

Open Source y Open Data

Open Data o Datos Abiertos actualmente es un término común en los medios de comunicación, pero por supuesto donde es habitual es en ciencia sobre todo como parte de la denominada revolución de los datos y de la expansión de la Data Science o Ciencia de los Datos o las Humanidades Digitales. De hecho, Google^{1*} para la acepción inglesa de datos abiertos ofrece aproximadamente ciento treinta y tres millones de resultados. Además, hoy en día la denominación “datos abiertos” es habitual en los titulares periodísticos^{2*} donde se afirma por ejemplo que los datos abiertos salvan vidas o generan riqueza. Incluso existe^{3*} el día mundial de los datos abiertos. Asimismo, en 2018 se celebrará^{4*} la quinta Conferencia Internacional de Datos Abiertos, una conferencia bianual y que tiene en esta edición el lema ‘El Futuro está Abierto’. Todos estos elementos nos ayudan a entender la trascendencia de los datos abiertos en la actualidad.

Open Data, los datos abiertos, entronca con los movimientos de Open Source^{5*} o código abierto para software, Open Access* que se refiere al acceso gratuito a la información y al uso sin restricciones de los recursos digitales como textos y bases de datos, soportes de audio, vídeo o multimedia en general, así como el Open Knowledge*. La Open Knowledge International, se trata de una organización global sin ánimo de lucro que persigue mostrar el valor de los datos abiertos para la sociedad civil, ayudándola a acceder y a usar estos datos, con herramientas y habilidades técnicas, para afrontar problemas sociales. De esta manera, se libera una información que permitirá crear y compartir conocimiento. En realidad, estos movimientos^{6*} se retroalimentan los unos a los otros.

Veamos en detalle cuál es el significado del concepto ‘open data’. Data o datos, del latín datum significa ‘lo que se da’, una etimología muy acorde con el concepto de dato abierto. Como Data se suele entender el material en bruto producto de la abstracción del mundo en categorías, medidas u otras formas de representación como los números, caracteres, símbolos, imágenes, sonidos, bits, etc. que metafóricamente podrían ser los ladrillos que permiten construir/crear el conocimiento.^{7*} Los datos representan la naturaleza y por ejemplo permiten medir la edad de las personas, su altura, su peso, su opinión, sus hábitos, la temperatura, etc. Estos pueden estar en soporte físico o tangible o en formato analógico o digital. En verdad, existe un continuo incremento de la cantidad de datos sobre nuestro universo ya sean creados por las diferentes instancias gubernativas (estatal, regional o local), por instituciones de investigación, por el sector privado o por los individuos a través de sensores que se encuentran en millones de dispositivos, máquinas, vehículos, etc. que convenientemente analizados mediante

técnicas propias del Big Data ofrecen muchas posibilidades. De hecho, Big Data y Open Data también tienen sus concomitancias.

Los datos se definen como abiertos, según el Manual de Datos Abiertos^{8*} de la ya mencionada Open Knowledge International, cuando cualquier persona puede usarlos, reutilizarlos, redistribuirlos y cuando más se encuentran sujetos al requerimiento de atribución. Además, deben poderse compartir de la misma manera en que aparecen. Idealmente estos datos, aunque también puedan ser personales, pero completamente anonimizados, deberían estar disponibles en su totalidad y ser descargables desde Internet preferiblemente mediante un formato que sea machine readable y que se les puedan integrar otros conjuntos de datos buscando la interoperabilidad de los datos, lo cual tiene detrás la idea de linked data. Los datos abiertos podrán ser utilizados potencialmente por cualquier persona o grupo de personas en cualquier sector, ciencia, empresa, etc.

Muchas veces cuando se habla de abrir datos se habla generalmente de los datos generados por parte de las instancias del gobierno en el ejercicio de sus atribuciones como por ejemplo información fiscal, de empleo/desempleo, censal, escolar, demográfica, indicadores económicos, medición de las temperaturas, etc. con el fin de aumentar la transparencia, la participación ciudadana en las decisiones y/o la eficiencia del gobierno. Sin embargo, la creación masiva de datos desde instituciones dentro del ámbito de la investigación para sus proyectos ofrece un importante campo para la aplicación de la abertura de datos.^{9*} Muchos de los datos producidos en los proyectos de investigación pueden ser de interés para otras investigaciones que persiguen finalidades similares o distintas a los que originaron su creación. También pueden ser de utilidad para la industria o incluso para el ciudadano de a pie. De esta manera, una nueva utilización, es decir la reutilización de estos datos, supondría rentabilizar los costes de la investigación, lo que representaría un mayor retorno a la sociedad y permitiría la mejora de la calidad como consecuencia de su uso reiterado. Más específicamente la reutilización de estos datos para la ciencia [como se expuso hace unos años (2007) en el artículo Open Data for Global Science de Uhler, P.F. & Schröder, P. publicado en la revista Data Science Journal, 6 , pp . OD36–OD53 . DOI: <http://doi.org/10.2481/dsj.6.OD36>]:

- reforzaría la investigación comparativa.
- promovería nuevas investigaciones que podrían aplicar nuevos métodos de análisis o métodos alternativos.
- permitiría la verificación de los resultados previos o la exploración de temas no previstos por los investigadores/as iniciales facilitando estudios interdisciplinarios, inter-institucionales y por supuesto internacionales.
- ayudaría a la formación de nuevos investigadores.

- posibilitaría la creación de nuevos datos mediante el cruce de datos producidos en diferentes investigaciones aumentando la representatividad de los datos y sin duda evitaría la duplicación de costosas operaciones de recogida de datos.

En definitiva, con estos datos abiertos se podría seguir contribuyendo al progreso científico y a la innovación.

Algunas áreas científicas ya hace años que siguen esta estela en el acceso abierto de datos, un caso paradigmático es el campo de la Demografía con la creación de múltiples infraestructuras de investigación a partir de la creación de bases de datos de censos de población o de encuestas sociales por ejemplo de salud. Véase el caso de IPUMS, una de las infraestructuras de investigación del Minnesota Population Center de la Universidad de Minnesota en Estados Unidos y en concreto el caso de IPUMS internacional^{10*} que integra los datos censales individuales de 82 países desde 1960 hasta la actualidad. Para su utilización no es necesario ningún registro, esto sí, debe tenerse cierta formación en el manejo de datos. Desde la Demografía Histórica existen también múltiples ejemplos de la creación de bases de datos individuales por parte de investigadores particulares, lo que se podría denominar *small data*, que muchas veces con el tiempo se han convertido en infraestructuras de investigación ya que han ido aumentando en volumen gracias al trabajo continuado en sus instituciones de origen^{11*} como la Demographic Data Base in Umeå (Sweden), the Utah Genealogical Database (United States), the Scania Database in Lund (Sweden) o la base de datos del and the database of parish records in the Programme de recherche en démographie historique of the Université de Montréal (Canada) para citar algunos ejemplos.

De esta manera, la Demografía Histórica y su necesidad de disponer de datos para sus investigaciones resultan ser un buen laboratorio para la creación de datos abiertos con la participación de ciudadanos voluntarios a partir del paradigma de *crowdsourcing*, los cuales ayudan a la producción de los datos en formato digital para que luego toda la ciudadanía, donde también se incluye a los investigadores, pueda disfrutar de estos datos de manera directa o mediante visualizaciones creadas *ad hoc* a partir de mapas, pirámides de población o buscadores onomásticos. Es decir, se crea conocimiento a través de unos datos creados por la ciudadanía, para la ciudadanía. En este sentido, puede verse el proyecto^{12*} *Redes: Tecnología e innovación ciudadana en la construcción de redes sociales históricas para la comprensión del legado demográfico*, financiado por el programa Recercaixa en su modalidad de Humanidades Digitales y desarrollado por el Centro de Visión por Computador y el Centro de Estudios Demográficos de la Universidad Autònoma de Barcelona.

Ya para terminar esta breve introducción, mencionar las críticas que recibe este movimiento de *open data*. El hecho de que cualquier persona tenga el potencial para

usar estos datos es ciertamente una falacia ya que no todo el mundo dispone de la infraestructura digital necesaria, del hardware o el software imprescindible o de la formación indispensable que permita manejar datos. Como consecuencia de lo anterior, algunos autores se han llegado a preguntar si este tipo de iniciativas abiertas no estarán empoderando más a los más poderosos y de hecho ya se han visto casos. La otra gran diatriba a los datos abiertos es como se financia su producción y sobre todo su mantenimiento, los cuales por definición no tienen que tener costes para el usuario final.

Bibliografia de referencia

Anglada i de Ferrer, Lluís M., and Ernest Abadal. "¿Qué es la ciencia abierta?." Anuario ThinkEPI, 2018, vol. 12, p. 292-298(2018).

Gurstein, Michael B. "Open data: Empowering the empowered or effective data use for everyone?." First Monday 16.2 (2011).

Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. "Benefits, adoption barriers and myths of open data and open government." Information systems management 29.4 (2012): 258-268.

Kitchin, Rob. The data revolution: Big data, open data, data infrastructures and their consequences. Sage, 2014.

Mancebo, M. F. P., Benavent, R. A., Gonzalvo, S. D., & Sapena, A. F. (2017). Experiencias y lecciones aprendidas en datos abiertos en investigación. El grupo Datasea. PH: Boletín del Instituto Andaluz del Patrimonio Histórico, 25(92), 224-225

Peset, F., Aleixandre-Benavent, R., Blasco-Gil, Y., & Ferrer-Sapena, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. In Anales de Documentación (Vol. 20, No. 1).

Ruijter, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. Government Information Quarterly, 34(1), 45-52.

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. Information Polity, 22(1), 1-24.

Safarov, Iqbal, Albert Meijer, and Stephan Grimmelikhuijsen. "Utilization of open government data: A systematic literature review of types, conditions, effects and users." Information Polity 22.1 (2017): 1-24.

Páginas electrónicas:

[-http://progcity.maynoothuniversity.ie/2013/11/four-critiques-of-open-data-initiatives/](http://progcity.maynoothuniversity.ie/2013/11/four-critiques-of-open-data-initiatives/)

[- https://www.ted.com/playlists/495/how_data_can_save_lives](https://www.ted.com/playlists/495/how_data_can_save_lives)

[- http://opendatahandbook.org/](http://opendatahandbook.org/)

[- https://www.youtube.com/watch?v=6cRtbA_d4RI](https://www.youtube.com/watch?v=6cRtbA_d4RI)

Autora: Joana Maria Pujadas-Mora
Centre d'Estudis Demogràfics. UAB

