

BIG DATA EN LAS HUMANIDADES DIGITALES. NUEVAS CONVERSACIONES EN EL CONTEXTO ACADÉMICO GLOBAL

ANTONIO ROJAS CASTRO [@ROJASCASTROA](#)

Antonio Rojas Castro es doctorado en Humanidades por la Universitat Pompeu Fabra (Barcelona, 2015). En esta misma universidad fue investigador predoctoral, becario FPI del grupo de investigación Todo Góngora II y profesor de asignaturas sobre escritura académica y estudios literarios. En 2015 co-coordinó un monográfico dedicado a las humanidades digitales para la revista *Ínsula*. Actualmente es editor de The Programming Historian en español, está al cargo de la comunicación de la Asociación Europea de Humanidades Digitales (EADH) y trabaja como investigador posdoctoral en el Cologne Center for eHumanities (Alemania).

Introducción

Navidades de 2016. Un momento perfecto para recapitular, resumir y publicar listas con los principales acontecimientos ocurridos durante el año. Google Trends publica cuáles han sido las búsquedas más frecuentes, ordenadas por categorías como «Noticias», «Personas», «Tecnología», «Películas», «Música», «Deportes» o «Fallecimientos». Pocos días antes, la empresa sueca Spotify, que da acceso en línea a millones de canciones, lanza una campaña publicitaria basada en los datos producidos por los usuarios. Así, algunos de los gigantescos carteles que invaden las calles de Londres contienen mensajes como: «Querida persona que se puso *Sorry* 42 veces en San Valentín..., ¿qué has hecho?»; o bien: «Queridas 3749 personas que se pusieron *It's the End of the World As We Know It* el día de la votación por el Brexit..., seguid así».



Campaña publicitaria de Spotify 2016 © Adweek] (Nudd, 2016).

La campaña de Spotify sorprende y resulta eficaz porque juega con la complicidad del receptor. Ahora bien, ¿qué tiene que ver todo esto con las disciplinas humanísticas dedicadas al estudio de documentos, textos e imágenes del pasado? O, mejor dicho, ¿en qué medida la gestión de grandes datos amasados por las empresas puede ayudarnos a comprender mejor los límites de nuestro pensamiento, la lengua, los acontecimientos históricos, en fin, todas las *expresiones de nuestra mente humana*?

Si admitimos que las disciplinas humanísticas, como la filosofía, la filología o la historia, no solo

se caracterizan por un objeto de estudio concreto sino por un método que persigue comprender casos particulares, singulares, incluso únicos, a través del comentario de texto, la respuesta será, sin duda, negativa: «nada, o muy poco». Ahora bien, como ha defendido recientemente el profesor Rens Bod (2013), desde la Antigüedad, los humanistas también han buscado principios generales, leyes y patrones para explicar nuestra cultura, y han cambiado en numerosas ocasiones (para bien o para mal) nuestra forma de percibir el mundo.

Las disciplinas humanísticas, como la filosofía, la filología o la historia, no solo se caracterizan por un objeto de estudio concreto sino por un método que persigue comprender principios generales, leyes y patrones para explicar nuestra cultura.

Debemos empezar por rechazar ciertos clichés en torno a las humanidades e interrogarnos sobre los objetos clásicos de las letras humanas teniendo en cuenta los métodos disponibles actualmente. Este requisito tampoco es ajeno al trabajo de los humanistas, pues siempre han estado en contacto con otras disciplinas limítrofes como la antropología, el marxismo o los estudios de género. Así, los humanistas han establecido en los últimos años un diálogo fructífero con la informática y las ciencias sociales —lo que se ha llamado «giro computacional» (Berry, 2011)—. En este contexto académico, la expresión «Big Data» se ha infiltrado de manera directa en los debates sobre la «escala» —¿cómo estudiar todas las novelas de los siglos XVIII y XIX escritas en Inglaterra, Francia, Alemania, Estados Unidos, España o Japón?—; o bien, más a menudo, de manera oblicua mediante conceptos más familiares para los humanistas, como «lectura distante» (Moretti, 2007) o «macroanálisis» (Jockers, 2013).

Estos cambios han sido posibles gracias a que los métodos estadísticos e informáticos, y otros métodos de las ciencias sociales, se han transformado y han sabido adaptar sus modelos conceptuales a la complejidad de los textos (English y

Underwood, 2016). En otras palabras, estamos ante una verdadera conversación en la que los distintos interlocutores dialogan y se escuchan.

De lo particular en lo universal

Desde el año 2011 la expresión «Big Data» se ha extendido en las ciencias experimentales y los medios de comunicación como si el incremento de los datos disponibles fuera la próxima revolución científica. Se trata de una expresión utilizada en la academia, la industria y los medios de comunicación..., pero ¿qué es exactamente? ¿Es un objeto de estudio, un método, un conjunto de tecnologías o una disciplina?

Uno de los pocos artículos que han arrojado algo de luz al respecto lleva por título «Undefined By Data: A Survey of Big Data Definitions». Los autores (Ward y Barker, 2013) se dedican a rastrear las distintas definiciones de «Big Data» proporcionadas por grandes empresas tecnológicas como Oracle, Intel y Microsoft, y algunos informes previos. En términos generales, las definiciones aúnan dos ideas importantes: por un lado, el almacenaje de un gran volumen de datos (algunos autores hablan de 500 Terabytes semanales); por el otro, el análisis cuantitativo y visual de estos datos con el fin de encontrar patrones, establecer leyes y predecir conductas.

Big Data implica tanto el almacenaje de un gran volumen de datos como el análisis cuantitativo y visual de esos datos con el fin de encontrar patrones, establecer leyes y predecir conductas.

La definición clásica de «Big Data» consta de una fórmula fácil de entender y memorizar —las tres V—: el *Volumen* (Terabytes, Petabytes, Exabytes), la *Velocidad* (datos en continua generación) y la *Variación* (textos, imágenes, sonidos) (Ward y Barker, 2013). Posteriormente, algunos informes han añadido a la definición una cuarta V correspondiente al término *Veracidad*. En cualquier caso, esta definición del Big Data, anclada

en el volumen, solo tiene sentido si tenemos en cuenta como principal fuente de datos los blogs, las redes sociales y los sensores.

Por el contrario, el objeto clásico de las humanidades suelen ser textos e imágenes analógicas que, con fortuna, han sido digitalizadas y publicadas en un formato legible por el ordenador. En otras palabras, si atendemos a las tres V, debemos admitir que en las humanidades no podemos hablar de Big Data en sentido estricto. Por un lado, los clásicos de la poesía española del Siglo de Oro caben en un pendrive de 4GB; por el otro, los archivos y bibliotecas no producen de manera continua nuevos datos y a gran velocidad sobre nuestros poetas, literatos o artistas (o, mejor dicho, estos no son accesibles para el investigador). En cuanto a la variedad, estaríamos ante archivos de imágenes en TIFF, JPEG o algún formato similar, y de texto semiestructurado en formato XML o sin marcado alguno, en formato TXT.

Antes del advenimiento de Google Books en 2004, los humanistas digitales han estado trabajando para digitalizar corpus de textos e imágenes en forma de ediciones, bibliotecas y archivos digitales. La Asociación Europea de Humanidades Digitales (EADH) nos ofrece una buena muestra del tipo de proyectos desarrollados. Desde el año 2015, la asociación ha dedicado un espacio en su página web para documentar y promover el acceso a proyectos europeos de humanidades digitales realizados durante los últimos cinco años. La iniciativa tiene naturaleza participativa porque cualquier investigador (sea miembro de la asociación o no) puede rellenar el formulario accesible en la web y enviar una descripción de su proyecto con información sobre el nombre del mismo, un resumen descriptivo, las instituciones colaboradoras o el equipo responsable, entre otros campos. En el momento en que escribo este artículo, la asociación ha recibido 175 envíos. Si se analizan los títulos y los resúmenes con Voyant, una herramienta que cuenta las palabras más frecuentes, se puede percibir sin dificultad que abundan palabras relacionadas con el tema analizado en este

el proceso de recolección. Así, por ejemplo, la investigadora Johanna Drucker (2011) rechaza el término «data» —que en latín significa «lo que nos es dado»— y utiliza en su lugar el término «capta», es decir, «lo que ha sido capturado o recogido»; como es obvio, esta intervención crítica pone de relieve la naturaleza imparcial e incompleta de los datos.

Los humanistas digitales también han insistido en la temporalidad de los datos —pues todos tienen fecha de creación y caducidad— y en la falacia consistente en separar los datos de los metadatos (esto es, datos como título, creador, tema, descripción, fecha, formato, identificador, fuente, lengua, etc.). En realidad, no existe el segundo grado de los datos, que se cristaliza en la raíz *meta-*; los metadatos son tan importantes, selectivos e imparciales como los datos porque son producidos por seres humanos (o bien por algoritmos diseñados por seres humanos). Tampoco es válida la distinción, que se remonta al triángulo culinario de Lévi-Strauss, entre «datos en crudo» y «datos cocinados», o entre «datos», «materia prima» e «información».

De hecho, para investigadores como Tom Boellstorff (2013), los datos tienen una naturaleza densa, interpretativa y contextual, y por eso es preferible hablar de «thick data». Parafraseando al antropólogo Clifford Geertz, los datos deben tomarse como «construcciones propias basadas en las construcciones de otras personas sobre objetos imaginados por una comunidad determinada».

Por ejemplo, la Text Encoding Initiative es una organización sin ánimo de lucro que tiene por objetivo publicar unas *Recomendaciones* sobre cómo codificar textos humanísticos con lenguaje de marcado XML para que sean intercambiables y, más o menos, estándares. Se trata de una organización participativa en la que cualquier investigador puede proponer cambios o mejoras basados en su experiencia sobre el conjunto de etiquetas definidas por el consorcio. Hasta el año 2012, sin embargo, ninguno de sus miembros había puesto en cuestión que la etiqueta <sex>

utilizada para describir el género de una persona mencionada en un texto, siguiera la norma ISO/IEC 5218:2004 y, por tanto, tuviera como valor del atributo @value las cifras 1 (hombre), 2 (mujer), 9 (no aplicable) y 0 (desconocido).

La situación fue reformulada cuando una investigadora denunció que esta tipología era sexista, porque situaba a la mujer en segundo lugar respecto al hombre y codificaba con lenguaje de marcado las estructuras patriarcales (Terras, 2013). Con esto no quiero restar valor a la importancia de la TEI, sobre todo en la configuración de las humanidades digitales, sino poner de relieve que la tecnología, los datos, los algoritmos y los estándares son producto de una interpretación del mundo y llevan marcas culturales. En conclusión, los datos no deben verse como verdades absolutas sino interrogarse de manera crítica.

La tecnología, los datos, los algoritmos y los estándares son producto de una interpretación del mundo y llevan marcas culturales. Los datos no deben verse como verdades absolutas sino interrogarse de manera crítica.

En defensa de la teoría

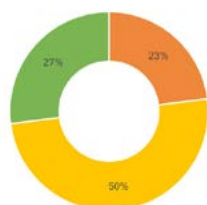
En la bibliografía sobre el Big Data también es frecuente encontrarse un descrédito de la teoría. La argumentación, a grandes rasgos, es la siguiente: si disponemos de grandes cantidades de datos y de métodos estadísticos eficaces, no necesitamos teorías, modelos e hipótesis, que deban demostrarse o refutarse con experimentos. Dicho en otras palabras, en la era del Petabyte, el método científico resulta obsoleto (Anderson, 2008). El rechazo a las teorías y modelos no solo ha recibido crédito en el mundo empresarial; también ha sido aceptado en algunos trabajos de naturaleza humanística. Así, por ejemplo, Jean-Gabriel Ganascia (2015: 632-633) afirma que ya no es necesaria una teoría o hipótesis previa si se analizan todos los datos

existentes y no una muestra o un grupo reducido, como se solía hacer hasta ahora.

Frente a esta perspectiva no han sido pocos los trabajos que han tenido por objetivo revalidar la importancia de las teorías, los modelos y las hipótesis para investigar. Por un lado, hay que recordar que nuestro patrimonio cultural (documentos, textos, pinturas, imágenes, sonidos) no está digitalizado por completo pese a los esfuerzos colectivos de iniciativas como Europea. Según el último informe realizado por el proyecto de la Comisión Europea llamado ENUMERATE (Nauta y Wietske, 2015), en la actualidad solo un 23% de las colecciones europeas ha sido digitalizado. La encuesta fue respondida por unas mil instituciones europeas como bibliotecas, museos, archivos, etc. Estas instituciones aún tienen pendiente digitalizar un 50% de sus colecciones y admiten que un 27% de los objetos que preservan no serán digitalizados. Las cifras ponen de relieve que gran parte de nuestro patrimonio no es accesible en Internet.

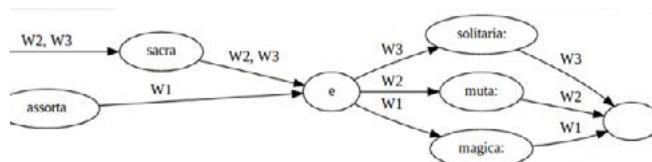
La digitalización del patrimonio cultural en Europa (2015)

■ Patrimonio cultural digitalizado ■ Patrimonio cultural pendiente de digitalizar ■ Patrimonio cultural que no se planea digitalizar



Por un lado, el proceso de digitalización siempre implica una selección basada en los recursos disponibles por la institución o el grupo de trabajo encargado de digitalizar los documentos; pero también responde a razones ideológicas e identitarias. No hay que olvidar que los museos, bibliotecas y archivos son instituciones financiadas con dinero público y que tienen por función preservar y difundir el patrimonio cultural de una comunidad (por ejemplo, una nación). Por el otro, los formatos, lenguajes de marcado y algoritmos también están inscritos en una cultura y una ideología determinadas, y llevan consigo adheridos muchos presupuestos que varían en función del contexto.

Desde un punto de vista humanístico, pues, cuesta creer que el análisis de grandes cantidades de datos vuelva inútil el método científico, ya sea porque jamás disponemos de todos los datos existentes —una de las premisas del Big Data es la *Velocidad* con que se generan nuevos datos—, ya sea porque los datos son erróneos o ambiguos, ya sea porque el procesamiento (automático o no) de los datos está determinado por nuestra cultura y, por tanto, contiene sesgos ideológicos. Por ejemplo, tomemos el caso de CollateX; esta herramienta está diseñada para cotejar textos que presentan ligeras variaciones y alinear las partes de los textos que difieren. Entre otros presupuestos del algoritmo, hay que destacar que para CollateX no es relevante distinguir entre una transposición o cambio de lugar de una porción de texto (por ejemplo, en un poema, una estrofa que aparece desplazada o en distinto lugar) y una sustitución (es decir, la eliminación de una estrofa en un sitio y la adición de los mismos versos en otro lugar) (Van Zundert, en prensa); la cuestión aquí no es determinar si el algoritmo de CollateX es correcto. Los investigadores pueden estar de acuerdo o no, pero la clave estriba en conocer esta elección, esta preferencia, y ser consciente de que condiciona los resultados y la interpretación.



Fragmento de un grafo de variantes creado con la herramienta CollateX.

Algunos autores, de hecho, defienden que, en la era del Big Data, las teorías y los modelos cobran aún mayor importancia porque es necesario explicar y entender los fenómenos analizados mediante abstracciones. En las humanidades digitales el concepto de «modelo» está muy extendido porque ayuda a explicar el núcleo del trabajo de digitalización. Los modelos se entienden como herramientas, esquemas o diseños utilizados en un contexto específico con unos fines determinados, a veces prácticos (poner en línea un conjunto de textos), pero, a menudo,

sobre todo en el ámbito académico, especulativos (entender la estructura de los textos). Más que un producto acabado, lo que importa en las humanidades digitales es el proceso creativo que tiene lugar cuando se «modela» un fenómeno porque la finalidad es obtener nuevo conocimiento, nuevos significados, a través de la generación de un objeto externo que lo representa.

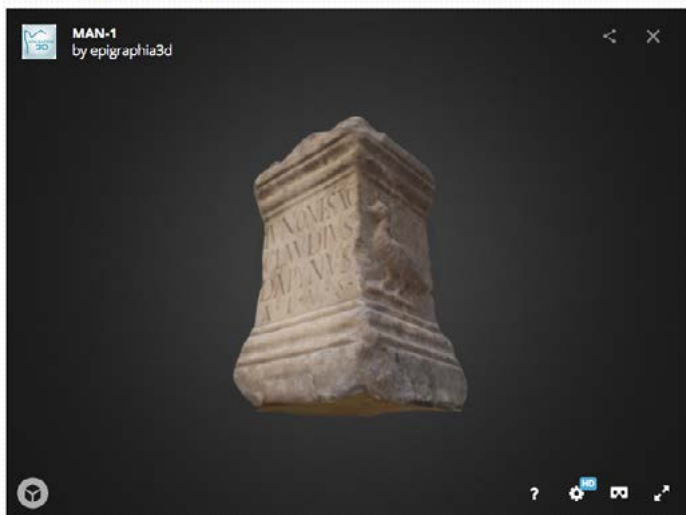
La relación entre el objeto externo (por ejemplo, una inscripción epigráfica) y la representación (una reconstrucción 3D que permite ver la lápida desde distintas perspectivas y aumentando el detalle) se basa en la similitud; de ahí que sea importante situar la reflexión sobre el «modelado» en la tradición de semiótica y de la ciencia de los signos (Ciula y Eder, 2016). Por supuesto, existen distintos grados de similitud; la relación

de similitud puede ir desde la identidad completa hasta la metáfora pasando por cierta analogía entre las cualidades del objeto representado y la representación digital.

Los modelos digitales son, pues, iconos que nos ayudan a pensar y conocer mejor el original, el objeto analógico. Este tipo de pensamiento ha sido identificado como «abducción» porque se encuentra a medio camino de la inducción y de la deducción, y descansa en la intuición y la experiencia de la persona que «modela» (Bryant y Raja, 2014). En otras palabras, el proceso de modelado está influenciado por elementos contextuales como las hipótesis de partida, los presupuestos teóricos, los métodos científicos, los formatos y las tecnologías.

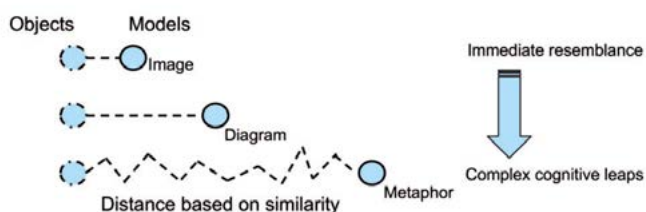
Los modelos digitales son iconos que nos ayudan a pensar y conocer mejor el original, el objeto analógico. Este tipo de pensamiento ha sido identificado como «abducción».

1. Dedicatoria a Juno



Augusta Emerita , Mérida (Badajoz)
Museo Arqueológico Nacional
Nº inventario: 34449

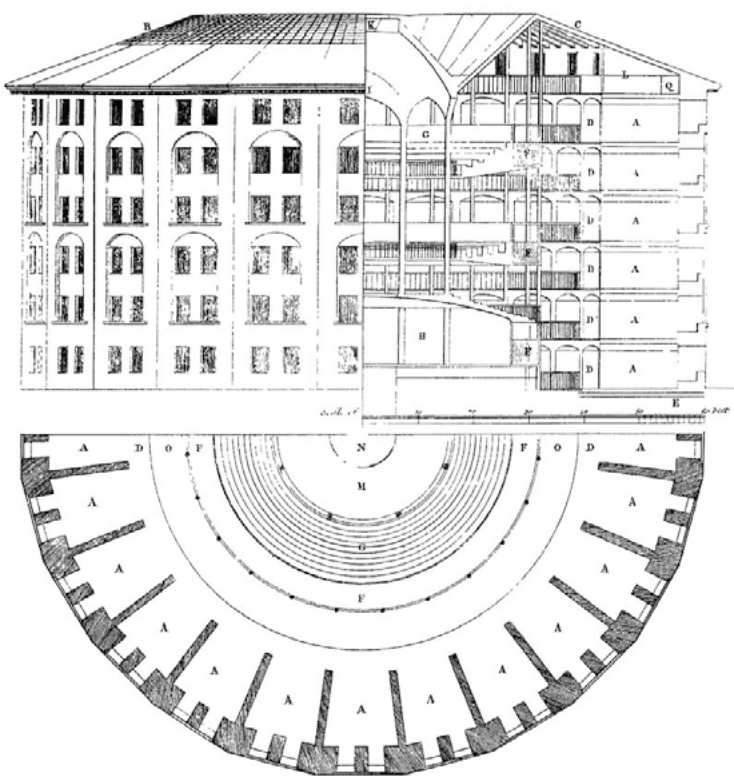
Inscripción epigráfica modelada en 3D. © Epigraphia 3D
<http://www.epigraphia3d.es/>



Grados de iconicidad © Ciula y Eide, 2016.

Dentro del Panóptico

La producción de grandes cantidades de datos en tiempo real y de manera continua a través de las redes sociales también tiene su contrapartida oscura. No son raras las comparaciones del Big Data con el Big Brother o, mejor incluso, con el Panóptico —un tipo de arquitectura carcelaria ideada por Jeremy Bentham en el siglo XVIII que crea la sensación de estar en constante vigilancia—, sobre todo a raíz del caso Edward Snowden. Los Gobiernos controlan a los ciudadanos para garantizar su seguridad; esto no es ninguna novedad, sino que forma parte de la historia de las estructuras de poder estudiadas por Michel Foucault, entre otros. En el Estado moderno el sujeto es vigilado y, al mismo tiempo, incitado a revelar sus más íntimos secretos a través de la confesión, la terapia psicoanalítica o, en la actualidad, la publicación de «estados» en Facebook.



Diseño del panóptico de Bentham.

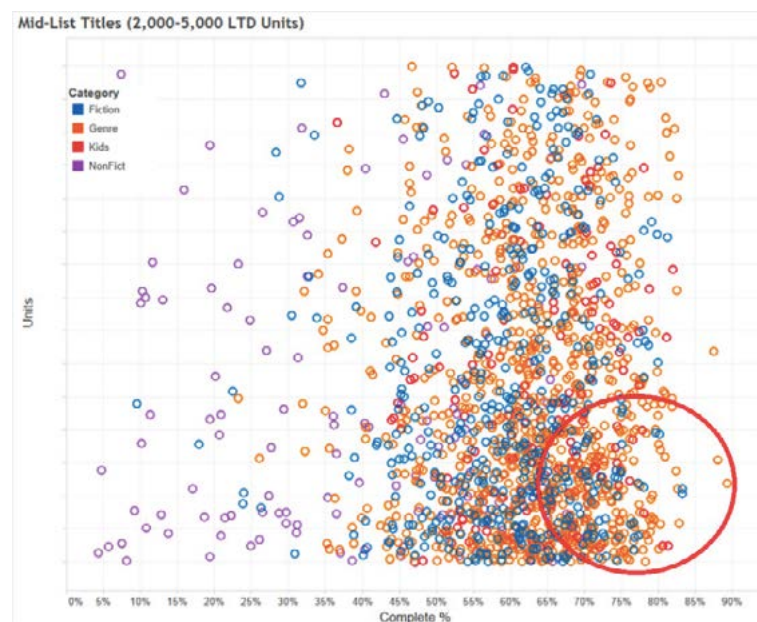
Como ya se ha visto, el objeto de las humanidades suele ser externo, autónomo y terminado —un documento histórico, un texto literario, una representación visual—, por lo que la investigación no suele plantear dilemas éticos sobre la privacidad de los creadores y los receptores. Ahora bien, como consumidores de cultura, nuestros actos son registrados cada vez que buscamos en Internet un libro, una película, una canción, clicamos sobre el producto y lo compramos; lo mismo ocurre cuando visitamos un museo —la cámara de vigilancia está ahí para proteger nuestro patrimonio de delitos y robos, pero también para controlar al visitante—; por último, cuando cogemos en préstamo un libro de una biblioteca pública queda un registro en la base de datos.

El caso de las bibliotecas públicas es especialmente interesante porque son infraestructuras de proximidad a las que cualquier persona, con independencia de su clase económica, puede acceder. Los bibliotecarios registran en sus bases de datos cada uno de los préstamos realizados, en qué momento y por quién. Este tipo de datos,

sin embargo, no son accesibles porque las bibliotecas municipales tienen una larga tradición relacionada con la protección de datos (Starr, 2004). En cambio, sí que es pública una lista de libros más prestados, que funciona como indicador del gusto contemporáneo. Para que fueran objeto de investigación estos datos deberían publicarse en un formato abierto como XML o CSV, e incluir una serie de metadatos, como el lugar y el momento en que se produjo el préstamo, algo que pondría en peligro la privacidad de los usuarios.

Como consumidores de cultura, nuestros actos son registrados cada vez que buscamos en Internet un libro, una película, una canción y clicamos sobre un producto y lo compramos.

Para los investigadores interesados en los hábitos de lectura, sin duda, resultaría una revolución disponer de tantos datos. Por ejemplo, sería posible determinar cómo el cine, la televisión y la publicidad condicionan los gustos y los hábitos lectores de la población. Los fabricantes de libros electrónicos, por ejemplo, ya están utilizando estadísticas de lectura para descubrir qué libros pueden ser considerados buenos, porque los lectores los han terminado, pero no han



Recepción positiva, ventas bajas © Kobo.

cosechado grandes ventas; o bien para identificar al próximo Dan Brown teniendo en cuenta el grado de satisfacción de los lectores con libros escritos por autores desconocidos (Kobo, 2014). En fin, todos los datos generados con nuestro libro electrónico son amasados por las empresas editoriales para conocer mejor la relación entre ventas/satisfacción; de esta manera, es más fácil justificar decisiones económicas sobre el futuro editorial de un determinado autor, saga literaria o género.

Con esto no quiero decir que las bibliotecas públicas y los museos deban actuar como las empresas, sino señalar que el estado de vigilancia —al igual que los espacios de resistencia— es previo a las redes sociales. Del mismo modo que empresas como Twitter han sido acusadas de ejercer un poder coercitivo sobre la investigación en ciencias sociales (Reichert, 2015), debemos preguntarnos cómo los humanistas pueden estudiar los hábitos culturales de los ciudadanos, en diálogo siempre con las bibliotecas y los museos, y utilizando métodos para anonimizar los datos. El objetivo, desde mi punto de vista, es que empresas como Spotify o Amazon no sepan más sobre una determinada sociedad —sobre nuestros gustos, inquietudes, estados de ánimo— que los integrantes de la misma.

Conclusiones

Desde el año 2011 la expresión «Big Data» se ha extendido en las ciencias experimentales y los medios de comunicación como si el incremento

de los datos disponibles fuera la próxima revolución científica. Aunque hay mucho «hype» al respecto, las humanidades no han sido ajenas a este fenómeno; de manera más específica y pese a que la digitalización de nuestro patrimonio cultural es incompleta, en las humanidades digitales se pueden encontrar varias publicaciones que conversan con el Big Data y las ciencias sociales. En el ámbito académico europeo, son destacables numerosos proyectos que procesan grandes cantidades de datos para estudiar la lengua, la literatura o el arte gracias a técnicas como el procesamiento del lenguaje natural, la visión automática por ordenador, el modelado de tópicos y la estilometría.

Tras analizar el significado de la expresión «Big Data», con este artículo se ha puesto el acento en la naturaleza cultural de los datos y se ha defendido la vigencia de las teorías, los modelos y las hipótesis para llevar a cabo una investigación científica. Por último, se ha tratado la dialéctica privacidad-control. Este asunto escapa, en cierto modo, al ámbito tradicional de las humanidades, pero también merece nuestra atención como ciudadanos del siglo XXI interesados en las prácticas culturales del presente. Los humanistas, sin duda, tienen mucho que aportar a los debates éticos y epistemológicos sobre el uso de los datos generados por la ciudadanía, recordando la condición «capturada» y cultural de los mismos y proporcionando su experiencia para analizar casos particulares teniendo en cuenta el contexto general.

Bibliografía

- Anderson, Chris (23.06.2008), «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete», *Wired*. <https://www.wired.com/2008/06/pb-theory/>.
- Berry, D. M. (2011), «The Computational Turn: Thinking about the Digital Humanities», *Culture Machine*, 12. <http://www.culturemachine.net/index.php/cm/article/viewarticle/440>.
- Blei, David M. (2012), «Probabilistic Topic Models», *Communications of the ACM*, 55.4, pp. 77-84. <http://cacm.acm.org/magazines/2012/4/147361-probabilistic-topic-models/fulltext>
- Bod, Rens (2013), *A New History of the Humanities*, Oxford University Press.
- Boellstorff, Tom (2013), «Making big data, in theory». *First Monday*, 18.10. <http://firstmonday.org/ojs/index.php/fm/article/view/4869>
- Bryant, Anthony y Raja, Uzma (2014), «In the realm of Big Data...», *First Monday*, 19.2. <http://firstmonday.org/ojs/index.php/fm/article/view/4991>
- Burrows, John (2002), «“Delta”: a measure of stylistic difference and a guide to likely authorship», *Literary and Linguistic Computing*, 17.3, pp. 267-287.
- Ciula, Arianna y Eder, Øyvind (2016), «Modelling in the digital humanities: Signs in context», *Digital Scholarship in the Humanities*.
- Drucker, Johanna (2011), «Humanities Approaches to Graphical Display», *DHQ: Digital Humanities Quarterly*, 5.1. <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
- English, James F. y Underwood, Ted (2016), «Shifting Scales: Between Literature and Social Science», *Modern Language Quarterly*, 77.3, pp. 278-295. <http://mlq.dukejournals.org/content/77/3/277.full>
- Ganascia, Jean-Gabriel (2015), «Les Big Data dans les Humanités», *Critique*, 818-819, pp. 627-636. <https://www.cairn.info/revue-critique-2015-8-page-627.htm>
- , Glaudes, Pierre y Del Lungo, Andrea (2015), «Automatic detection of reuses and citations in literary texts», *Literary and Linguistic Computing*, 29.3, pp. 412-421.
- Jänicke, S., Franzini, G., Faisal, C., Scheuermann, G. (2016), «Visual Text Analysis in Digital Humanities», *Computer Graphics Forum*, 35.2. DOI: 10.1111/cgf.12873.
- Jockers, Matthew (2013), *Macroanalysis. Digital Methods and Literary History*, University of Illinois Press.
- Kobo (2014), «Publishing in the Era of Big Data». http://news.kobo.com/_ir/159/20149/Publishing%20in%20the%20Era%20of%20Big%20Data%20-%20Kobo%20Whitepaper%20Fall%202014.pdf
- Moretti, Franco (2007), *Graphs, Maps, Trees: Abstract Models for Literary History*, Londres, Verso.
- Nauta, Gerhard Jan y Heuvel, Wietske van den (2015), *Survey Report on Digitisation in European Cultural Heritage Institutions 2015*, ENUMERATE. <http://dataplatfom.enumerate.eu/reports/survey-report-on-digitisation-in-european-cultural-heritage-institutions-2015/detail>
- Nudd, Tim (2016), *Spotify Crunches User Data in Fun Ways for This New Global Outdoor Ad Campaign*, *Adweek*. <http://www.adweek.com/adfreak/spotify-crunches-user-data-fun-ways-new-global-outdoor-ad-campaign-174826>

Reichert, Ramón (2015), «Big Data. Digital Media Culture in Transition», en Sabine Himmelsbach y Claudia Mareis (eds.), *Poetics and Politics of Data. The Ambivalence of Life in a Data-Driven Society*, Basilea, Christoph Merian Verlag, pp. 147-166.

Rosado Rodrigo, Pilar (2015), *Formas latentes: protocolos de visión artificial para la detección de analogías aplicados a la catalogación y creación artísticas* (tesis doctoral), Barcelona, Universitat de Barcelona. <http://www.tdx.cat/handle/10803/300302>

Starr, Joan (2004), «Libraries and national security: An historical review», *First Monday*, 9.12. <http://firstmonday.org/ojs/index.php/fm/article/view/1198/1118>

Terras, Melissa (27.03.2013), «On Changing the Rules of Digital Humanities from the Inside», <https://melissaterras.org/2013/05/27/on-changing-the-rules-of-digital-humanities-from-the-inside/>

Van Zundert, Joris (en prensa), *Digital Scholarship in the Humanities*.

Ward, Jonathan Stuart y Barker, Adam (2013), «Undefined By Data: A Survey of Big Data Definitions». <https://arxiv.org/abs/1309.5821>

Recursos digitales

Alliance of Digital Humanities Organizations: <http://adho.org/>

Europeana: <http://www.europeana.eu/portal/es>

European Association for Digital Humanities: <http://eadh.org/>

FreeLing: <http://nlp.lsi.upc.edu/freeling/node/1>

Asociación Humanidades Digitales Hispánicas: <http://www.humanidadesdigitales.org/inicio.htm;jsessionid=FDC5ED5Boo5786714E-45936B6E127DF8>

Text Encoding Initiative: <http://www.tei-c.org/index.xml>

The Programming Historian: <http://programminghistorian.org/>

Stylo R: <https://sites.google.com/site/computationalstylistics/stylo>

Voyant: <http://voyant-tools.org/>

Google Arts & Culture: https://www.google.com/culturalinstitute/beta/u/o/?utm_campaign=cilex_v1&utm_source=cilab&utm_medium=artsexperiments&utm_content=freefall

Tuiteros

Ted Underwood: [@Ted_Underwood](https://twitter.com/Ted_Underwood)

Lev Manovich: [@manovich](https://twitter.com/manovich)

Nuria Rodríguez Ortega: [@airun72](https://twitter.com/airun72)

Greta Franzini: [@GretaFranzini](https://twitter.com/GretaFranzini)

Dev Verhoeven: [@bestqualitycrab](https://twitter.com/bestqualitycrab)

Frank Fischer: [@umblaetterer](https://twitter.com/umblaetterer)

Matthew Lincoln: [@matthewdlincoln](https://twitter.com/matthewdlincoln)

José Calvo: [@eumanismo](https://twitter.com/eumanismo)

Elena González Blanco: [@elenagbg](https://twitter.com/elenagbg)

Dan Cohen: [@dancohen](https://twitter.com/dancohen)