



# La planeación estratégica de la investigación social en escenarios no convencionales

Rafael A. González<sup>1</sup>

En esta ponencia voy a presentar varias herramientas y métodos que se han ido desarrollando en los últimos años, los cuales pueden tener un impacto particular para la investigación en ciencias sociales y humanidades, pero que no son de exclusividad de este tipo de disciplinas. En ese sentido, voy a hacer énfasis, más bien, en que el trabajo debe ser interdisciplinario, a partir de las tendencias en investigación social relacionadas con la noción de individuo cuantificado. De acuerdo con tales tendencias, expondré las nuevas técnicas emergentes para el análisis de textos, de datos cualitativos, desde el análisis computacional, y cómo estas han conducido a una reformulación de la investigación; es decir, cómo han constituido el paso de la investigación orientada por la teoría a la investigación dirigida por datos. Aquí examinaremos especialmente qué oportunidades brindan estas nuevas tendencias a la investigación social.

La transformación del objeto de estudio o de aquellas fuentes de información que utilizamos es quizá el punto de partida fundamental, no solo en cuanto tendencia científica, sino, también, en términos de

<sup>1</sup> Ingeniero de Sistemas de la Universidad Javeriana, MSc en Ciencias de la Computación y Doctor en Ingeniería de Sistemas (*cum laude*, Premio Aart Bosman) de la Universidad de Delft en Holanda. Ha sido profesor e investigador en sistemas de información, pensamiento sistémico y gestión del conocimiento, además de haber sido consultor para los sectores público y privado. Sus intereses académicos giran en torno al desarrollo de sistemas de información desde un enfoque centrado en el diseño, centrado en problemas de coordinación, complejidad y de interrelación entre tecnología y sociedad. [ragonzalez@javeriana.edu.co](mailto:ragonzalez@javeriana.edu.co)

tendencia social. Esto significa que hay una transformación del individuo, que ahora pasa a ser ubicado en una moderación que lo convierte en perfil. Este perfil está montado a partir de técnicas recientes de ingeniería de datos, ingeniería de texto, aprendizaje de máquina, etc. En últimas, el perfil pasa a ser, sobre todo, predictivo. Un ejemplo puede ser el siguiente: Google tiene un servicio que, basado en todos los datos de la tarjeta de crédito (compras, transacciones, etc.) de una persona, puede predecir, con un 80 % de confianza, si esa persona se va a divorciar o no de su pareja.

Esto quiere decir que conocemos tanto de la persona que la podemos predecir, pero la predecimos justamente porque está completamente alejado de lo que el individuo es; únicamente está amarrado a lo que esa persona hace, y lo que esa persona hace en relación con máquinas (actividad en *chats* y redes sociales, compras, etc.).

Al respecto, yo preguntaría: ¿será que es únicamente en relación con las máquinas? Es decir, cuando alguien utiliza su teléfono para mandar un mensaje por WhatsApp está comunicándose a través de una máquina, está interactuando con una máquina, estrictamente hablando. Pero en la práctica le está enviando mensajes a la novia, a la mamá, al hermano o al amigo; es decir, siguen siendo interacciones fundamentalmente sociales. Por tanto, mal haríamos en verlas como interacciones únicamente entre humanos y máquinas. Lo mismo ocurriría con la idea de percibir las netamente como datos cuantitativos. Quizá en el caso de Google se puede decir que es estrictamente cuantitativo, pero en la generación de contenidos web estaríamos hablando de textos, de videos, de imágenes; es decir, de todo un cúmulo de información que se procesa de manera muy distinta que la información cuantitativa.

De ahí la importancia del concepto de *quantified self*<sup>2</sup> o individuo cuantificado. Este posee un lado positivo y un lado oscuro. Si se mira desde el lado oscuro, el individuo cuantificado convierte a la persona en sus datos, hasta el punto de vaciar los datos de sentido; es decir, ya no están amarrados a la personalidad del individuo, sino que únicamente hablan por sí solos. El filósofo coreano Byung-Chul Han, en su obra *Psicopolítica* (2014), denomina este fenómeno como el *dataísmo* o la barbarie digital; esto es, una nueva forma de sociedad en la que los datos cuantificados son los que determinan lo que constituye a la sociedad en lugar de los individuos. Algunos dirán que este fenómeno se trata de un ejemplo claro del biopoder de Foucault (1976); sin embargo, a mi juicio, esto no es más que otra forma de aproximarse a la construcción de la subjetividad.

La principal razón para sostener mi punto y refutar la idea del biopoder es que la coacción no resulta ser fundamental en este tipo de prácticas. Por ejemplo, cuando uno entrega datos en internet, cuando uno comparte información por redes sociales, usualmente no es por obligación o por prohibición, sino por seducción. Esto se asemeja a lo que Byung-Chul Han llama *psicopolítica*; es decir, el control no opera a escala corporal, sino psíquica. Lo más importante es que esto ocurre de manera invisible; no hay alguien ejerciendo poder sobre la psique, es uno mismo ejerciendo el poder sobre uno mismo. Uno mismo es el que siente la necesidad de publicar cosas en internet, y es una seducción que está amarrada a las recompensas psicológicas que se asocian con dicha acción.

La noción de la *psicopolítica*, en este punto, se puede relacionar con el crecimiento de la llamada *ludificación*. La ludificación es básicamente usar lo que se conoce acerca de cómo funciona la adicción y adaptarlo al funcionamiento de las dinámicas de los sistemas informáticos, con el fin de hacerlos adictivos.

---

2 El concepto *quantified self* se relaciona también con el *life-logging*, *digital footprint* y *sousveillance*; el *individuo cuantificado* se asocia con el yo cuantificado que Gary Wolf y Kevin Kelly popularizaron en la revista *Wired*, en el 2007, y divulgaron posteriormente en los eventos de la organización TED (Tecnología, Entretenimiento, Diseño).

Estas tendencias empezaron con los videojuegos y ahora están en el diseño de las redes sociales. De este modo, van generando motivaciones intrínsecas, con el fin de que no sea un mandato externo el que nos obliga a publicar información, sino que la entreguemos de manera voluntaria. Lo que se pone en juego aquí, principalmente, es el valor afectivo del intercambio de informaciones y generaciones de contenidos. De ahí la tendencia en investigación que busca entender, desde la psicología personal y grupal, el funcionamiento y dinámicas de la interacción en redes sociales e internet.

Otra tendencia que está asociada con el *big data* y el *neuromarketing* es la de convertir todo en correlaciones. Convertir todo en correlaciones quiere decir que se le da principal importancia a los comportamientos estadísticos que ligan el hecho *A* con el hecho *B*, sin importar su origen y/o contexto. En este tipo de técnicas investigativas, lo único que importa es la cantidad de datos observables en su ordenamiento estadístico, con el fin de determinar los comportamientos de consumo de la gente, y, con base en ellos, generar modelos predictivos de dicho comportamiento. En últimas, esta tendencia de investigación no explica ni el porqué ni el cómo de los comportamientos colectivos.

Esta tendencia mayoritaria en la investigación ha generado un vacío por llenar; es decir, existe una necesidad por conocer los porqués y los cómo de estos comportamientos; una necesidad por establecer causalidades.

Este vacío por llenar está ligado con las tendencias sobre ciencia abierta. La propuesta de la ciencia abierta es comunicar, y publicar no solo el resultado de las investigaciones, al estilo de la ciencia clásica newtoniana, sino, especialmente, los procedimientos y los datos, con el fin de que cualquier investigador interesado en el tema pueda procesar los datos, y, especialmente, procesarlos desde múltiples fuentes, con distintos fines. En otros términos, lo que busca la ciencia abierta es generar una replicabilidad científica de los datos, con miras a generar una transparencia que permita a futuros investigadores tomar esos datos, examinarlos y profundizar en ellos en nuevas investigaciones.

Dicha noción de ciencia abierta —vale la pena aclarar— es apenas una tendencia emergente. En efecto, se han estado constituyendo tendencias que no se pueden llamar como tal ciencia abierta, pero sí *crowdsourcing*; esto ocurre cuando la participación es abierta, pero los datos son cerrados. Por ejemplo, un científico tiene un conjunto de datos climáticos recogidos y decide hacer un concurso, con el fin de invitar a sus colegas a encontrar relaciones entre sus variables; por tratarse de un concurso, la política es que solo las personas que participen tienen acceso a los datos, lo que quiere decir que los datos son cerrados, pues solo quienes participan tienen acceso a ellos.

Sin embargo, cuando se habla de *crowdscience* o ciencia abierta se debe tener en cuenta que esta posee distintas modalidades, en las que se les piden a los participantes distintas habilidades. En algunos casos, no se piden mayores requerimientos y habilidades especializadas a los participantes, no más que poseer un título de bachiller; normalmente, en este tipo de proyectos la participación de la gente es bastante mecánica. Por ejemplo, se le pide a la gente que revise una cantidad determinada de fotos y las dividan entre las que tienen personas y las que no.

En otros casos, en cambio, se puede necesitar una participación especializada; por ejemplo, que el investigador necesite personas que sepan leer una radiografía u otro tipo de imágenes de exámenes médicos, y que las clasifiquen con un conocimiento de esas imágenes diagnósticas. Así mismo, las necesidades acerca de las habilidades de las personas que se necesitan pueden seguir creciendo; continuando con el anterior ejemplo, el investigador no solo necesita que las personas puedan interpretar una radiografía, sino, especialmente, imágenes del cerebro.

Otro factor que varía de modalidad en modalidad es la metodología y las relaciones entre las tareas de los participantes. Es decir, cuando las tareas son independientes y altamente estructuradas, como en el caso de las fotos, el grado de coordinación entre participantes es casi nulo y el trabajo es simple. Sin embargo, hay otros casos en que la tarea es interdependiente, o sea, hay distintos tipos de tareas y puede que esas tareas sean poco estructuradas; en este tipo de casos se exige una participación mucho más activa de las personas involucradas y, como vemos, es en el área en que menos se ha logrado avanzar, en términos prácticos. De este modo, en la medida en que cambian las exigencias de las habilidades especializadas de los participantes, las herramientas de análisis de datos y la estructuración de los roles en el proceso de investigación, cambia la modalidad.

Tenemos el caso de *citizenscience*, que es una modalidad de *crowdscience*. Esta modalidad hace referencia a la elaboración de proyectos de investigación en los que el resultado de alguna manera tiene que estar enmarcado en una problemática social y en el que usualmente se necesita mayor alcance geográfico. Casos típicos de esta modalidad se dan, por ejemplo, cuando el investigador necesita monitorizar datos ambientales y no tiene los suficientes sensores para hacerlo; entonces, lo que él hace es ir a las comunidades y pedirles ayuda con la recolección de datos. Esto puede ser algo tan sencillo como que los vecinos de un río le tomen fotos periódicamente al río y se las envíen al investigador.

El *crowdfunding* es una modalidad de participación indirecta sobre los proyectos de investigación de cualquier tipo; normalmente esta modalidad funciona a partir de plataformas de financiación de proyectos. El problema de esta modalidad es que no ha logrado construir un nuevo tipo de financiación y de ciencia que no se base en la competitividad; es decir, una ciencia en la que los grupos de investigación o investigadores independientes no tengan que competir por los premios y becas en convocatorias de financiación. Esta dinámica de la ciencia clásica se fundamenta en una visión que claramente dice: “si gano yo, pierdes tú, pero no podemos ganar los dos”.

En esta nueva visión hay que tratar de entender cómo funcionarían los incentivos colaborativos en los que efectivamente “yo no tengo una competencia donde gano yo porque publico y pierdes tú porque no publicas”, sino que ganamos entre todos, y eso quiere decir que empezamos a ganar otras cosas distintas a solo publicaciones. Todo esto presenta grandes retos y el avistamiento de nuevas problemáticas. Por ejemplo, se han presentado las propuestas de generar métricas alternativas, en las que ya no solo importan cuántas publicaciones se tiene, sino, también, cuántos *retweets* o cuántos *likes* le han dado a la publicación. Otra problemática ya mencionada es la financiación: cómo construir nuevas estrategias de financiación. Por último, está la grave cuestión acerca de la calidad de los resultados en este tipo de ciencias. Claramente, hay muchas promesas, pero yo no les puedo garantizar que en este momento la ciencia sea más efectiva y de mayor calidad por esta vía.

Por otro lado, tenemos acceso a un sinnúmero de contribuyentes, a habilidades especializadas que tal vez localmente no están disponibles, a que la participación sea más diversa, a que se contribuya a la transparencia, al uso de plataformas tecnológicas, etc.; esto, en cuanto a la contribución de las plataformas tecnológicas, que tienen un potencial inmenso en términos de eficiencia y de motivación. En muchos casos la gente participa no porque le interese el tema de la investigación, sino porque a veces la plataforma es tan divertida que motiva a participar.

Por último, tenemos la tendencia de investigación, que estudia cómo este tipo de interacción social desde plataformas tecnológicas potencia, dentro de los esquemas de ciencia abierta, la colaboración. Lo que se ha encontrado es que, en lugar de desfavorecer la colaboración, la plataforma tecnológica, en cuanto intermediaria, hace que los científicos interactúen más.

Ahora bien, dejando un poco de lado el tema de las tendencias en investigación, pasemos ahora a las técnicas a las que están ligadas estas prácticas investigativas desde la ciencia abierta. Tal vez se pueda llegar a pensar que la ciencia abierta ha planteado también técnicas de investigación innovadoras, pero, en realidad, casi todas las técnicas utilizadas tienen décadas de existencia. Me refiero a técnicas de procesamiento de lenguaje natural, técnicas de aprendizaje de máquina, técnicas de ingeniería de texto. Quizá algunos se preguntarán: ¿qué ha cambiado, entonces? Lo que ha cambiado es que hay más datos, de más fácil acceso, en los cuales aplicar dichas técnicas, y, en adición, hay más herramientas de fácil acceso económico, con interfaces más sencillas de navegar, acompañadas de mayor capacidad de cómputo.

En términos de análisis de contenido, por ejemplo, se viene haciendo —tanto en ciencias sociales como naturales— el trabajo de codificación y toda la mecánica aburrida de analizar una entrevista, de analizar veinte años de noticias de periódico, etc.; de ahí que tareas como la de reconocimiento de entidades en un contexto como la cuantificación de todo lo que sean lugares, todo lo que sean personas, todo lo que sean momentos del día, se pueda realizar a través de *softwares* especializados, en un segundo.

En cuanto a la minería o ingeniería de textos, básicamente se trata de utilizar herramientas o técnicas de estadística pura y clásica, y mecanismos de inteligencia artificial. Se le llama *minería* principalmente porque lo que hace el programa es encontrar los diamantes entre el tierrero del texto. Es decir, continuando con la metáfora de la minería, el investigador le da las características del diamante y el programa lo busca automática o semiautomáticamente; eso quiere decir que puede haber una especie de supervisión, o sea que el investigador designa a un experto el trabajo de clasificar los textos de entrenamiento, para que distinga en cuáles hay diamantes y en cuáles no. La supervisión consiste en darle a la máquina unos criterios de búsqueda, de acuerdo con la clasificación previa de los textos. El problema principal de este tipo de programas es que predominan en inglés, por lo cual el trabajo de ingeniería lingüística en la lengua española tiene todavía mucho trabajo por hacer.

El reto más importante que suponen estas técnicas ligadas con estos *softwares* quizá sea que no hay un método, no existe una sola forma de hacer esto en una investigación; por tanto, se torna imprescindible combinar métodos y combinar disciplinas. Así mismo, invita a indagar y profundizar en las investigaciones, más allá de los diseños de modelos predictivos, a partir de la correlación de variables; es decir, se trata de empezar a preguntarnos por los porqués, las causas y los procesos, más que en el qué pasará.

En este sentido, el análisis de datos deja de estar guiado por modelos hipotético-deductivos y pasa a ser guiado por los datos mismos; es decir, por modelos inductivos y abductivos. El peligro de esto no reside en el cambio hacia la inducción, sino en el hecho de pensar que como los programas ya son capaces de organizar y sistematizar datos a partir de la inducción, estos procesos hacen de la investigación algo totalmente imparcial. En otros términos, no podemos olvidar que el que le da sentido a los datos procesados por la máquina y, en últimas, está comandando la máquina es el investigador. Por ello, el trabajo de pensar e ir más allá del procesamiento de los datos es aquel elemento básico que le agrega el investigador a la investigación.

Para terminar, les voy a exponer muy brevemente el trabajo que hemos estado desarrollando en algunos proyectos, relacionado con lo anterior.

El primer proyecto se adscribe a la red Patrimonio Histórico Iberoamericano (PHI)<sup>3</sup>. Dicha red está conformada por una serie de universidades iberoamericanas, entre ellas la Pontificia Universidad Javeriana.

---

3 Véase <https://phi.aq.upm.es/static/about.html>.



Digamos que la idea central es generar una especie de plataforma de ciencia abierta, pero con el patrimonio histórico; es decir, lo que se quiere compartir son los proyectos de intervención sobre el patrimonio histórico, la identificación de los métodos utilizados, el procedimiento de publicación de los resultados y la discusión de cómo difundir ese conocimiento.

En este último punto no solamente hemos desarrollado unas herramientas informáticas que les ayudan a los proyectos a evaluar su compatibilidad y su posible vinculación mutua, sino, también, a que las personas participen de una manera más efectiva, en un contexto sociomaterial. La idea de *sociomaterial* se refiere a que nosotros reconocemos que la interacción entre las personas no depende únicamente de los artefactos o aparatos que intermedian la experiencia, sino que depende principalmente de la *sociomaterialidad* en la que está inmersa una persona. En este sentido, la sociomaterialidad flexibiliza los medios y las dinámicas a partir de las cuales nos comunicamos.

Con el Centro Nacional de Memoria Histórica (CNMH)<sup>4</sup>, en Colombia, se está llevando a cabo un proyecto que consiste en hacer un análisis de las historias de las víctimas. Esto tiene como objetivo aproximarse a desarrollar un reconocimiento de identidades, fundamentalmente para aprovechar o para apoyar el proceso de discernimiento de las víctimas en un escenario de posconflicto. Nuestro trabajo consiste, por ejemplo —para que puedan hacerse una idea más clara—, en clasificar qué tipo de hechos relata tal o cual historia; es decir, se busca saber si en el relato de la víctima se está hablando de un asesinato selectivo o de una masacre. Claramente, en ambas circunstancias sabemos que alguien fue asesinado, pero la clave es construir los criterios para identificar cuáles son las condiciones que determinan que eso fue una masacre o que se dio la aniquilación de un individuo.

Finalmente, cierro con este proyecto que se está llevando a cabo en Cartago, con las bordadoras. Resulta que en Cartago hay un bordado único, basado en el calado. Nosotros decidimos entrar a la convocatoria de diálogo de saberes de Colciencias con este proyecto de apoyo a las bordadoras de Cartago, y fue aceptado. Lo que se busca es encontrar cómo en la multidisciplinariedad y en la multimetodología se puede identificar un rol de la tecnología en la colaboración de un intercambio de conocimientos entre otras personas. Lo más importante por recalcar de este proyecto es que los investigadores no llegamos a desempeñar el rol de expertos portadores de conocimiento, sino a aprender, a escuchar y a construir conjuntamente con las personas de la región; es decir, es un diálogo de saberes.

Esta dinámica de trabajo permitió que las mujeres tejedoras pasaran de tener un cierto recelo o temor hacia la tecnología, a tener una cierta pasión e incluso a interesarse por las posibilidades que la tecnología tendría, directa o indirectamente, sobre sus actividades de bordar. En últimas, pienso yo, esto es un ejemplo de cómo el trabajar activamente con las personas y entender sus problemas ayuda a construir nuevas posibilidades de vida para ellas.

## Referencias

- Byung, C. H. (2014). *Psicopolítica: neoliberalismo y nuevas técnicas de poder*. Barcelona: Herder.
- Foucault, M. (1976). *Histoire de la sexualité. Volumen I: La volonté du savoir*. Paris: Gallimard.

---

4 Véase <http://www.centrodememoriahistorica.gov.co/>.