

**Esteban Romero Frías y  
María Sánchez González (editores)**  
**Prólogos de Paul Spence y de Nuria Rodríguez Ortega**

Esteban Romero Frías; Octavio Rubiera; Belén Álvarez; Carlos Arcila; Mabel Calderín Cruz; Luis Núñez; Ysabel Briceño; Erika Ortega; Silvia Eunice Gutiérrez; María Sánchez González; Daniel Villar-Onrubia; Juan M. García Sierra; Manuel D. Pérez Heredia; Javier Jiménez Cuadros; Patricia Lázaro Pernias; Daniel Tena Parera; David Roca Correa; Josep María Blanco Pont; Alfonso González Quesada; Álvaro Baraibar; Victoria Tur-Viñes; María del Carmen Quiles-Soler; José Manuel de Pablos; Ana María Rodera Bermúdez; Ana María González Ramos; Ernesto Priani Saisó

# **Ciencias Sociales y Humanidades Digitales**

**Técnicas, herramientas y experiencias de  
*e-Research* e investigación en colaboración**

**CAC, Cuadernos Artesanos de Comunicación / 61**



Universidad  
de La Laguna



Universitat d'Alacant  
Universidad de Alicante



UNIVERSIDAD  
DE MÁLAGA



Sociedad Latina de  
Comunicación Social



# El *Big Data* como fenómeno y herramienta para la *e-Research* en entornos infosaturados y complejos

**María Sánchez González.** Profesora asociada doctora e investigadora en Periodismo, Universidad de Málaga (España).  
m.sanchezgonzalez@uma.es/ Técnico Área de Innovación  
Universidad Internacional de Andalucía (España). m.sanchez@unia.es

## 1. Introducción: evolución de datos en el ecosistema digital

### 1.1. El valor de los datos en Red

COMO apunta Paniagua (2013), los datos han estado y están presentes en todos los aspectos de nuestra cotidianeidad. Pero es sobre todo con la expansión del uso cotidiano de Internet cuando la actividad de entidades e individuos online deja una huella en forma de datos cuyo valor vienen tratando de aprovechar, mediante la puesta en marcha de sistemas tecnológicos de recogida y análisis de datos, instituciones o entidades del sector de la Economía Digital.

Así, por ejemplo, empresas como Google, concienciadas del valor de las búsquedas online, han ido creando sistemas inteligentes que, basados en algoritmos, se adelantan al usuario y las completan, en función, por ejemplo, de las más frecuentes que el sistema tenga registradas; o presentan unos resultados más adaptados a su perfil en función de variables como su localización y sus preferencias según búsquedas anteriores. Redes sociales como Facebook incluyen hoy algoritmos que, además de jerarquizar la aparición de las publicaciones de otros en el muro del usuario en función de su grado

de interacción con el emisor, ofrecen a éste páginas o publicaciones sugeridas y publicidad personalizada. Compañías como Amazon han incrementado su rentabilidad gracias a sistemas de compra por recomendación. Y la monitorización de las webs visitadas por un usuario o la actividad realizada por éste sobre las mismas sirve luego a las empresas para adelantarse a sus demandas con productos o servicios personalizados<sup>73</sup>.

Todas estas prácticas, sin las que, siguiendo a expertos como Mayer-Schönberger, del Internet Institute de la Universidad de Oxford, no existirían hoy, o al menos, no de la misma forma, tales empresas (Gonzalo, 2013) se basan en el mismo principio: el almacenamiento y la utilización de datos masivos mediante tecnologías y estándares propios de la web semántica, que permite agregar o conectar datos relacionados bien semánticamente bien por sus atributos y, con ello, mejorar la búsqueda y explotación de resultados en la Web<sup>74</sup>. con la expansión de las redes sociales y otras plataformas de *social media*, sus lenguajes, sin los cuales no se entendería el auge del *Big Data*, vienen siendo usados en un número creciente de aplicaciones.

## 1.2. Web al cuadrado y crecimiento exponencial de datos

Ha sido especialmente en el último lustro, ante la consolidación, pues, de *social media* contruidos sobre aplicaciones que, como las propias redes sociales, incorporan opciones de geolocalización, y la eclosión de dispositivos móviles inteligentes que permiten una conexión

---

<sup>73</sup> Resulta relevante, como ejemplo en este sentido, el episodio narrado en *El poder de los hábitos* por Charles Duhigg que muestra cómo una empresa norteamericana, Target, fue capaz de detectar que una mujer estaba embarazada, monitorizando su actividad online, antes que su propia familia (Goitia y Petry, 2013).

<sup>74</sup> Según la Wikipedia, la web semántica es “una ampliación de la Web, por medio de la que se intenta realizar un filtrado de manera automática pero precisa de la información”. Consiste en añadir a ésta, de manera formal y estandarizada, metadatos semánticos y ontológicos, como informaciones adicionales que describen el contenido, el significado y la relación de los datos y posibilitan evaluarlas automáticamente por máquinas de procesamiento (Cfr. [http://es.wikipedia.org/wiki/Web\\_sem%C3%A1ntica](http://es.wikipedia.org/wiki/Web_sem%C3%A1ntica). Última edición 5 de enero de 2014. Consultado el 14 de febrero de 2014), esto es, hacerlas comprensibles a esas máquinas.

permanente y ubicua a las mismas, cuando se ha producido una multiplicación exponencial del volumen de datos online. En su obra *Big, Data, Big Analytics* (...), Minelli, Chambers y Dhiraj (2013: 1-2), describen este fenómeno como consecuencia de “three converging ‘perfect storms’”, referidas a tres áreas, *computing*, *data* y *convergence*, y llegan a apuntar un crecimiento anual del volumen de datos generados online de un 40% (op.cit.: 5). Al tiempo que informes recientes muestran que buena parte de éstos proceden de usuarios a través de redes sociales y otros espacios de la Web social<sup>75</sup>.

De ahí que los propios autores que acuñaran, hace prácticamente una década, el término de Web 2.0, aseguren que esa etapa ha quedado obsoleta para dar paso a la llamada “Web Squared” o Web al Cuadrado (O’Reilly y Battelle, 2009), que combina participación con sistemas de inteligencia artificial. Aplicaciones como las propias redes sociales y otras basadas en la semántica, el filtrado o la inteligencia colectiva, son dirigidas no exclusivamente por los usuarios, sino también por sensores que, cuando se conectan desde distintos dispositivos, recogen datos, sobre su actividad online e incluso su ubicación, mostrándolos además en tiempo real. A mayor participación sobre esta web, más sensores alimentando estas aplicaciones, de forma que la web crece no ya de forma aritmética sino exponencial, y de ahí el término de Web al cuadrado. Aunque estos sensores han existido desde hace décadas, factores como el abaratamiento de los dispositivos y las comunicaciones inalámbricas (Soraya, 2013) y la llamada “Internet de las cosas”, esto es, objetos cotidianos identificados y capaces de conectarse para intercambiar información<sup>76</sup>, facilitan la interacción y hacen que la web sea cada vez más inteligente, más rápida, más personalizada y más móvil. De esta forma, “everything we share, everywhere we go, everything we say and everyone we follow or connect with, generates valuable information that can be used to improve consumer experiences and ultimately improve products and services.” (Solis, 2012)

---

<sup>75</sup> Los datos de la UIT recogidos por Paniagua (2013) evidencian que buena parte de estos datos proceden, entre otros, de sitios como Facebook, Twitter, Youtube o Flickr.

<sup>76</sup>[http://es.wikipedia.org/wiki/Internet\\_de\\_las\\_Cosas](http://es.wikipedia.org/wiki/Internet_de_las_Cosas) (Última edición 24 de enero de 2014. Consultado el 14 de febrero de 2014).

Esta Web al Cuadrado se convierte en una magnífica fuente de datos; y la idea de Web Semántica o Web 3.0, a la que ya se refería Berners-Lee en 1998 como Web de datos global cuya finalidad no es sólo ser útil para la comunicación entre humanos sino también entre máquinas, adquiere su sentido pleno a través del conocido como *Linked Data*, que posibilita enlazar información externa con datos estructurados unidos, a su vez, con otros datos estructurados. Hasta tal punto que algunos autores hablan, indiferentemente, de Web semántica, Web 3.0 o Web de los datos enlazados.

Los datos, masivos, han dejado, por tanto, de ser dependientes o independientes para convertirse en interdependientes (Minelli *et al*, 2013: 13), y asistimos a “the Age of *Big Data Analytics*” (op. cit.: 3), como nueva forma de tratamiento y análisis de datos ante esta complejidad (Calero, 2013). Una fórmula que abre nuevas vías, en cuanto a que promueve el desarrollo de técnicas y herramientas para la investigación en Ciencias Sociales y Humanidades Digitales (en adelante, CSyHD). Al tiempo que se erige, por su novedad y por tratarse de un fenómeno prácticamente inexplorado desde una perspectiva académica, en objeto de estudio en sí mismo.

## **2. Planteamiento: objetivos y metodología**

El presente capítulo plantea un análisis exploratorio, desde una perspectiva amplia, descriptiva y cualitativa, del *Big Data*, partiendo, entre otros, de estos objetivos:

1. Acotar su significación, conforme al contexto actual y diferenciándolo de otros relacionados con el auge de los datos en red.
2. Identificar colectivos relacionados con la difusión y el análisis del fenómeno del *Big Data*, así como los nuevos perfiles profesionales y de investigadores que, como el llamado científico de datos, comienzan a demandarse.
3. Explorar la incidencia del *Big Data* en la llamada Ciencia de los Datos y su función en la Nueva Ciencia de Redes.
4. Describir técnicas de procesado y representación de datos masivos en red, como el llamado *Visual Analytics*, que facilitan la labor de investigadores.

5. Detectar y analizar tendencias y casos significativos de uso del *Big Data* en los principales campos de conocimiento de las CSyHD.

Además del análisis bibliográfico, mediante diversas fuentes hemerográficas online se han localizado artículos de académicos, entrevistas a expertos o reseñas de iniciativas significativas de aplicación al *Big Data* que aportan luz en torno al estado de la cuestión del fenómeno. A partir de estas fuentes se han localizado y analizado casos relevantes de proyectos e iniciativas basados en *Big Data*, considerando variables como las siguientes:

- Función del *Big Data*. Se diferencian los casos en los que éste se emplea como materia prima para el desarrollo de metodologías o herramientas (como sistemas de visualización de datos online), de aquellos en los que se convierte en objeto de análisis en sí mismo.
- Disponibilidad de resultados/herramientas en red y, por ende, potencial replicabilidad por otros usuarios. Se observa en qué casos es posible, además, su uso o reutilización para la *e-Research*, cuestión fundamental para que la ciencia sea acumulativa.
- Destinatarios de los proyectos. Se analiza si se trata de iniciativas específicamente orientadas a investigadores; o si se dirigen a la ciudadanía en general o a colectivos específicos (por ejemplo, periodistas), pero son también de utilidad, por su naturaleza y por estar en abierto, en el ámbito de la *e-Research*.
- Promotores del proyecto. Se distinguen los surgidos en el seno de proyectos/grupos oficiales de investigación en universidades u otros centros educativos; de aquellos que son resultado de colectivos al margen de la universidad, ya sean por iniciativa de determinadas entidades (fines comerciales), ya sean proyectos colaborativos (normalmente, sin ánimo de lucro) a cargo de colectivos ciudadanos, como activistas online.

Por cuestiones de espacio se exponen únicamente los casos más relevantes. Algunos se citan a lo largo del artículo. Mientras que el resto, considerando, además, su finalidad y campos de aplicación, se expone sobre una taxonomía de posibles aplicaciones del *Big Data* para la investigación en CSyHD.

### **3. Desarrollo**

#### **3.1. *Big Data*: concepción y auge en el contexto actual**

Las primeras alusiones al *Big Data* lo describían como “data that exceeds the processing capacity of conventional database systems” (Dumbill, 2012: 23), y de ahí que, siguiendo sus palabras, “to gain value from this data, you must choose an alternative way to process it” (ibid.). Aunque se trata de un concepto aún, incipiente, la idea de *Big Data* trasciende hoy el mero concepto de datos masivos o macrodatos, a veces equiparados.

La mayoría de expertos y académicos coinciden en señalarlo como especie de combinación entre algoritmos, tecnologías y estrategias, capaces de capturar y analizar grandes volúmenes de datos provenientes de múltiples fuentes heterogéneas a una alta velocidad, de forma económica y a tiempo real (Paniagua, 2013), para convertirlos en conocimiento. Muchos se basan en las cuatro dimensiones, conocidas como V's del *Big Data*, postuladas por IBM, una de sus grandes propulsoras, hacia 2012: “Volume”, “Velocity” y “Variety” inicialmente, a las que sumaría más adelante “Veracity”, incidiendo en la necesidad de que los datos sean correctos, como puede apreciarse en las infografías que la empresa ofrece desde su sitio web (IBM, s.d.).

A estas cuatro cabría agregar una quinta, sugerida por el periodista Mario Tascón en el dossier que la revista *Telos* dedicaba, a inicios de 2013, al *Big Data*: “Visualization”, también porque “muchas de las imágenes que nos traen a la memoria el trabajo con *Big Data* tienen que ver con estas nuevas formas de ‘ver’ estos datos” (Tascón, 2013: 47). Precisamente por la importancia que adquiere su representación visual, algunos califican la nueva Ciencia de los Datos producto del *Big Data*, descrita más adelante, como híbrido entre la informática tradicional, las matemáticas y el arte.

Así, por ejemplo, desde que en 2010 adquiriera Metaweb, empresa creadora de Freebase -una ingente base de datos abierta-estructurada y líder en Web Semántica- (Editorial RWWES, 2010), el propio Google viene realizando numerosos experimentos basados en *Big Data* y en sistemas interactivos de visualización de datos online. Entre ellos, el Gráfico del Conocimiento, una base de conocimiento creada en 2012 que, basada en la semántica, mejora los resultados del motor de búsqueda<sup>77</sup>. También destacan Google Flu y Dengue Trends, que, respectivamente, monitorizan personas con síntomas del virus H1N1 casi en tiempo real o muestran datos sobre la evolución del dengue en el mundo a través de sus búsquedas, y que se engloban en el proyecto Google.org<sup>78</sup>.

O el llamado Our Mobile Planet<sup>79</sup>, que además de proveer de información sobre la distribución de consumo de móviles a nivel global permite consultas personalizadas y que el usuario cree gráficos personalizados con los datos que le interesen. Google ha desarrollado, además, la tecnología subyacente a lo que se conoce como Hadoop, herramienta crucial para trabajar con datos masivos surgida en 2006 y que supuso, para muchos, el despegue del *Big Data*<sup>80</sup>.

Y desde 2012 y como parte de su “cloud platform” ofrece, además, Google BigQuery, un servicio de análisis de *Big Data* sobre bases de datos en la nube, como la Wikipedia, que procesa, a partir, por ejemplo, de búsquedas por palabras, grandes volúmenes de información rápidamente.

Las aplicaciones de esta analítica inteligente de datos alcanzan hoy, no obstante, a sectores no necesariamente relacionados con la Economía Digital, como la medicina y la sanidad, la banca, el marketing y la

---

<sup>77</sup> <http://www.google.com/insidesearch/features/search/knowledge.html>

<sup>78</sup> <http://www.google.org/projects.html> (consultado el 14 de febrero de 2014).

<sup>79</sup> <http://www.thinkwithgoogle.com/mobileplanet/en-gb/> (consultado el 14 de febrero de 2014).

<sup>80</sup> Creada en 2006 y empleada hoy por numerosos sitios de redes sociales y entidades para sus espacios online, se trata de una multiplataforma implementada en Java que soporta aplicaciones distribuidas bajo una licencia libre y permite, por tanto, trabajar con miles de nodos y volúmenes de datos del orden de petabytes (Cfr. <http://es.wikipedia.org/wiki/Hadoop>, última edición 9 de octubre de 2013, consultado el 14 de febrero de 2014).

comunicación, la educación, las telecomunicaciones, la moda o la industria alimentaria.

De aquellas relacionadas con la investigación en Ciencias Sociales y Humanidades nos encargaremos más adelante. Pero el *Big Data* (en conexión con disciplinas relacionadas, como la nueva Ciencia de Redes) es útil, además, como herramienta para el análisis de organizaciones y la planificación estratégica de cualquier organización y, por ende, para reducir la incertidumbre ante la complejidad actual, puesto que además, permite, en cierta forma, trazar tendencias.

De ahí que numerosas consultoras en los últimos años se hayan especializado en ofrecer, empleando *Big Data*, servicios de análisis predictivos a nivel global; y que, como apuntaba García Cantero (2013: 51-56) en un artículo reciente bajo el título de *Infoexplosión. Nuevas estrategias de gestión de la información*, cada vez más entidades comiencen a adoptar “soluciones de *Big Data* que permitan no solo manejar grandes cantidades de información, sino también integrar distintos formatos de la misma y, por supuesto, hacerlo en tiempo real”<sup>81</sup>.

### **3.2. Movimientos impulsores del Big Data: Open Data, Gobierno Abiertoy transparencia de la información pública**

A la expansión del *Big Data* han contribuido, en cierto sentido, movimientos como el *Open Data*, referido a la libre disposición de datos online, sin restricciones de copyright y bajo condiciones técnicas que permitan su uso y reutilización en el sentido más amplio por otros usuarios.

Así como el surgimiento de plataformas o herramientas *OpenSource* en red que, como respuesta a la demanda de usuarios, permiten recoger,

---

<sup>81</sup> En la infografía de la consultora Fico (2013), disponible desde <http://bit.ly/1842rsk> (consultado el 14 de febrero de 2014), se resumen algunas de las posibles aplicaciones, tanto presentes como futuras, en este “Big Bang de la Analítica” de datos.

analizar y visualizar de forma síncrona estos datos<sup>82</sup>. Su aparición, pese a que la ingente cantidad de datos online escape aún, a priori, del control de los usuarios (Paniagua, 2013), ha impulsado iniciativas que abren la posibilidad al llamado *Big-Linked-Open Data*, esto es, a disponer de un ecosistema digital compuesto por grandes bases de datos interconectados y abiertos de acceso público<sup>83</sup>.

Sirvan como referencia la traducción de bases de datos públicos a estándares semánticos, que propician su apertura e interoperatividad, y la posterior puesta a disposición en red, mediante APIs, de *data sets* en abierto que pueden ser consultados y reutilizados por los ciudadanos desarrollados en el marco de programas de Gobierno Abierto y fomento de la transparencia pública<sup>84</sup>. Aunque los pioneros en desarrollarlos, en 2009, fueron Gran Bretaña y Estados Unidos, en España, además de actuaciones regionales en el País Vasco, Asturias o Cataluña, se dispone ya de portales de *Open Data* a nivel nacional que, como Datos.gob.es<sup>85</sup>. Junto a ello, colectivos activistas como Hazte Oír o Access InfoEurope vienen aportando nuevos conjuntos de datos abiertos de utilidad para ciudadanos y profesionales de diversos sectores. Es el caso del periodístico, donde, en los últimos tiempos, este *Open Access* de datos masivos ha impulsado el llamado periodismo de datos, cuyo producto final son, a menudo, sistemas interactivos de visualización de datos que pueden erigirse en instrumentos para la *e-Research*<sup>86</sup>. El propio Google ofrece, en la misma línea, utilidades como

---

<sup>82</sup> Un ejemplo de proyecto OpenSource que permite que cualquier usuario sin conocimientos técnicos específicos pueda reutilizar los datos publicados y compartirlos entre ellos, es GORQL, surgido en 2012 y cuya web, en el momento de escribir estas líneas, no está operativa (<http://gorql.com/>). Mientras que otras herramientas comerciales de visualización y análisis de datos online que empiezan a popularizarse, como Tableau Software, disponen también de soluciones para Big Data (<http://www.tableausoftware.com/solutions/big-data-analysis>) (consultado el 14 de febrero de 2014).

<sup>83</sup> A este término se refería, por ejemplo, Erik Mannens, investigador de la Universidad de Gent, en Bélgica, en un coloquio organizado por una universidad alemana en diciembre de 2013 (Universität Passau, 2013).

<sup>84</sup> Ello implica que se ofrezcan en formatos abiertos y editables, como .csv.

<sup>85</sup> <http://datos.gob.es/datos> (consultado el 14 de febrero de 2014).

<sup>86</sup> En ocasiones los medios emplean, para desarrollar estos sistemas, herramientas online que están también al acceso de investigadores, como

su explorador de datos públicos Google Data Public, que incluye diversas fuentes estadísticas oficiales<sup>87</sup>.

### 3.3. Nuevos perfiles profesionales e iniciativas y colectivos orientados a su análisis y divulgación

Consecuentemente, nuevos perfiles profesionales expertos en *Big Data*, como los científicos de datos o “data scientific”, aparecen entre las profesiones más demandadas de los próximos años, especialmente en el ámbito norteamericano, donde el *Big Data* tiene mayor desarrollo (Paniagua, 2012b). Aunque revisando la literatura no existe total coincidencia al describir su perfil, se trata de profesionales que, además de competencias técnicas específicas vinculadas a las matemáticas (estadística y programación), requieren de una personalidad determinada (curiosidad científica) y de otras competencias transversales (capacidad de comunicación, enfoque distinto en análisis y resolución de problemas y habilidad para trabajar en equipo, con profesionales como los expertos en información visual o los “artistas de datos”).

Las universidades y otras entidades comienzan a ofrecer formación orientada a la capacitación de estos profesionales, aunque lo cierto es que en los momentos de escribir estas líneas la mayoría son aún primeras ediciones. Están, además, centradas en la formación técnica, como vía de especialización profesional, no tanto en la capacitación de investigadores, y menos aún, en el campo de las CSyHD.

Paralelamente, en los últimos años, diversas instituciones vienen impulsando, a nivel mundial, el conocimiento en torno a *Big Data* y *Data Science*, mediante publicaciones y la organización de eventos científicos. Es el caso del grupo O’Reilly y sus Conferencias Strata, iniciadas en 2011 en California. En el caso español, existen ya algunas agrupaciones, como la *Asociación Española de Linked Data* (AELID)<sup>88</sup>, o colectivos de profesionales, como el grupo *Macrodatos*, con el objetivo

---

Tableau Public (<http://www.tableausoftware.com/es-es>) (consultado el 14 de febrero de 2014).

<sup>87</sup> <http://www.google.com/publicdata/directory> (consultado el 14 de febrero de 2014).

<sup>88</sup> <http://www.aelid.es/web/guest/home> (consultado el 14 de febrero de 2014).

de promover la divulgación y el conocimiento en torno al mundo de los datos masivos. Una tarea a la que también contribuyen fundaciones de grandes corporaciones, como Telefónica o BBVA, o universidades que, como la Oberta de Cataluña, comienzan a organizar congresos sobre esta temática<sup>89</sup>.

### **3.4. Exploración en torno al fenómeno *Big Data* desde la Academia**

Parece que, como apuntaban Viktor Mayer-Schönberger, profesor de la Universidad de Oxford, y Kenneth Cukier, editor de datos de la revista *The Economist*, en su obra *Big Data. La revolución de los datos masivos*, “ninguna esfera de la sociedad quedará intacta” a este fenómeno (2013). Tampoco el ámbito de la investigación. Como sostenía McCandless en una charla para Ted sobre la belleza de la visualización de datos, “los datos no son solo el petróleo del siglo XXI sino la tierra más fértil para que crezca nuestro conocimiento, creatividad e innovación” (2010). En la web de IBM se recogen algunos proyectos universitarios que, basados en soluciones *Big Data*, denotan el modo en que los propios investigadores académicos están comenzando a usar esta fuente de valor. Entre ellos, en el ámbito de las Ciencias Sociales y Humanidades, el *Language, Interaction and Computation Laboratory (CLIC)* de la Universidad de Trento en Italia, orientado al estudio de la comunicación verbal y no verbal tanto con métodos computacionales como cognitivos (Barranco, 2012).

Sin embargo, como fenómeno en sí mismo el *Big Data* ha sido poco explorado desde la perspectiva de la investigación académica (menos aún en lo referido a su potencial para la *e-Research*). Más allá de los informes, análisis o reflexiones sobre iniciativas determinadas, la mayoría son bien obras técnicas que aportan luz sobre su funcionamiento, bien ensayos desde la perspectiva del *Intelligence Business* o de los *Predictive Analytics* (Siegel, 2013).

En lo que respecta a las CSyHD, cabe destacar sobre todo la labor de determinados centros de investigación asociados a universidades que, como el ya referido Oxford Internet Institute (OII), de la Universidad

---

<sup>89</sup> Cfr. P. ej. Balcells, J. *et al.* (coords.) (2013).

de Oxford, aparecen, por su propia naturaleza, muy vinculados al *Big Data*.

El OII resulta significativo, en primer lugar, como escenario en el que éste se convierte en el propio centro de la investigación. Como ejemplo, el proyecto “Accessing and Using *Big Data* to Advance Social Science Knowledge”, coordinado por Eric Meyer, que, a partir del análisis de casos, tiene como finalidad, según lo publicado en la web del OII, descubrir “what are the social and scientific implications of large-scale 'Big Data' as it becomes more widely available to social scientists in academia, public institutions, and the private sector”<sup>90</sup>.

En segundo lugar, otros grupos de investigadores del OII vienen desarrollando proyectos basados en estrategias y técnicas de extracción, procesado y análisis de datos masivos, e incluso, sus propias herramientas y aplicaciones web basadas en *Big Data* (el caso de NameGenDev, de Hogan, para Facebook)<sup>91</sup>. Vinculados a éstos, vienen celebrándose seminarios, simposios u otras actividades de análisis y debate en torno al estado de la cuestión y a su posible utilidad para la investigación. Y en tercer lugar, desde la web del centro se ofrecen no sólo los resultados de estas actividades, en forma de publicaciones o recursos multimedia en abierto, sino que los propios sistemas empleados sirven a la institución<sup>92</sup> y quedan disponibles para su reutilización en futuros análisis por otros usuarios.

### **3.5. Implicaciones del *Big Data* para la *e-Research*: nuevas Ciencias y Técnicas**

#### **3.5.1. La llamada Ciencia de los Datos y su transformación con el *Big Data***

La Ciencia de los Datos o *Data Science* lleva existiendo, en el mundo de la computación e Internet, desde hace ya mucho tiempo. Pero es el contexto de la Web al Cuadrado, como se adelantó y como afirmaba

---

<sup>90</sup> <http://www.oii.ox.ac.uk/research/projects/?id=98> (consultado el 14 de febrero de 2014).

<sup>91</sup> <http://apps.facebook.com/namegendev/> (consultado el 14 de febrero de 2014).

<sup>92</sup> <http://www.oii.ox.ac.uk/vis/> (consultado el 14 de febrero de 2014).

el vicepresidente de Oracle, Sutherland, durante una visita a Madrid en 2012, cuando “El *Big Data* nos da la posibilidad de hacer un análisis mucho más certero (...). Un nivel más profundo en el entendimiento de nuestras organizaciones, es el siguiente paso en las analytics” (Carnelo, 2012).

Siguiendo a Paniagua, la “importancia del *Big Data* radica en la capacidad de transformar grandes volúmenes de datos en inteligencia” (2013: 94). En el sentido de que, aplicando los principios de la Ciencia de los Datos, se puede obtener información sobre tendencias en tiempo real y, también, adelantarse a futuras tendencias y generar modelos predictivos.

En uno de los artículos de su blog, esta autora recopila la visión de distintos expertos en torno a esta Ciencia y a la revisión conceptual generada, en parte, por la incidencia del *Big Data* y el papel del científico de datos. Así, para Vallante, uno de los fundadores de Wikibon, la Ciencia de los Datos puede ser dividida en cuatro partes esenciales: “Mining Data”, como recopilación y formateado de la información; “Statistics”, referida a su análisis; “Interpret”, como visualización de los datos a través de infografías, imágenes o gráficos; y “Leverage”, que alude, por último, a las implicaciones de estos datos representados (Paniagua, 2011).<sup>93</sup>

Al igual que la Minería de Datos, el *Big Data* utiliza, así, métodos de la Inteligencia Artificial y la Estadística para analizar los patrones de las bases de datos con las que trabaja (Tascón, 2013: 47-50). Y es gracias a tecnologías que, como se explica más adelante, permiten realizar representaciones visuales de procesos complejos o hacen abordable la consulta, online, de forma interactiva y personalizada de datos, cuando éstos promueven una nueva forma de comprender la realidad al alcance de científicos y ciudadanos.

---

<sup>93</sup> Paniagua referencia, además, las consideraciones de Loukides en torno a la idea de que “el futuro pertenece a las empresas y personas que puedan transformar datos en productos”, recogidas en el *e-book Big Data Now, Current perspectives* de O’Reilly Media (disponible en <http://shop.oreilly.com/product/0636920022640.do>) (consultado el 14 de febrero de 2014).

### 3.5.2. La sinergia entre *Big Data* y la Nueva Ciencia de Redes

La disponibilidad de datos de redes de gran magnitud, como consecuencia de la eclosión del *Big Data*, juega, asimismo, un papel esencial en lo que autores como Calero denominan Nueva Ciencia de las Redes, disciplina científica emergida en la última década que “estudia cómo las redes emergen en la naturaleza, la tecnología y la sociedad” (2013: 87). El análisis de estos datos mediante el empleo de determinadas soluciones tecnológicas posibilita, así, “el estudio experimental de las mismas, así como la comprobación de la validez de determinadas leyes comunes”, de forma más ágil (ibid.). De esta forma, se identifican patrones y parámetros de centralidad o se descubren anomalías y además se pueden efectuar simulaciones sobre estos parámetros para, en fin, conocer su posible evolución.

Se trata, siguiendo al mismo autor, de una ciencia horizontal, que puede aplicarse a todas las áreas y problemas, más aún ante la proliferación de *data set* para el modelado de datos y la simulación y de tecnologías de visualización automática e inteligente de estas redes, online. Ejemplo de ello es el estudio realizado por Esteban Moro y otros investigadores de la Universidad Carlos III que, sobre un experimento real de IBM y una modelización matemática, analiza cómo influye el comportamiento de los internautas en la difusión de información a través de redes sociales (Iribibarren y Moro, 2009).

### 3.5.3. *Big Data* y análisis de redes en la investigación aplicada

El referido análisis de redes se está empezando a usar, más allá de la investigación básica, como herramienta para la investigación aplicada en diversos sectores profesionales. Especialmente en la sanidad, donde el *Big Data* posibilita predecir, prevenir y personalizar (Feldman *et al.*, 2012). O en la comunicación y marketing digital, donde permite monitorizar a tiempo real la reputación digital de una entidad o la visibilidad de determinados mensajes en redes sociales y comunidades virtuales, detectar temas que interesan al público o líderes de opinión asociados a ciertos asuntos, analizar preferencias y hábitos de consumo, o mejorar la gestión de las relaciones con los clientes (CRM) y ofrecerles servicios o productos personalizados. Como apunta Suta (2013), además de “datos de interacción” (gustos y comportamientos), muchas empresas disponen ya de *Big Data* propios

que incluyen “transacciones tradicionales del cliente” (compras, devoluciones, reclamaciones...).

Su aplicación se extiende también, según se adelantó, al mercado de valores o al ámbito empresarial, como técnica para el análisis de las organizaciones o la planificación estratégica; e incluso, para la solución práctica de problemas sociales que posibilita, por ejemplo, combatir con el ciberterrorismo o los cibercrimes mediante la identificación y seguimiento en tiempo real de patrones anómalos online. Se trata, en síntesis, de una ciencia cuya visión, “en clave de red”, es diferente, y que permite “un nuevo tratamiento de la complejidad” (Calero, 2013: 87) que, en un mundo cada vez más interconectado, reduce el nivel de incertidumbre y ayuda, por tanto, a la toma de decisiones a directivos y profesionales.

Las soluciones, algunas gratuitas, que emergen, en la red, contribuyen a su uso masivo por parte de directivos, profesionales, ciudadanos e investigadores. Algunas aplicaciones trabajan exclusivamente datos disponibles online, pero otras permiten a los usuarios introducir y almacenar nuevos datos para su análisis, seguimiento y visualización. En la medida en que esto suceda, la cantidad de macrodatos continuará su crecimiento exponencial online. Y cuando los resultados científicos queden publicados en abierto, se contribuirá a que la Ciencia sea acumulativa.

#### **3.5.4. *Visual analytics* como técnica de visualización y análisis de grandes cantidades de datos**

Ante el auge del *Big Data* y la referida aplicación de la Nueva Ciencia de las Redes, emergen, como se ha comentado, diversas utilidades que extraen información de cuestiones complejas y la representan visualmente, facilitando su comprensión para nuestro estudio o toma de decisiones. Se basan en los llamados “Visual Analytics” o Análisis Visuales, que Thomas y Cook definían, así, como “la ciencia del razonamiento analítico facilitada por interfaces visuales interactivas”, y que, como campo multidisciplinar, se relaciona con diversas áreas:

- “analytical reasoning techniques that let users obtain deep insights that directly support assessment, planning, and decision making;

- visual representations and interaction techniques that exploit the human eye's broad bandwidth pathway into the mind to let users see, explore, and understand large amounts of information simultaneously;
- data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis; and
- techniques to support production, presentation, and dissemination of analytical results to communicate information in the appropriate context to a variety of audiences" (2005: 10-11)

Como recogen Alonso y García Olaizola (2013: 49-52) en un monográfico sobre el tema, existen diversas formas de visualización. Entre las más relevantes, las “bases de datos orientadas a grafos”, en las cuales la información se representa a través de nodos (el caso de las gráficas de diagramas o barras interactivas); la “información geoposicionada”, que parte de cartografías a las que se agregan grandes cantidades de datos y otros elementos; y la “información espacio-temporal”, esto es, representaciones visuales cíclicas o lineales de la evolución de hechos o fenómenos.

Aunque se trata de fórmulas tradicionalmente empleadas en áreas como la meteorología o las ciencias medioambientales o de la tierra, o la economía y las finanzas, entre otras, hoy son de uso habitual también, en ámbitos de las Ciencias Sociales y Humanidades, como la Sociología, Política o Comunicación. En gran medida por la proliferación, gracias al HTML 5, de herramientas de visualización web, del tipo Bird Eyes (sobre Adobe Flex), adecuadas para la visualización científica y *Visual Analytics* en múltiples dispositivos.

### **3.6. *Big Data* y *e-Research* en CSyH: posibles aplicaciones, objetos de estudio y radiografía de casos**

#### **3.6.1. Estudios macroscópicos: macrodemografías, macroanálisis de contenidos...**

En una conferencia organizada en 2012 por el OII bajo el título de “Internet, politics and policy: *Big Data*, bigchallenges?”, Duncan

Watts, de Microsoft Research, aseguraba que el modo en que el *Big Data* podría estar cambiando las ciencias sociales “is that it allows researchers conducting massive observation all field studies on macro-level, about collective phenomena, which could hardly be studied other wise” (Poutanen, 2012)<sup>94</sup>.

Es evidente, así, que el *Big Data* posibilita estudios “macroscópicos” que necesitan muchos datos para alcanzar sus objetivos o un elevado muestreo para hacerse extrapolables. Es el caso de demografías y análisis cuantitativos de perfiles en ámbitos como la sociología, la educación o la comunicación, entre otros, sobre los que pueden establecerse hipótesis correlacionales en torno a determinadas variables, describiendo tendencias presentes o anticipándose a las futuras. El *Big Data* sistematiza y automatiza, casi siempre, siguiendo a Watt, mediante variables previamente programadas por humanos, la obtención de datos. Así operan, por ejemplo, los servicios de análisis de *Big Data*, que, sobre bases de datos en la nube, permiten procesar grandes volúmenes de información en muy poco tiempo (como el ya comentado caso de Google BigQuery, sobre la Wikipedia). Sin embargo, el resultado es un contenido que, siguiendo a Watt, en ocasiones “is hard to analyze and understand”.

Especialmente porque, más allá de estudios puramente cuantitativos, el *Big Data* también puede llegar a emplearse para macroanálisis que contemplan lo cualitativo, como serían, en el área de los contenidos, las referidas al segundo nivel de *agenda setting* en noticias (análisis del discurso), o los comentarios de usuarios en redes sociales. La recogida y análisis de datos se hace a través, por ejemplo, de frecuencias de palabras (términos positivos o negativos), programadas a partir de los criterios de investigadores expertos en esas disciplinas. La presencia de estos expertos, más allá de para esta operacionalización y categorización de variables, resulta esencial también para otorgar luego un sentido a los macrodatos obtenidos, conforme al contexto y complementándolos, cuando corresponda, mediante el uso de otras técnicas. En análisis centrados en el comportamiento o en la actividad

---

<sup>94</sup> En esa ocasión (Poutanen, 2012) fueron objeto de reflexión cuestiones sobre a partir de qué tamaño de datos se considera *Big Data* (“what is big now, won't be in a few yeears”), o sobre qué cantidad son necesarios en una investigación (“big depends on the research questions”).

de los usuarios online o en investigaciones basadas en la simulación y el modelado de datos a través de sistemas de visualización online, esta interpretación posibilita el trazado de posibles escenarios futuros. Como apuntaba Boyd, investigadora de Microsoft y profesora en la Universidad de New York, “we must continue to ask ‘why’ questions that cannot be answered through traces alone, that cannot be elicited purely through experiments” (2010).

La creciente aparición de software que, orientado al análisis cualitativo, automatiza el macroanálisis extrayendo contenido generado en plataformas de *social media* la vez que permite luego tratar e interpretar, conforme al contexto, esa información por expertos (el caso de NVivo 10), facilita esta labor.

### **3.6.2. Exploraciones sobre la marca digital e influencia de personas o entidades en redes sociales**

En la última etapa han comenzado a proliferar las investigaciones que, en campos como la comunicación, persiguen el estudio de la identidad y reputación digital de personas o empresas, o de su influencia, como posibles líderes de opinión, en redes sociales. El *Big Data*, conforme integre datos correspondientes a variables en torno al perfil de los usuarios, al contenido publicado por éstos y por otros sobre los mismos o a los nodos e interrelaciones, entre otras, se erigirá como una herramienta de gran utilidad para la *e-Research* en este sentido.

Existen, de momento, herramientas y aplicaciones online, no orientadas exclusivamente a investigadores pero que pueden servir para análisis cuantitativos sobre perfiles de usuarios, su actividad y conexiones o, incluso, su influencia en red. Entre las gratuitas, Klout<sup>95</sup>, basada en *visual analytics*. Emplearlas requiere conocer la forma en que miden los datos, para interpretar, con las correspondientes limitaciones, los resultados; y aunque son poco flexibles, son útiles para hacer comparativas exploratorias.

---

<sup>95</sup><http://klout.com/home> (consultado el 14 de febrero de 2014).

### 3.6.3. Análisis de percepciones, clima social u opinión pública mediante sistemas de escucha activa online

En la medida en que puede recogerse y analizarse la percepción que los usuarios online tienen sobre determinadas personas o entidades, el *Big Data* se conforma en herramienta que facilita los macroanálisis sobre clima social u opinión pública, de gran utilidad en el ámbito de la sociología y la comunicación política.

En el segundo caso, especialmente desde 2012 y con motivo de diversas elecciones nacionales e incluso autonómicas se han desarrollado sistemas métricos que, basados en *Big Data*, ofrecen una visión inmediata del impacto de las campañas electorales a través de sistemas de visualización; y hay diversas herramientas que ofrecen rankings de popularidad e influencia de políticos y otros personajes públicos<sup>96</sup>. Frente a las anteriores, que proporcionan análisis cuantitativos, los denominados sistemas de escucha activa indagan en la polaridad o el sentimiento de los mensajes en torno a determinadas campañas políticas. Sería el caso del creado para las elecciones catalanas de 2012, Eleccions25n.cat, que funciona mediante rastreo, etiquetado automático de información por sentimiento e idioma y posterior revisión del etiquetado por analistas<sup>97</sup>.

La fiabilidad, especialmente de aquellas que realizan estos análisis cualitativos de forma totalmente automatizada, es menor, y en cualquier caso, la necesidad de la intervención humana limita el tamaño de los análisis. Quizás ante las próximas elecciones estadounidenses o europeas el avance del *Big Data* y la Web semántica posibilite el desarrollo de sistemas de escucha activa que ofrezcan

---

<sup>96</sup> Como herramientas para una visión inmediata del impacto de determinadas campañas electorales pueden citarse, por ejemplo, Elections20 (<http://elections20.com>) para las elecciones federales de Alemania de 2013, o la utilidad “elections” de Socialbakers (<http://www.socialbakers.com/elections>), para las presidenciales de Estados Unidos de 2012. Mientras que entre las que ofrecen rankings a partir del impacto/popularidad de los políticos en redes sociales, puede citarse el caso, en España, Politweets.es. Esta herramienta nació a partir de las elecciones locales de 2011 y que, tras cambiar de dueños en agosto de 2013 no está, en el momento de escribir esta consulta, operativo (Cfr. <http://pabloroman.es/blog/2012/12/26/se-vende-politweets-es/>) (consultado el 14 de febrero de 2014).

<sup>97</sup> <http://eleccions25n.cat/> (consultado el 14 de febrero de 2014).

microanálisis válidos y exhaustivos en el ámbito de la comunicación política y campañas electorales.

Con todo, son muchos los expertos que coinciden en que ésta es una de las grandes áreas de expansión del *Big Data*. También para los propios políticos, en el sentido de que permite, en parte, controlar o anticiparse al voto de un país. Las visiones de diversos expertos y académicos que se refieren acerca de la forma en la que el *Big Data* contribuyó, supuestamente, a la victoria de Obama en las elecciones norteamericanas de 2012, mediante el análisis de grandes bases de datos online, son muestra de ello<sup>98</sup>.

#### **3.6.4. Investigación sobre audiencias sociales**

Ya se comentó como, mediante sistemas de *Big Data*, pueden llevarse a cabo macroanálisis de perfiles y comportamientos de los usuarios, de gran utilidad en el ámbito de la investigación aplicada en campos de las Ciencias Sociales como la comunicación, la sociología o la antropología, entre otros.

Así, Gallego Calonge reflexionaba, en un artículo, sobre la forma en que, en un contexto de consumos transmediáticos, “la conexión entre el mundo de la televisión y Twitter permite extraer análisis cualitativos casi en tiempo real a través del procesado de ingentes cantidades de nuevas métricas” (2013: 68), y apuntaba la necesidad de contar con nuevos perfiles que, como el analista digital, se encarguen de convertirlos en información valiosa “para clientes, plataformas publicitarias, agencias y empresas de comunicación”. La investigación de las audiencias sociales se conforma, así, como un interesante complemento de los tradicionales sistemas de audiometrías, y recientemente comienzan a surgir herramientas orientadas a su

---

<sup>98</sup>Al respecto, por ejemplo, en una monografía sobre las elecciones estadounidenses de 2012, Antonio Gutiérrez Rubí (2012) sitúa el Big Data entre los “12 factores decisivos” de la victoria de Obama, en el sentido de que “por primera vez, se logró no solo recopilar sino organizar y convertir en votos una cantidad de datos de todo tipo sobre cada votante”, hecho que, para este autor, se convertirá en un referente para las próximas elecciones presidenciales en todo el mundo.

monitorización, como Trend TV, lanzado en octubre de 2013 por Vertele<sup>99</sup>.

### **3.6.5. Estudios antropológicos y sociológicos con *Big Data*: comprender el comportamiento humano**

A este respecto, resulta significativo “The Human Face of *Big Data*”<sup>100</sup>, iniciativa colaborativa arrancada a finales de 2012 por Rick Smolan, creador de la serie “Day in the Life” con el patrocinio de Cloud Computing EMC y que, bajo el eslogan “learn about yourself by exploring *Big Data*” y a través de una red de voluntarios que fueron captando y compartiendo online, datos cotidianos, pretendía demostrar que mediante el análisis de datos y con la ayuda actuales, se pueden extraer gran cantidad de informaciones que nos ayuden en nuestra vida diaria, y cuyo análisis posibilita, además, predecir múltiples situaciones futuras<sup>101</sup>.

### **3.6.6. Estudios geolocalizados y mapeados sobre usuarios en redes sociales para analizar posibles correlaciones**

Aunque existen varios proyectos en esta línea, citaremos, como caso significativo en el ámbito de la investigación sociológica y que enlaza, también, con otras disciplinas más allá de las Ciencias Sociales, el denominado “Hedonometerproject”. Se trata de una aplicación que, desarrollada por el Centro de Computación Avanzada de la Universidad de Vermont, utiliza el *Big Data* y los datos de redes sociales para analizar la felicidad de los ciudadanos. El sistema rastrea más de 400 palabras clave, relacionadas con estados de ánimo, enviadas por ciudadanos, principalmente de Estados Unidos, desde Twitter y las presenta luego a través de un sistema de visualización que incluye un ranking por palabras, la posibilidad de descubrir

---

<sup>99</sup> <http://www.vertele.com/trendtv/> (consultado el 14 de febrero de 2014).

<sup>100</sup> <http://thehumanfaceofbigdata.com> (consultado el 14 de febrero de 2014).

<sup>101</sup> El proyecto, cuyos productos son un documental y un libro, se basaba, así, en la premisa de que la visualización en tiempo real de datos recopilados por satélites, millones de sensores, etiquetas RFID, y *smartphones* y cámaras con GPS en todo el mundo permite a la humanidad percibir, calcular, comprender e influir en aspectos de la existencia como nunca se hubiera imaginado.

niveles de felicidad de forma geolocalizada (a través de un mapa con zonas más alegres o más tristes), de correlacionarla con otras variables, como la salud de los habitantes de estas zonas o la felicidad/tristeza según el tweet fuera enviado cerca o lejos del hogar<sup>102</sup>, o de filtrarla por fechas (como curiosidad, el día más triste hasta ahora ha sido el de los atentados de Boston).

### **3.6.7. Herramienta para la investigación sobre e-learning y enseñanza-aprendizaje en red**

El *Big Data* ha sido también aplicado para la investigación en educación, especialmente porque, cuando provee datos detallados sobre la actividad de usuarios en entornos de enseñanza-aprendizaje en red, con los resultados de su análisis se intenta proporcionar una experiencia más personalizada a las necesidades y demandas de los estudiantes, contribuyendo por tanto a la creación de los llamados *personal learning environments* (PLEs). Empresas dedicadas a la formación, como Knewton y Junyo, han desarrollado ya algunos sistemas en esta línea (Paniagua, 2012a), mientras que los propios gestores de formación comienzan a experimentar con métricas basadas en *Big Data* en plataformas como Moodle (Reig, 2013). Y más allá, nuevas tendencias educativas como los llamados MOOCs, cursos masivos online en abierto, suponen una magnífica oportunidad para los investigadores en e-learning, empleando técnicas basadas en *Big Data*. Este fenómeno se conoce como *Data Mining in Education* o, más recientemente, *Learning Analytics*, y ha sido objeto de estudio por diversos investigadores (p.ej. Romero y Ventura, 2013)

### **3.6.8. Herramienta para la estrategia y prospectiva: modelos predictivos basados en *Big Data* en CSyHD**

Además de las iniciativas reseñadas, cuyos resultados pueden ser de utilidad, en muchos casos, para análisis estratégicos y prospectivos, y de las propias herramientas que, entidades como Google, han desarrollado y están al alcance de los usuarios, existen varias investigaciones que precisamente inciden en el potencial del *Big Data* para anticiparse al futuro.

---

<sup>102</sup>Cfr. Wakefield (2013).

Así, el grupo de investigadores británicos y estadounidenses liderados por Preis y Mota, de la Boston University publicaban en la revista *Nature*, en 2012, los resultados de un trabajo que comparaba el número de búsquedas en Google relativas a 2009 y a 2011 realizadas por usuarios de 45 países durante 2010. Concluyeron entonces que los países con mayor PIB (Alemania, Suiza, Japón, etc.) tienden a realizar búsquedas online sobre el futuro, mientras que los de economías menos desarrolladas hacen lo propio sobre el pasado (Preis *et al.*, 2012).

Poco después se recoge, también en *Nature*, otra investigación del mismo equipo que asegura que las búsquedas online de términos financieros actuaron, en escenarios como la crisis de 2008, como “alertas tempranas” de sacudidas de los mercados (Preis *et al.*, 2013).

A partir de la selección de casi un centenar de términos de búsqueda, de entre los más empleados por el *Financial Times* y los sugeridos por Google Trends, y de su evolución, simularon estrategias de inversión de distintos portafolios entre 2004 y 2011. El resultado fue que la estrategia basada en la evolución de búsquedas en Google del término “deuda” fue la que mejor se comportó, en cuanto a rentabilidad de la inversión. Para los investigadores, los datos de la herramienta de Google, así, no sólo reflejaban el estado de los mercados financieros en ese momento (seguimiento a tiempo real), sino que también podrían servir para anticipar tendencias futuras.

Con todo, después de “deuda”, aparecían búsquedas sobre términos como “color” o “restaurante”, resultados que, como publicaba el diario *Abc* en una reseña sobre este estudio, “ilustran también las limitaciones del *Big Data* y los elevados niveles de ruido estadístico” que dan lugar a conclusiones, en ocasiones, irrelevantes en este tipo de investigaciones (Bergareche, 2013). De ahí que muchos se muestren escépticos en su utilidad para predecir los movimientos de la economía.

#### 4. Discusión y conclusiones: oportunidades y retos del *Big Data* en CSyHD

Como consecuencia del *Big Data*, el *Linked Data* y el *Open Data*, asistimos a una web de datos abiertos masivos, inteligentes, conectados a tiempo real y en continua interacción. En el ámbito que nos ocupa, el de las Ciencias Sociales y Humanidades Digitales, ello incide en varios aspectos.

Primero, y como muestran los casos reseñados, la era del *Big Data* Análisis supone una transformación sobre el modo de investigar online, con la posibilidad de emplear técnicas de análisis inteligente de datos masivos sobre diversos fenómenos así como sistemas de visualización en red, ya sea como fuentes para extraer datos, ya sea para representar resultados de investigaciones sobre datos masivos.

Segundo, desde la perspectiva de los investigadores, se hace necesario que éstos desarrollen nuevas competencias, incluyendo la asunción de aspectos éticos asociados al *Big Data*. Y sobre todo, el *Big Data* implica la colaboración entre áreas y profesionales de diverso perfil. No sólo porque con frecuencia se trata de investigaciones complejas o en las que resulta necesario desarrollar soluciones tecnológicas propias, que requieren incorporar, también en el ámbito de las ciencias sociales y humanidades, a expertos en *Big Data* o a científicos de datos. Sino también porque, al contrario, los análisis sobre *Big Data* en otras disciplinas necesitan, a menudo, de la colaboración de científicos sociales para aportar valor y significado, mediante un análisis del contexto, a los meros datos.

Algunas voces sostienen que la utopía de la Web semántica de Berners Lee, como base de datos global del conocimiento humano que permite, además, generar nuevo conocimiento, empieza a hacerse realidad, al vislumbrarse una posible puesta en común y aprovechamiento de recursos que distintos individuos y colectivos aportan a la Web (Reig, 2010), incluyendo a investigadores.

Hay, sin embargo, determinadas circunstancias que limitan, aún, la expansión del *Big Data* y su empleo en el ámbito de la *e-Research*. Entre ellas, la escasez de datos abiertos en determinados contextos, a lo que contribuye en parte, siguiendo el artículo previo a la conferencia Strata 2014 de Croll (2013), a modo precisamente de prospectiva sobre *Big Data*, las barreras culturales “to data-driven change” y el

retraso y temores en adoptar las correspondientes tecnologías por parte de determinadas entidades.

Es, además, necesario reflexionar sobre cuestiones relacionadas con la solvencia de los datos accesibles desde sistemas de consulta de datos masivos o de visualización online, muchos aportados por fuentes no oficiales o entidades con determinados intereses, para lo que resulta fundamental disponer, en todo momento, de información sobre su procedencia. No menos relevante es el filtrado de datos, tarea compleja que nos lleva a insistir, de nuevo, en la importancia de los expertos.

Otro dilema está relacionado con la escasa fiabilidad o validez de los resultados de determinadas investigaciones basadas en *Big Data*, especialmente análisis prospectivos; o el riesgo de hacer interpretaciones erróneas limitando las conclusiones a lo cuantitativo. Un factor que hace imprescindible la llamada, siguiendo la terminología del referenciado artículo sobre Strata 2014, “antropología de datos”. El ejemplo de las inundaciones de Haití expuesto en el mismo resulta significativo: sus datos mostraban que la población no se mudaba a una zona más segura. Se concluyó que todos debían estar enfermos de cólera y por ello no se mudaban, pero la realidad es que las carreteras estaban destruidas y no podían, por tanto, salir (Croll, 2013).

Algunas voces señalan que la clave para aumentar la fiabilidad de los análisis reside en el llamado *Small Data*, esto es, “the amount of data you can conveniently store and process on a single machine, and in particular, a high-end laptop or server” de acuerdo a un artículo publicado, en abril de 2013, sobre esta cuestión en el blog de la Open Knowledge Foundation (Pollock, 2013b). Se trata de pequeños *datasets* que pueden ser almacenados y procesados a través de un simple ordenador personal, y en los que se basan, por ejemplo, las iniciativas de mapeo online realizadas por colectivos activistas o de ciudadanos a nivel hiperlocal.

Desde esta perspectiva, el *Small Data* podría entenderse, bien como un instrumento de visualización y empoderamiento colectivo del *Big Data*, bien como la última fase de los procesos asociados a éste, en la cual los grandes volúmenes de datos se traducen en aplicaciones concretas asociadas a determinados objetivos (Pollock, 2013a). En

cualquier caso, ambos modelos, *Small* y *Big*, no son excluyentes, y su combinación abre nuevas posibilidades a la *e-Research* sobre Data.

## Referencias bibliográficas

Alonso Cerrato, J.& García Olaizola, I. (2013). “Análisis visual en el campo del Big Data Visual Analytics”. *Bit*, 193: 49-52. Disponible en: <http://www.coit.es/publicaciones/bit/bit193/PDF/monograficojavier.pdf> (consultado el 1 de diciembre de 2013)

Balcells, J. *et al.* (coords.) (2013). *Big Data. Retos y Oportunidades*. Actas del IX Congreso Internacional Internet, Derecho y Política. Universidad Oberta de Catalunya, Barcelona, 25 y 26 de junio de 2013: [http://edcp.uoc.edu/proceedings\\_idp2013.pdf](http://edcp.uoc.edu/proceedings_idp2013.pdf) (consultado el 14 de febrero de 2014).

Barranco Fragoso, R. (2012). “¿Qué es el Big Data? Todos formamos parte de ese gran crecimiento de datos”. *IBM.com*, 18 de junio de 2012. Disponible en: <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/> (consultado el 1 de diciembre de 2013).

Bergareche, B. (2013). “Los límites del Big Data”. *ABC.es*, 11 de mayo de 2013. Disponible en: <http://www.abc.es/sociedad/20130511/abci-google-data-201305102129.html> (consultado el 14 de febrero de 2014).

Boyd, d. (2010). “Big Data: Opportunities for Computational and Social Sciences”. *Zephoria.org*, 17 de abril de 2010. Disponible en: <http://www.zephoria.org/thoughts/archives/2010/04/17/big-data-opportunities-for-computational-and-social-sciences.html> (consultado el 1 de diciembre de 2013).

Calero, A. (2013). “Nueva Ciencia y tecnología de Redes. Una nueva forma de tratar la complejidad”. *Telos*, 95: 87-90. Disponible en: [http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/REVISTA/Dossier/DetalleArticuloTELOS\\_95TELOS\\_DOSSIER\\_PV1/seccion=1266&idioma=es\\_ES&id=2013062110120002&activo=6.do](http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/REVISTA/Dossier/DetalleArticuloTELOS_95TELOS_DOSSIER_PV1/seccion=1266&idioma=es_ES&id=2013062110120002&activo=6.do) (consultado el 1 de diciembre de 2013).

Carnelo, E (2012). “El Big Data optimiza todo lo que toca”. *TICBeat.com*. 1 de abril de 2012. Disponible en:

<http://bigdata.ticbeat.com/big-data-optimiza-todo-toca/>  
(consultado el 14 de febrero de 2014).

Croll, A. (2013): “Predicting the future: Strata 2014 hot topics. Eleven areas of focus for deeper investigation”. *O’Reilly Radar*, 7 de agosto de 2013. Disponible en:  
<http://radar.oreilly.com/2013/08/strata2104hottopics.html>  
(consultado el 1 de diciembre de 2013).

Dumbill, E. (2012). “Big Data Market Survey”. *Planing for Big Data. A CIO's Handbook to the Changing Data Landscape*. O'Reilly Strata Making Data Work. Disponible en:  
<http://strata.oreilly.com/2012/03/planning-big-data.html>  
(consultado el 14 de febrero de 2014).

Editorial RWWES (2010). “Google adquiere Metaweb, un paso más hacia la Web semántica”. *TICBeat*, 19 de julio de 2010. Disponible en:  
<http://www.ticbeat.com/novedades/google-adquiere-metaweb-paso-hacia-web-semantica/> (consultado el 14 de febrero de 2014).

FICO (2013). “El big bang de la analítica” (infografía).  
<http://bit.ly/1842rsk> (consultado el 14 de febrero de 2014).

Gallego Calonge, F. (2013). “Twitter y televisión. El papel de la gestión analítica de las audiencias sociales”. *Telos*, 95: 68-76.  
Disponible en:  
[http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/REVISTA/Dossier/DetalleArticuloTELOS\\_95TELOS\\_DOSSIER3/seccion=1266&idioma=es\\_ES&id=2013062110120001&activo=6.d](http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/REVISTA/Dossier/DetalleArticuloTELOS_95TELOS_DOSSIER3/seccion=1266&idioma=es_ES&id=2013062110120001&activo=6.d)  
[o](#) (Consultado el 1 de diciembre de 2013).

IBM (s.d.). “The FOUR V’s of Big Data” (infografía).  
*IBM.com*. Disponible en:  
<http://www.ibmbigdatahub.com/infographic/four-vs-big-data>  
(consultado el 4 de febrero de 2014).

Iribibarren, J.L. & Moro, E. (2009). “Impact of human activity patterns on the dynamics of information diffusion”.  
*PhysicalReviewLetters*, 103. Disponible en:  
<http://markov.uc3m.es/~emoro/ps/PRL09.pdf> (consultado el 14 de febrero de 2014)

- Feldman, B.*et al.* (2012). “Big Data in Healthcare Hype and Hope”. Dr.Bonnie 360º: <http://es.scribd.com/doc/107279699/Big-Data-in-Healthcare-Hype-and-Hope> (consultado el 14 de febrero de 2014).
- Goitia, F. & Petry, A. (2013). “La revolución (o dictadura) de los datos 'big data'”. *XL Semanal*, 11 de agosto de 2013. Disponible en: <http://www.finanzas.com/xl-semanal/conocer/20130811/revolucion-dictadura-datos-data-5993.html#VwZ1rnWvPQUvcowF> (consultado el 14 de febrero de 2014).
- Gonzalo, M. (2013). ““Los datos masivos (o big data) son el nuevo oro”” (entrevista a Viktor Mayer-Schönberger). *Diario.es*, 5 de agosto de 2013. Disponible en: [http://www.eldiario.es/turing/Big-data\\_0\\_161334397.html](http://www.eldiario.es/turing/Big-data_0_161334397.html) (consultado el 14 de febrero de 2014)
- Gutiérrez Rubí, A. (2012). “eBook ‘Elecciones USA 2012: los 12 factores decisivos’”. *Gutierrez-rubi.es*. Disponible en: <http://www.gutierrez-rubi.es/2012/12/04/ebook-elecciones-usa-2012-los-12-factores-decisivos/> (consultado el 14 de febrero de 2014)
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data: la revolución de los datos masivos*. Madrid: Turner Noema.
- McCandless, D. (2010). “La belleza de la visualización de datos”. *TEDGlobal*. Disponible en: [http://www.ted.com/talks/lang/es/david\\_mccandless\\_the\\_beauty\\_of\\_data\\_visualization.html?embed=true](http://www.ted.com/talks/lang/es/david_mccandless_the_beauty_of_data_visualization.html?embed=true) (consultado el 1 de diciembre de 2013).
- Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. Wiley CIO, vol. 578. Disponible en: [http://ebookey.org/Big-Data-Big-Analytics-Emerging-Business-Intelligence-and-Analytic-Trends-for-Today-s-Businesses\\_2302463.html](http://ebookey.org/Big-Data-Big-Analytics-Emerging-Business-Intelligence-and-Analytic-Trends-for-Today-s-Businesses_2302463.html) (consultado el 1 de diciembre de 2013).
- O’Reilly, T. & Battelle, J. (2009). *Web Squared: Web 2.0 Five Years On*. En O’Reilly.com. Disponible en: [http://assets.en.oreilly.com/1/event/28/web2009\\_websquared-whitepaper.pdf](http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf) (consultado el 11 de marzo de 2013).
- Paniagua, S. (2011). “La Ciencia de los Datos (BD II)”. *Sorayapanigua.com*, 1 de noviembre de 2011. Disponible en:

<http://www.sorayapaniagua.com/2011/11/01/la-ciencia-de-los-datos-bdii/> consultado el 1 de diciembre de 2013).

Paniagua, S. (2012a). “Llega un nuevo modelo de formación online basado en big data”. *SorayaPaniagua.com*, 4 de marzo de 2012.

Disponible en: <http://www.sorayapaniagua.com/2012/03/04/llega-un-nuevo-modelo-de-formacion-online-basado-en-big-data/> (consultado el 14 de febrero de 2014).

Paniagua, S. (2012b): “Big Data, nuevas competencias y oportunidades de empleo”. *Sorayapanigua.com*, 10 de junio de 2012.

Disponible en: <http://www.sorayapaniagua.com/2012/06/10/big-data-nuevas-competencias-y-oportunidades-de-empleo/> (consultado el 14 de febrero de 2014).

Paniagua, S. (2013). “Un mundo de sensores. De los datos al Big Data”. *Telos*, 95: 94-96. Disponible en:

[http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/NMEROSANTERIORES/pocactual/DetalleAnteriores\\_95TELOS\\_DOSSIERPV3/seccion=1268&idioma=es\\_ES&id=2013062110130001&activo=6.do](http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/NMEROSANTERIORES/pocactual/DetalleAnteriores_95TELOS_DOSSIERPV3/seccion=1268&idioma=es_ES&id=2013062110130001&activo=6.do) (consultado el 30 de noviembre de 2013).

Pollock, R. (2013a). “Forget Big Data, Small Data is the real revolution”. *OpenKnowledge Foundation Blog*, 22 de abril de 2013.

Disponible en: <http://blog.okfn.org/2013/04/22/forget-big-data-small-data-is-the-real-revolution/> (consultado el 14 de febrero de 2014).

Pollock, R. (2013b). “What do we mean by Small Data”.

*OpenKnowledge Foundation Blog*, 26 de abril de 2013. Disponible en:

<http://blog.okfn.org/2013/04/26/what-do-we-mean-by-small-data/> (consultado el 14 de febrero de 2014).

Poutanen, P. (2012). “Social sciences going for big data”.

*Blogs.helsinki.fi/Pkpoutan*, 26 de septiembre de 2012. Disponible en:

<http://blogs.helsinki.fi/pkpoutan/social-sciences-going-for-big-data/> (consultado el 14 de febrero de 2014).

Preis, T. *et al.* (2012). “Quantifying the Advantage of Looking Forward”. *Nature, Scientific Reports*, 2, doi: 10.1038/srep00350.

Disponible en:

<http://www.nature.com/srep/2012/120405/srep00350/full/srep00350.html> (consultado el 14 de febrero de 2014).

- Preis, T. *et al.* (2013). “Quantifying Trading Behavior in Financial Markets Using Google Trends”. *Nature, Scientific Reports*, 3. doi:10.1038/srep01684. Disponible en: <http://www.nature.com/srep/2013/130425/srep01684/full/srep01684.html> (consultado el 14 de febrero de 2014).
- Reig, D. (2010). “Vídeo con subtítulos en español: Metaweb y otros sobre la web semántica”. *El Caparazón*, 24 de julio de 2010. Disponible en: <http://www.dreig.eu/caparazon/2010/07/24/video-con-subtitulos-en-espanol-metaweb-y-otros-sobre-la-web-semantica/> (consultado el 1 de diciembre de 2013).
- Reig, D. (2013). “Las analíticas de aprendizaje como Macrotendencia”. *El Caparazón*, 24 de mayo de 2013. Disponible en: <http://www.dreig.eu/caparazon/2013/05/24/analiticas-de-aprendizaje-learnovationda/> (consultado el 14 de febrero de 2014).
- Romero, C. & Ventura, S. (2013). “Data mining in education”. *WIRES Data Mining and Knowledge Discovery*, 3(1): 12-27. doi: 10.1002/widm.1075
- Siegel, E. (2013). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Solís, B. (2012). “The Human Algorithm: Redefining the Value of Data”. *Briansolis.com*, 11 de diciembre de 2012. Disponible en: <http://www.briansolis.com/2012/12/the-human-algorithm-redefining-the-value-of-data/> (consultado el 1 de diciembre de 2013).
- Suta, C.M. (2013). “Los beneficios del Big Data aplicados a Redes Sociales”. *Vuelo digital*, 19 de septiembre de 2013. Disponible en: <http://www.vuelodigital.com/2013/09/19/los-beneficios-del-big-data-aplicados-a-redes-sociales/> (consultado el 1 de diciembre de 2013).
- Tascón, M. (2013). “Introducción: Big Data. Pasado, presente y futuro”. *Telos*, 95: 47-50. Disponible en: [http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/NMEROSANTERIORES/pocaactual/DetalleAnteriores\\_95TELOS\\_DOSSIER0/seccion=1268&idioma=es\\_ES&id=2013062110090002&activo=6.do](http://sociedadinformacion.fundacion.telefonica.com/DYC/TELOS/NMEROSANTERIORES/pocaactual/DetalleAnteriores_95TELOS_DOSSIER0/seccion=1268&idioma=es_ES&id=2013062110090002&activo=6.do) (consultado el 30 de noviembre de 2013).

Thomas, J.J. & Cook, K.A. (Eds.) (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Press. Disponible en: [http://www.infovis-wiki.net/index.php/Visual\\_Analytics](http://www.infovis-wiki.net/index.php/Visual_Analytics) (consultado el 1 de diciembre de 2013).

Universität Passau (2013). “What if ... Open Data, Linked Data, and Big Data are needed at the same time?”. Coloquio con Erik Mannens, 16 de diciembre de 2013. Disponible en: [http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/graf/Mannens\\_16\\_Dez\\_13\\_Kosch\\_Kolloqu.pdf](http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/graf/Mannens_16_Dez_13_Kosch_Kolloqu.pdf) (consultado el 14 de febrero de 2014).

Wakefield, J. (2013). “Tomorrow's cities: How big data is changing the world”. *BBC News Technology*, 27 de agosto de 2013. Disponible en: <http://www.bbc.co.uk/news/technology-23253949> (consultado el 14 de febrero de 2014).