestra atención a nivel de red y adoptar una visión de conjunto, menos centrada en individuos concretos, a empezar a imaginar de qué manera podemos conectar los datos que ya existen en nuestro campo de investigación.

Entre otras muchas apuestas que deberían hacerse está la de la creación de APIs para los proyectos y los datos en HD. Este modelo desplazaría, en cierta medida, el típico modelo de búsqueda tradicional, por la consulta libre de los datos albergados de un proyecto. Las APIs permiten tanto la descarga de grandes cantidades de contenidos tanto aquellos creados por los autores como por los usuarios de un sitio, y además facilitan búsquedas personalizadas que quizás ni los mismos investigadores habían previsto. Algunas de las bibliotecas que han apostado por los datos enlazados y por este sistema de consulta son la misma BNE, y a nivel internacional, el Schoenberg Institute for Manuscript Studies[37].

4. CONCLUSIONES

A lo largo de estas páginas, he intentado aproximarme al concepto de dato en las Humanidades desde perspectivas diferentes: histórica, disciplinar, técnica y crítica. Por un lado, la historia del término *dato* ofrece como primeras acepciones la idea de *antecedente*, *hecho concreto*, pero también la de *documento*, *testimonio*, *prueba*, y algunos ejemplos corresponden al documento físico, a coordenadas geográficas e incluso al precio de algunas mercancías. Este significado constituye la clave para entender el dato informático como una representación simbólica y codificada de un hecho, valor u objeto empírico. La historia proporciona, en este caso, el significado más materialista del término de la representación digital y discreta de nuestro objeto de estudio. Por otro, el papel de los datos dentro de las HD y su evolución ha

[31] Accesible desde: https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.
[32] Accesible desde: https://www.wikidata.org/wiki/Wikidata:Main_Page.
[33] Accesible desde: http://es.dbpedia.org/.
[34] Accesible desde: http://data.cervantesvirtual.com.
[35] Accesible desde: http://datos.bne.es/inicio.html.
[36] Accesible desde: https://www.museodelprado.es/modelo-semantico-digital/estandares-semanticos-y-datos-enlazados.
[37] Esta institución recibió financiación de la convocatoria *Digging into Data*, accesible desde: http://mappingmanuscriptmigrations.org y la interfaz de consulta por API se encuenta accesible desde: https://sdbm.library.upenn.edu/sparql-space.

sido el de proporcionar el acceso al patrimonio y a la información cultural; esto ha sido posible a través de una serie de etapas que implican siempre el trabajo con datos: la captura o la creación, la gestión, la manipulación, el análisis, el modelado, la publicación o el almacenamiento de los datos[38].

Desde el punto de vista informático, la estructuración de los datos es esencial para su procesamiento. Los sistemas de clasificación en los entornos digitales son omnipresentes, y abrazan desde la clasificación del sistema de archivos, hasta los modelos de los estándares de metadatos. En tanto que modelos de conocimiento, los sistemas de clasificación –como hemos visto– no son neutros o siempre objetivos y responden no solo a un determinado contexto cultural sino también ideológico o incluso disciplinar (piénsense en algunos sistemas de metadatos, como el MARC para las bibliotecas). En un nivel más práctico, el humanista lidia con la clasificación y estructuración de su objeto de estudio, además de con los datos informáticos propiamente dichos. Por ello, parece necesario tener una perspectiva genérica de qué tipos hay, qué estructuras existen, qué organización pueden tomar, cuál puede ser su naturaleza (y la diferencia entre datos y metadatos), pues esto ayuda al humanista a imaginar, concebir y transformar su objeto de estudio en datos concretos.

Hay además una serie de cuestiones críticas que deben considerarse al trabajar en HD y concretamente con datos informáticos. En primer lugar, la objetividad de los datos es a veces solo aparente pues estos son manufacturados y generados por personas que dependen de un contexto cultural o ideológico. Existen, como se ha expuesto, otras cuestiones como la imposibilidad de captar la totalidad del objeto, los límites de la codificación de la dimensión interpretativa de nuestra disciplina o la falta de datos especialmente de aquellos históricos. A ello se suman factores externos que dificultan el trabajo digital, como son los prejuicios hacia los diferentes procesos que implica la labor digital, y la necesidad urgente de conferir un estatus reconocido y científico a la publicación y al almacenamiento de los datos.

Relacionado con el concepto de *dato* está su uso a gran escala y a través de ciertas tecnologías, como en el caso de big data. Queda todavía mucho por hacer pero algunas tendencias empiezan a aparecer más claramente, como el uso de ciertas tecnologías de minería de textos, de Procesamiento del Lenguaje Natural, o la apuesta por la web semántica, el uso de APIs, o la actualización de recursos ya existentes.

Se ha dicho que el futuro está en los datos (Rosenberg, 2013) y como tal, debemos apostar por la formación de nuestros investigadores en Humanidades para saber recolectarlos, crearlos, manipularlos, analizarlos y explotarlos al máximo.

[38] Véase para todos estos procesos las definiciones propuestas por TaDiRAH, *Taxonomía sobre actividades de investigación digital en* Humanidades, accesible en español desde: https://www.vocabularyserver.com/tadirah/es/index.php.

REFERENCIAS BIBLIOGRÁFICAS

Baca, M. (Ed.) (2008). *Introduction to Metadata* (2nd ed.). Los Ángeles: The Getty Research Institute.

Barlow, M. (2013). *The Culture of Big Data*. Recuperado de https://bit.ly/2IF1Gfl el 07/10/2019.

Borges, J. L. (1952). *Otras inquisiciones (1937-1952)*. Buenos Aires: Sur.

Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge: MIT Press.

_____ (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge: MIT Press.

Consultative Committee for Space DataSystem Secretariat (CCSDS) (2012). *Reference Model for an Open Archival Information System (OAIS)*. Recuperado de https://public.ccsds.org/pubs/650x0m2.pdf el 07/10/2019.

Coromines, J. (1954). *Diccionario crítico etimológico de la lengua castellana*. Madrid: Gredos.

Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, *5* (1). Recuperado de https://bit.ly/2d8fJf1 el 07/10/2019.

_____ (2015). Graphical Approaches to the Digital Humanities. En S. Schreibman, R. Siemens y J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 238-250). Londres: Blackwell Publishing.

Foucault, M. (1966). *Les mots et les choses*. París: Gallimard.

Gitelman, L. (2013). *"Raw data" is an Oxymoron*. Cambridge: MIT Press.

Gómez de Silva, G. (1988). *Breve diccionario etimológico de la lengua española: 10.000 artículos, 1.300 familias de palabras* (1ª ed.). México: Fondo de Cultura Económica.

Graham, S., Milligan, I. y Weingart, S. (2015). *Exploring Big Historical Data. The Historian's Macroscope*. Recuperado de http://www.themacroscope.org/2.0/ el 07/10/2019.

Hockey, S. (2004). The History of Humanities Computing. En S. Schreibman, R. Siemens y J. Unsworth (Eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. Recuperado de https://bit.ly/35pNrol el 07/10/2019.

Ide, N. M. y Sperberg-McQueen, C. M. (1995). The TEI: History, Goals, and Future. *Computers and the Humanities*, *29*(1), 5-15.

Jones, S. E. (2016). *Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*. Abingdon: Routledge.

Kaplan, F. (2015). A Map for Big Data Research in Digital Humanities. *Frontiers in Digital Humanities*, 2, 1-7. doi:10.3389/fdigh.2015.00001.

Kirschenbaum, M. (2010). What Is Digital Humanities and What's It Doing in English Departments? *ADE Bulletin*, *150*, 55-61. Recuperado de https://bit.ly/2VA8Mr4 el 07/10/2019.

Laney, D. (2001). *3D-Data Management: Controlling Data: Volume, Velocity and Variety*. Recuperado de https://gtnr.it/2cFHqsu el 07/10/2019.

Laue, A. (2004). How the Computer Works. En S. Schreibman, R. Siemens, y J. Unsworth (Eds.), *A Companion to Digital Humanities* (pp. 145-160). Oxford: Blackwell. doi:10.1002/9780470999875.ch13.

Levi, A. S. (2013). Humanities 'big data': Myths, Challenges, and Lessons. En *2013 IEEE International Conference on Big Data* (pp. 33-36). California: IEEE. doi:10.1109/BigData.2013.6691667.

Manovich, L. (1999). Database as Symbolic Form. *Convergence*, *5*(2), 80-99. doi:10.1177/135485659900500206.

_____ (2001). *The Language of New Media*. Cambridge: MIT Press.

_____ (2011, junio 20). Trending: The Promises and the Challenges of Big Social Data [Entrada de blog]. Recuperado de https://bit.ly/2M7GBg1 el 07/10/2019.

_____ (2016, mayo 23). The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics. *Journal of Cultural Analytics*. doi:10.22148/16.004.

Marche, S. (2012, octubre 28). Literature Is not Data: Against Digital Humanities [Entrada de blog]. Recuperado de https://bit.ly/1UAqPZ5 el 07/10/2019.

Mayer-Schönberger, V. y Cukier, K. (2013). *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Londres: John Murray.

National Research Council (1999). *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington: The National Academies Press. doi:10.17226/9692.

O'Carroll, E. (2018, mayo 16). When the Humanities Meets Big Data. *Christian Science Monitor*. Recuperado de https://bit.ly/2IwU7Wd el 07/10/2019.

Owens, T. (2011). Defining Data for Humanists: Text, Artifact, Information or Evidence? *Journal of Digital Humanities*, *1*(1). Recuperado de https://bit.ly/2IBuuN7 el 07/10/2019.

Presner, T. y Shepard, D. (2016). Mapping the Geospatial Turn. En S. Schreibman, R. Siemens y J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 199-212). Londres: Blackwell Publishing.

Real Academia Española (s.f.). Corpus diacrónico del español (CORDE). Recuperado de http://corpus.rae.es/cordenet.html el 07/10/2019.

Rio Riande, G. del (2015). Humanidades Digitales. Mito, actualidad y condiciones de posibilidad en España y América Latina. *ArtyHum*, *1*, 7-19. Recuperado de https://bit.ly/2Mz46h3 el 07/10/2019.

_____ (2016). ¿De qué hablamos cuando hablamos de Humanidades Digitales? En *Actas I Jornadas de Humanidades Digitales de la AAHD* (pp. 50-62). Buenos Aires: Editorial de la Facultad de Filosofía y Letras. Recuperado de https://bit.ly/35oyWBv el 07/10/2019.

_____ (2018). Humanidades Digitales: Cuando lo local es global. En *Humanidades Digitales. Construcciones locales en contextos globales. Actas del I Congreso Internacional de la Asociación Argentina de Humanidades Digitales* (pp. 1-15). Buenos Aires: Editorial de la Facultad de Filosofía y Letras. Recuperado de https://bit.ly/2ovl7PV el 07/10/2019.

Rio Riande, G. del, De Léon, R. y Ferreyra, D. (2015). *Taxonomía sobre actividades de investigación digital en Humanidades, TaDIRAH (DARIAH)*. Recuperado de http://www.vocabularyserver.com/tadirah/es/ el 07/10/2019.

Rojas Castro, A. (2017). Big Data in the Digital Humanities. New Conversations in the Global Academic Context. En R. Good, R. Carreras, E. Snijders, A. Rojas Castro, P. Diezma, y D. Ruíz Torres (Eds.), *AC/E Digital Culture 2017 Annual Report: Smart Culture: Analysis of Digital Trends. Focus: The Use of Digital Technologies in the Conservation, Analysis and Dissemination of Cultural Heritage* (1st ed.) (pp. 62-71). Madrid: Acción Cultural Española (AC/E).

Rosenberg, D. (2013). Data before the Fact. En L. Gitelman (Ed.), *"Raw data" is an Oxymoron* (pp. 15-40). Cambridge: MIT Press.

Sandoval de la Maza, S. (1995). *Diccionario etimológico de la lengua castellana*. Madrid: M. E. Editores.

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, *2*(3). Recuperado de https://bit.ly/191nMDf el 07/10/2019.

Schreibman, S., Siemens, R. y Unsworth, J. (Eds.). (2004). *A Companion to Digital Humanities*. Oxford: Blackwell. Recuperado de https://bit.ly/35pNrol el 07/10/2019.

Sperberg-McQueen, C. M. (2004). Classification and its Structures. En S. Schreibman, R. Siemens, y J. Unsworth *(Eds.), A Companion to Digital Humanities*. Oxford: Blackwell. Recuperado de https://bit.ly/2M3tn3A el 07/10/2019.

Spiro, L. (2011). *Why Digital Humanities?* GLCA's New Directions Digital Humanities Workshop, Holland, Michigan. Recuperado de https://bit.ly/2MtuKaZ el 07/10/2019.

_____ (2012). "This is Why We Fight": Defining the Values of the Digital Humanities. En M. K. Gold (Ed.), *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press. Recuperado de https://bit.ly/2B1usTy el 07/10/2019.

Terras, M. (2012, enero 20). Infographic: Quantifying Digital Humanities [Entrada de blog]. Recuperado de https://bit.ly/2B1usTy el 07/10/2019.

Vanhoutte, E. (2013). The Gates of Hell: History and Definition of Digital. En M. Terras, J. Nyhan, y E. Vanhoutte (Eds.), *Defining Digital Humanities: A Reader* (pp. 119-156). Recuperado de https://bit.ly/2pe8eL3 el 07/10/2019.

Ward, J. S. y Barker, A. (2013). Undefined By Data: A Survey of Big Data Definitions. *arXiv:1309.5821 [cs.DB]*. Recuperado de http://arxiv.org/abs/1309.5821 el 04/10/2019.

Weingart, S. (2015). Big Data. En S. Graham, I. Milligan y S. Weingart (Eds.), *Exploring Big Historical Data. The Historian's Macroscope*. Recuperado de http://www.themacroscope.org/?page_id=597 el 04/10/2019.