

Metodología de las ciencias sociales en la era digital: reflexiones sobre *Big data* para científicos sociales

M22

Dra. Schettini Patricia. LIMSyC, FTS, UNLP. patricia.schettini@gmail.com

Lic. Cuenca Adriana. ISTSyS, FTS, UNLP. cuencadriana59@gmail.com

ET4

INTRODUCCIÓN

La pandemia del Covid-19 aceleró procesos sociales que hubiesen demorado en impactar, o que ni siquiera hubiesen impactado de la forma y magnitud en que los vivimos actualmente. Uno de esos fenómenos es el avance de la era digital sobre la producción de conocimiento en las ciencias sociales.

En un artículo anterior (Cuenca y Schettini, 2020) nos preguntábamos sobre los efectos de la pandemia en la metodología de las ciencias sociales y cómo el aislamiento social, preventivo y obligatorio (ASPO) había modificado las actividades de investigación. En esta oportunidad, profundizaremos esas reflexiones en el convencimiento de que la pandemia aceleró un proceso que venía gestándose a partir de lo que denominamos la era digital o revolución digital.

La tecnología digital, permitió la masificación del uso de computadoras, de teléfonos celulares digitales y uso de Internet, que provocaron una revolución en la comunicación y en la información. Estos cambios produjeron importantes efectos socioeconómicos aumentando y acelerando la interconectividad y haciendo la comunicación más fácil. Pero también ocasionaron efectos no deseados como la sobrecarga de información (un volumen tal de información incapaz de ser procesada por la mente humana), la utilización en conspiraciones o espías (como el caso Snowden o la caída de la URSS) sin entrar en los temas de modificación de los empleos, todo un mundo para reflexionar que escapa al objetivo de este artículo.

Cuando hablamos de una revolución en la *información* y en la *producción de datos* necesariamente nos referimos a dos componentes importantes de la metodología de investigación en ciencias sociales. Como profesoras de Investigación Social entendemos que no podemos permanecer ajenas a este debate, es por lo que proponemos este recorrido que es teórico a modo de ensayo pero que se basa en lo que enseñamos. La pregunta que nos hacemos es ¿cómo impacta esta revolución en la metodología de las ciencias sociales? ¿Qué oportunidades pero también que limitaciones presenta este nuevo escenario en la investigación social?

Para esto comenzamos por caracterizar la era digital y en ese contexto la importancia de la emergencia de un dispositivo como *Big Data* que se utiliza en la investigación en general y en la investigación social, en particular; como una forma particular de producción, manejo y almacenamiento de grandes cantidades de datos. En este sentido, expresaremos a modo de reflexiones sobre el estímulo que ésta innovación produce en las ciencias sociales, los retos y conflictos que acarrea su utilización en la producción de conocimiento.

CARACTERIZACIÓN SOBRE LA ERA DIGITAL

La era digital marcó el comienzo de la Era de la Información y la Comunicación, también llamada Revolución Digital o Tercera Revolución Industrial o Tercera Revolución Científico Tecnológica en referencia a aquel desarrollo tecnológico que se produce entre las décadas de 1950 y 1970, con la innovación de la computación que supuso cambios profundos en la tecnología de la comunicación solo comparables a la revolución agrícola del paleolítico o la Revolución Industrial (entre los SXVIII y S XIX). Así, la nueva revolución tecnológica y el *modo informacional de desarrollo* reestructuraron el capitalismo de la segunda mitad del S XX a manera de una redefinición del paradigma tecno-económico (Castells, 1989). Estas tres grandes revoluciones tecnológicas tienen una trascendencia histórica ya que han transformado las dimensiones fundamentales de la vida humana: el tiempo y el espacio. Castells (1989) denomina a este proceso revolucionario del desarrollo "*modo informacional*" que refiere a una serie de innovaciones científicas y tecnológicas que constituyeron un nuevo paradigma tecnológico: la microelectrónica desarrollada desde el *transistor* (1947), los circuitos integrados (1957) al microprocesador (1971) que han permitido el desarrollo de computadoras personales. Tanto el *hardware* como el *software* modificaron las telecomunicaciones que *fueron el vector clave para la difusión y máximo aprovechamiento de las nuevas tecnologías* (Castells, 1989: 36).

Así, el producto de estas tecnologías es la información, justo en esto éste paradigma difiere de los anteriores. Pero a igual que las revoluciones anteriores el impacto es más en los procesos que en el producto. Esta consecuencia es muy importante ya que la innovación está orientada a los procesamientos que son los que se incorporan a todas las esferas de la actividad humana. El hecho de que las nuevas tecnologías estén enfocadas hacia el procesamiento de la información acarrea consecuencias en la metodología de las ciencias sociales. Una de estas se reconoce en la incorporación del *Big Data* como la base de la ciencia de los datos.

QUE ES BIG DATA?

Cada vez que utilizamos una pantalla, cada vez que ingresamos a una página web, comparamos algo o buscamos una dirección, cada vez que damos un *like* o subimos una foto a algún medio social, etc.,; cada una estas y otras acciones en la web generan datos. Esa inmensa generación de datos es información sobre nuestro comportamiento de forma instantánea y de miles de millones de usuarios. Tan importante como estos datos es la forma en que los procesamos (Sosa Escudero, 2020).

Big Data significa siempre *macrodatos* o datos masivos que quiere decir datos a gran escala; refiere a la recopilación de información de diferentes fuentes en un mismo lugar, por lo que se necesitan aplicaciones informáticas para procesarlos y almacenarlos. Grandes negocios se anuncian en la *web* sobre empresas que proponen "*recoger datos y analizarlos*" y aseguran que "*no puede llevarte mucho tiempo*" para vender una forma automatizada de creación de informes de datos "*incluso los más complejos*".

Aquel trabajo artesanal que un investigador realizaba durante meses de tareas al tratar de encontrar patrones y categorizarlos en grandes cantidades de información organizados en enormes matrices, hoy un *software* especializado lo hace en una velocidad que es siempre cada vez mayor. Para los científicos sociales manejar grandes cantidades de datos no solo era trabajoso, resultaba muy caro en término de recursos humanos que debían colaborar en la tarea y esto es siempre recursos económicos.

De Mauro y otros (2016) instauran la idea de que *Big Data* representa los activos de información que existen en Internet y están caracterizados por *volumen, velocidad y variedad* tan altos que requieren una tecnología específica y métodos analíticos para su transformación. Además, algunas organizaciones le agregan una nueva V, *veracidad*, siempre muy discutible; pero lo interesante es que se vuelve a hablar de esto. Las V (volumen, variedad, velocidad y veracidad) definen y caracterizan este tipo de procesamiento de datos (Sosa Escudero, 2020).

Hay en el mercado muchas herramientas tecnológicas que en general se pueden agrupar en tres tipos:

1. Datos estructurados o almacenes de datos; aquellos que ya están en una matriz organizados.
2. Datos no estructurados como los PDF, documentos multimedias, correos electrónicos o documentos de texto. Son datos que no se pueden descargar dentro de una tabla.
3. Datos semiestructurados: datos que no se limitan a campos determinados pero contienen marcadores como las hojas de cálculo.

Todos estos datos están en Internet, los generamos constantemente (en WhatsApp; Facebook, LinkedIn, Twitter, correo electrónicos), el problema está en cómo realizar su captura. Una vez encontrados hay que transformarlo de forma de organizarlos en tablas; y finalmente, almacenarlos. Estas son las operaciones básicas (captura, transformación y almacenamiento) de cualquier aplicación de Big Data (Robles y otros, 2020).

Los usos son infinitos. Ha sido utilizado por la industria de medios, las empresas (marketing), por los gobiernos; siempre es para saber qué piensa un determinado público y permite aumentar la eficiencia del mensaje que puede ser el consumo de un producto (material o simbólico).

REFLEXIONES SOBRE ESTUDIOS EN LAS REDES SOCIALES

Claramente estamos frente a un fenómeno que tiene grandes posibilidades y presenta, por supuesto también, enormes limitaciones para las ciencias sociales.

Todo un mundo se abrió al debate y al pensamiento cuando nos enfrentábamos a la idea y constitución del *dato*. ¿Qué es un dato? ¿Cómo se genera? ¿Dónde y cómo se le otorga validez? Fueron siempre temas muy debatidos y largamente tratados en el análisis de la información en las ciencias sociales (Schettini y Cortazzo, 2015). Hoy ese mundo está tratado en lo que se define como la *ciencia de los datos* que

refiere al *conjunto de métodos y procesos basados en técnicas estadísticas y de inteligencia artificial destinados a obtener información relevante a partir de un conjunto de datos* (Robles *et al*, 2020: 23). Esos resultados son el producto de la combinación de tres saberes diferentes: la rama de la matemática que se denomina estadística, la informática y, finalmente, las ciencias sociales. Robles *et al* (2020) afirman que es difícil encontrar una persona con conocimientos en las tres disciplinas, por lo que el mundo de *Big Data* tiende a la formación de equipos interdisciplinarios; por cuanto los autores afirman que la labor de un profesional en ciencias sociales no es tanto manejar la herramienta sino debe ser capaz de conocer sus posibilidades e identificar las soluciones de eventuales problemas emergentes en el proceso.

Estamos convencidas que es una herramienta muy potente para tratar volúmenes muy importantes de datos cosa que la sociología viene haciendo desde sus comienzos. La estadística siempre funcionó muy bien en los estudios de las ciencias sociales. Pero hasta ahora la tarea de generar esa información se resolvía elaborando un instrumento eficaz y someter a un sujeto a los estímulos necesarios para que respondiera en el sentido esperado (encuesta y objetivos de la encuesta). Hoy esta tarea no sería necesaria, la *web* se transforma en una fuente inagotable de información de gran interés. La actividad humana estaría en la interpretación de este volumen de información que necesitará un estudio de contexto que enmarque correctamente las acciones.

Van Dijck (2016: 17-18) afirmaba que durante los últimos diez años (en referencia a los primeros años del S XXI) *las vidas profesionales y personales se han visto paulatinamente inundadas por los medios sociales* (en referencia a Facebook, YouTube, Wikipedia, y tantas otras) donde las familias comparten *contenidos expresivos y comunicativos, construyen sus carreras profesionales y disfrutan de vidas sociales online*. Esto es lo que decimos que la pandemia encontró este campo desarrollado lo que permitió (a una porción importante de la población) trasladar sus actividades a entornos virtuales. Es lo que Van Dijck (2016) denomina el camino que va de la comunicación en red a la sociabilidad por plataformas, parte importante de *lo social* pasa por las plataformas y las relaciones sociales transformaron las *redes sociales* en digitales. Imprimiéndole al concepto solo esa acepción cuando deberíamos comprender que lo digital es solo un medio en el que se expresan las redes sociales (Schettini, 2019).

CONCLUSIONES

Entendemos que la nueva *ciencia de los datos*, apoyada en disciplinas como la estadística, la informática, el diseño influyen en la vida cotidiana modificando desde la política y la sociología hasta la medicina o la física, desde la empresa hasta el Estado (Sosa Escudero, 2020) y, en especial en las ciencias sociales, han encontrado en la tecnología digital nuevos modos de proceder en investigación. No es nuevo decir que las modas tienen gran influencia en el quehacer académico. Pero esto no puede hacerse abandonando los principios elementales de la investigación social que implica investigar con sujetos y no sobre sujetos.

No creemos que los grandes conjuntos de datos –Big Data- ofrezcan una forma superior de inteligencia y conocimiento. Ni que garanticen objetividad y precisión en los resultados como espíritus optimistas anuncian. Creíamos que estos debates habían sido superados. Big Data reeditó un giro conservador en el debate sobre la *verdad objetiva* y esto lo combina con los serios problemas éticos que devienen de la falta de autorizaciones y el filtrado automático de los datos.

Confundir comunicación con conectividad es un problema recurrente en el que podemos caer cuando se automatizan y se codifican en conceptos algorítmicos.

En el mundo offline, suele entenderse que las personas que están “bien conectadas” son aquellas cuyas relaciones resultan valiosas en virtud de su cualidad y condición, no de su cantidad. En el contexto de los medios sociales, el término “amigos” ha llegado a designar tanto vínculos fuertes *como* débiles, contactos íntimos como completos desconocidos (Van Dijck, 2016: 31).

Todos los autores que aquí se citan que han escrito sobre el tema preanuncian una realidad que vino para quedarse y que *ya invadió nuestras vidas* como afirma Sosa Escudero (2020) desde su título; hay que ver cuán profundas son estas transformaciones que calan en las ciencias sociales. Aún están por producirse.

BIBLIOGRAFÍA

- Castells, M (1989) La ciudad Informacional. Tecnologías de la información, reestructuración económica y el proceso urbano-regional. Madrid: Alianza Editorial.
- Cuenca, A. y Schettini P. (2020) Los efectos de la pandemia sobre la metodología de las ciencias sociales. Revista Escenarios N° 32; La Plata: Revista de la Facultad de Trabajo Social de la Universidad Nacional de La Plata. Disponible en: URL: <http://portal.amelica.org/ameli/jatsRepo/184/1841434023/index.html>
- De Mauro, G.; Grimaldi, A.; Marco, M. (2016). *A Formal definition of Big Data based on its essential Features*. Emerald Group Publishing.
- Gindin, I.L. y Busso M.P. (2018) Investigaciones en comunicación en tiempos de Big Data: sobre metodologías y temporalidades en el abordaje de redes sociales. adComunica. Revista de Estrategias, Tendencias e Innovación en Comunicación, nº15. Castellon: Asociación para el Desarrollo de la Comunicación adComunica y Universitat Jaume I, 25-43. DOI: <http://dx.doi.org/10.6035/2174-0992.2018.15.3>
- Robles, J.M.; Rodríguez T; Caballero R. y Gómez D. (2020) Big Data para científicos sociales. Una introducción. Madrid: Centro de Investigaciones Sociológicas (CIS).
- Schettini P. y Cortazzo I. (2015) Análisis de datos cualitativos en la investigación social. Procedimientos y herramientas para la interpretación de información cualitativa. La Plata: SEDICI Colección Libros de Cátedra. Universidad Nacional de La Plata. Disponible en: <https://doi.org/10.35537/10915/49017>
- [Schettini, P \(2019\)](#) Redes, acciones colectivas y ciclos de protesta. Un estudio de caso en un barrio del Gran Buenos Aires (1988-2015). Tesis doctoral Facultad de Ciencias Sociales, Universidad de Buenos Aires.
- Sosa Escudero, W (2020) Big data. Breve manual para conocer la ciencia de datos que ya invadió nuestras vidas. Buenos Aires: Siglo XXI.
- Van Dijck, J (2016) La cultura de la conectividad. Una historia crítica de las redes sociales. Buenos Aires: Siglo Veintiuno.