

Los big y open data¹ y la fascinación por los grandes números: aspectos éticos en su uso e interpretación

Jaime Vergara-Muñoz | arquitecto

Miguel Martínez-Monedero | Dpto. Proyectos Arquitectónicos, ETS de Arquitectura, Universidad de Granada

URL de la contribución <www.iaph.es/revistaph/index.php/revistaph/article/view/3942>

Estamos en la era del *big data*². La información que generamos en internet alcanza los 2,5 trillones de *bytes*. Esta cifra no para de crecer exponencialmente cada año y nos lleva hacia un “infinito” que a ciencia cierta no sabemos dónde acabará. Con un teléfono inteligente en casi todos los bolsillos, un ordenador y *tablet* en casi todos los hogares y un número cada vez mayor de dispositivos conectados a internet, el rastro tecnológico que vamos dejando en nuestro día a día aumenta rápidamente. La posibilidad de acceder a internet desde cualquier dispositivo ha permitido “democratizar” la información y que todos los individuos tengan el derecho a poder acceder a la misma cantidad y calidad de contenidos.

El análisis de estos datos es valioso para empresas y consumidores. Se comprueba cómo esta información guía el desarrollo de nuevos productos y servicios, puede predecir tendencias y moda, adapta mejor las demandas del ciudadano, y crea con todo ello nuevas oportunidades. Su análisis va más allá del ámbito comercial y está demostrando su validez para llegar incluso a entender mejor algo tan complejo como una ciudad o los movimientos migratorios.

Todos estos avances tecnológicos están produciendo enormes cambios en el seno de la actividad científica, de manera que la misma forma de acometer la investigación está cambiando desde que utiliza esta nueva fuente de conocimiento: “...cada vez hay más necesidad de comparar, preservar y gestionar grandes cantidades de datos” (VALENTÍN, 2013: 4).

No obstante, el uso del *big data*, y el más “social” *open data*, del mismo modo que favorece nuevas vías de conocimiento para la sociedad y la investigación, puede

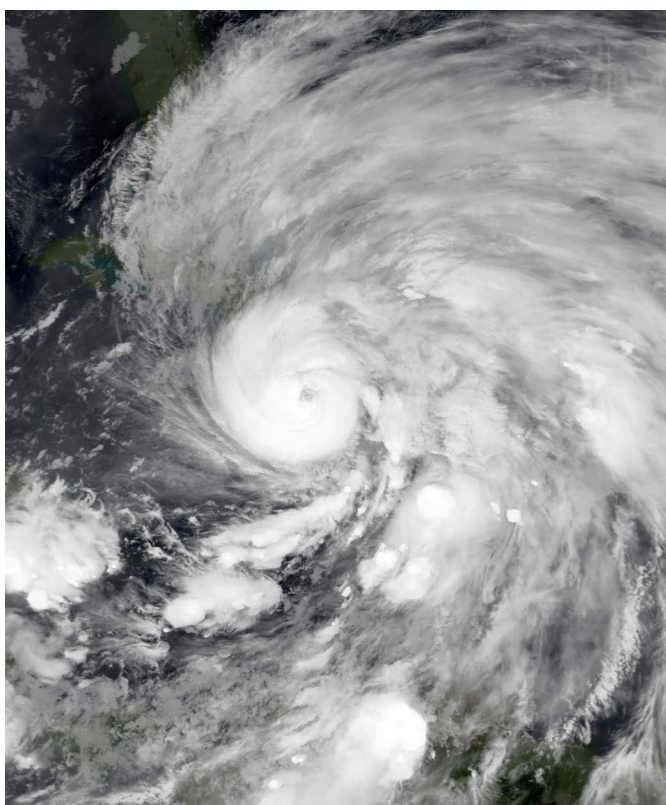
llegar a tener efectos perjudiciales cuando el manejo de estos datos no va ligado a una correcta interpretación. A continuación, se expone un claro ejemplo que nos puede hacer entender esta idea.

El huracán Sandy fue el más mortífero ciclón tropical de la temporada de huracanes de 2012. Afectó a numerosos países del mar del Caribe y dejó 253 víctimas mortales. Este hecho, entre el 27 de octubre y el 1 de noviembre de 2012, generó más de veinte millones de *tweets*. Si las organizaciones gubernamentales hubieran utilizado exclusivamente los datos obtenidos por esta fuente (Twitter) para determinar dónde se habían concentrado los daños más cuantiosos del huracán, con el objetivo de desplegar en ellos los servicios humanitarios, es muy posible que las personas que más lo necesitaran nunca los hubieran recibido. Muy pocos mensajes se enviaron en los lugares más afectados, que eran pequeñas poblaciones isleñas y costeras, sobre todo zonas donde los *smartphones* y el uso de Twitter eran menos habituales. En estos lugares, a medida que las reservas energéticas se agotaban, las baterías de los móviles y el uso de dispositivos con acceso a internet se fue limitando. Las zonas más gravemente afectadas tuvieron un índice de *tweets* cada vez menor. Se suponía que los datos reflejaban con exactitud las circunstancias sociales del momento, pero no se consideró que los más necesitados estaban en silencio tecnológico precisamente por la dureza de la tormenta tropical (RAMÍREZ; BRILL; OHLHAUSEN et ál., 2016: 27).

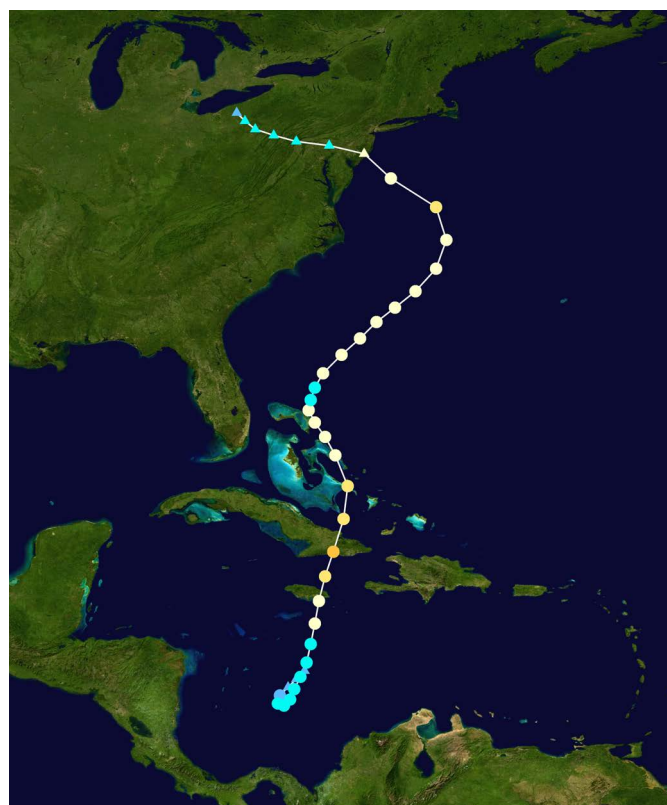
A tenor de lo anterior, la interpretación de los *big y open data* exige una nueva consideración. El uso de estos datos es un debate abierto, prueba de ello es la reciente Ley de 18/2015, por la que se modifica la anterior

El debate Datos abiertos en instituciones culturales

| coordina Isabel Medrano Corrales



Huracán Sandy a su paso por la costa del Caribe | foto NASA, MODIS/ LANCE, HDF File Data processed by Supportstorm [dominio público]



Recorrido del Huracán Sandy | foto Cyclonebiskit [dominio público]



La Guardia Nacional de Nueva Jersey se prepara para atender las zonas afectadas | foto Master Sgt. Mark C. Olsen [CC BY 2.0 o dominio público]

37/2007 sobre “reutilización de la información del sector público”.

El beneficio del *big data*, y la eliminación de las limitaciones para el acceso a sus datos (*open data*), favorece la multiplicidad de las investigaciones (MAYER, 2013: 200). Esta libertad lleva unida la responsabilidad en el uso y en la interpretación. Es necesario mantener el rigor y la veracidad como los valores éticos principales del investigador.

En este sentido, la competencia y la rivalidad (que dificultan la serenidad del juicio); las verdades parciales que los intereses personales, comerciales, empresariales o ideológicos puedan albergar; la carestía de presupuesto para continuar con una investigación (o peor aún, la dependencia de la continuación de una línea de inves-

tigación a un determinado resultado preconcebido); la rapidez con la que se exigen resultados, etc., son amenazas ciertas que pueden llegar a oscurecer el desarrollo de la investigación y la correcta interpretación de sus resultados.

Para minimizar los efectos negativos que el manejo de los grandes datos puede albergar y para favorecer su correcto uso, establecemos algunas pautas, criterios éticos de calidad, que ayudan a este fin. Estos son:

> Es necesaria la figura del catalogador de la información que proviene de estas fuentes. Debe ser cualificada, con una formación capaz de afrontar este desafío. El primer filtro, la primera selección y dirección de cómo y hacia dónde se deben manejar estos datos provendría de esta figura.

> Es necesario crear una estructura en la red, que consiga la interconexión de los datos que se obtienen a nivel nacional e internacional. De este modo no será un único campo de acción la tarea de estudio de estos datos, sino que se obtendrá una primera adecuación de los datos en función de su interés.

> Es necesario unificar o al menos compatibilizar los formatos que permita y facilite la interrelación entre repositorios, a nivel mundial. Es conocida la dificultad que se presenta cuando los datos son obtenidos de repositorios, almacenes digitales, nubes, etc., que son archivados con formatos no del todo compatibles; y en los que se produce, en su adaptación, errores o simplificaciones.

> Y finalmente, es necesario un compromiso ético global, emitido en primera instancia desde instituciones internacionales y refrendado posteriormente a nivel nacional con una legislación apropiada, y que permita, en definitiva, la correcta valoración de los resultados obtenidos en la investigación.

NOTAS

1. El concepto “datos abiertos” (*open data*, en inglés) es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen (<http://opendatahandbook.org/guide/es/introduction/>).

2. Denominamos grandes datos (*big data*, en inglés) a la gestión y análisis de enormes cantidades de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de *software* habitualmente utilizadas para la captura, gestión y procesamiento de datos.

BIBLIOGRAFÍA

- **MAYER SCHONGERGER, V.; CUKIER, K.** (2013) *Big Data. La revolución de los datos masivos*. Madrid: Turner, 2013
- **RAMÍREZ, E.; BRILL, J; OHLHAUSEN, M. et ál.** (2016) *Big Data: A Tool for Inclusion or Exclusion?* [en línea]. Washington: Federal Trade Commission, 2016 <<https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>> [Consulta: 6/06/2017]
- **VALENTÍN TORRECILLA, A.** (2013) *Gestión de datos de la investigación*. Tesina final de estudios Máster Oficial CALSI. Universidad Politécnica de Valencia, 2013
- **LEY 18/2015** por la que se modifica la Ley 37/2007 sobre reutilización de la información del sector público. *Boletín Oficial del Estado*, n.º 164, 10 de julio de 2015, pp. 57436-57450 <http://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-7731> [Consulta: 6/06/2017]