

Métodos digitales: construcción de un panel online basado en web tracking

*Digital methods: construction of an online panel
based on web tracking*

*Métodos digitais: construção de um painel online
baseado em web tracking*

**Razón
y Palabra**

e-ISSN: 1605 -4806
VOL 24 N° 111 Mayo - Agosto 2021 Varia pp. 329 - 344
Recibido 12-05-2021 Aprobado 28-05-2021
<https://doi.org/10.26807/rp.v25i111.1800>

Cristian Ruiz

Portugal
Centre for Innovation, Technology and Policy Research (IN+)
cristianocruiz@gmail.com

Branco Di Fátima

Portugal
Centro de Investigação e Estudos de Sociologia (CIES-Iscite)
brancodifatima@gmail.com

Filipe Montargil

Portugal
Escola Superior de Comunicação Social (ESCS-IPL)
fmontargil@escs.ipl.pt

Resumen

El objetivo del presente estudio es examinar el método denominado panel online desde la perspectiva de las ciencias sociales, como un recurso creado para aportar al debate del desarrollo de métodos digitales. Para ello, se ha adoptado la metodología del estudio de caso, examinando la experiencia del *Living Lab on Media Content and Platforms* (LLMCP), un proyecto de investigación sobre el comportamiento en línea, donde los datos se recopilan a través de la tecnología de *web tracking*. La recopilación de información se basó en la observación del panel, la lectura del proyecto, el examen de sus sistemas de información y acceso privilegiado a su equipo. Los resultados permiten considerar el panel online como una técnica de recopilación de datos con potencial para el desarrollo de

las ciencias sociales, que puede proporcionar información más detallada sobre el comportamiento en línea.

Palabras clave: internet; métodos digitales; panel online; web tracking

Abstract

This paper aims to discuss the method called online panel, from the social science perspective, a resource created to contribute to the debate of digital methods. A case study was adopted, examining the experience of the *Living Lab on Media Content and Platforms (LLMCP)*, a research project on online behaviour, where data is collected through web tracking technology. The collection of information was based on the direct observation of the panel, the reading of the project, the examination of its information systems and privileged access to the research team. Results allow to consider the online panel as a data collection technique with potential for the development of social sciences and this technique can also provide detailed data to investigate online behavior.

Keywords: internet; digital methods; online panel; web tracking

Resumo

Este artigo pretende discutir o método denominado painel online na perspectiva das ciências sociais, um recurso criado para contribuir com o debate dos métodos digitais. Foi adotado um estudo de caso, examinando a experiência do *Living Lab on Media Content and Platforms (LLMCP)*, um projeto de investigação sobre o comportamento online, onde os dados são extraídos por *web tracking*. A recolha de informações baseou-se na observação direta do painel, na leitura do projeto, no exame dos seus sistemas de informação e no acesso privilegiado à equipa de pesquisa. Os resultados permitem considerar o painel online como uma técnica com potencial para o desenvolvimento das ciências sociais por fornecer dados mais detalhados do comportamento dos utilizadores de internet.

Palabras chave: internet; métodos digitais; painel online; web tracking

1. Introducción

Los estudios de internet son un área desafiante y muy fértil dentro las ciencias sociales. Estos también son relativamente nuevos, y, por lo tanto, la discusión sobre derivar metodologías eficientes y adecuadas siguen siendo un tema vigente. Si bien es cierto que existe una tendencia marcada por la continuación del uso de métodos más longevos como las encuestas (digitalizadas o presenciales), también es verdad que los últimos 20 años han presenciado una búsqueda activa de métodos que aprovechen las ventajas de la era computacional a la que pertenece la internet. Hay autores, quienes

incluso sugieren la idea de una crisis en la sociología empírica (Savage & Burrows, 2007, 2009) y, por lo tanto, se evidencia la necesidad de nuevos enfoques.

Dentro de estos esfuerzos podemos encontrar iniciativas dirigidas a incentivar la creación de métodos enteramente de origen digital. Por ejemplo, desde Jones (1999) y Kozinets (2010) hasta Rogers (2019) se han adelantado esfuerzos para la creación de métodos adecuados a la comunicación mediada por computadoras. Este último autor propone los métodos digitales, a saber, la idea de una metodología enteramente guiada por el objeto de mediación que sea nativamente digital y no por el usuario (Rogers, 2013, 2019).

Este artículo tiene como objetivo documentar y revisar la experiencia realizada por el proyecto de investigación portugués *Living Lab on Media Content and Platforms* (LLMCP), alrededor del método denominado panel online. El LLMCP se basa en el uso de herramientas de *web tracking* para la extracción y recopilación de datos proporcionados por un determinado grupo de participantes en dicho panel. El artículo centra primero su atención en la revisión de algunos métodos recurrentes en los estudios de internet, identificando cuándo hablamos de métodos nuevos o nativos digitales y cuándo estamos frente a métodos migrantes ya usados en el pasado dentro de las ciencias sociales. Del mismo modo se hará una revisión teórica de lo que es (o debería ser) un método digital plausible para los estudios de internet. Posteriormente, se revisará el estudio de caso del panel online dirigido por el LLMCP, para determinar si éste encaja dentro de la definición de un método digital y si éste es un método viable para estudiar el fenómeno de la internet.

Los resultados permiten considerar el panel online como una técnica de recopilación de datos con potencial para el desarrollo de las ciencias sociales. Esta técnica puede proporcionar información más detallada sobre el comportamiento en línea. Así, la idea de un panel online basado en *web tracking* fundamenta su lógica en seguir el objeto digital que produce el usuario (enlaces), y no al usuario en sí mismo. En este sentido, el panel online es en esencia un método digital.

2. ¿Qué es un método digital?

Las diferentes disciplinas legitiman sus teorías en la medida en que estas son capaces de derivar metodologías de sus postulados. No es de extrañar que una disciplina que se quiera referenciar así misma como conocimiento científico imprima suficientes energías en la cuestión del método. En la búsqueda de proponer y desarrollar nuevos acercamientos a los distintos objetos de estudio se tienen en cuenta las tecnologías existentes para poder construir los instrumentos necesarios (Bordens & Abbott, 2014). Es menester por tanto resaltar que, si en determinados periodos de la historia se han utilizado determinados métodos, a veces reduccionistas o limitados, no es por una fe ciega en dichos métodos e instrumentos, sino porque son la mejor versión que el método científico en ese momento tiene. Con la llegada de los computadores, y su capacidad de

almacenar, tratar y procesar cantidades antes inimaginables de información, una nueva era en la academia fue posible (Hewson, 2008).

Las ciencias sociales evidentemente no fueron la excepción y pronto varios académicos comenzaron a plantearse el uso de las computadoras como herramientas para estudiar los diferentes fenómenos sociales. Uno de estos fenómenos es provocado por las mismas computadoras; estamos hablando de la internet, un sistema de comunicación que junto a la WWW hicieron posible la conexión de gran parte del mundo en menos de 40 años. La relevancia que internet tiene para la sociedad, por lo menos en países de renta alta y media (Castells, 2001), nos obliga a estudiarla e incluirla dentro de los campos de estudio de las ciencias sociales (Fuchs, 2017). Así, no han sido pocos los esfuerzos para obtener métodos (Bruns & Burges, 2016; Rogers, 2019) que, de la mano de las nuevas tecnologías sean capaces de darnos un acercamiento más fidedigno al comportamiento de los seres humanos en la red.

Una de las primeras estrategias fue usar el potencial de las computadoras para procesar información proveniente de encuestas y entrevistas. Borgman (2015) hace énfasis en que las encuestas han sido usadas por los sociólogos desde al menos la década de 1920 como un método más maduro de investigación. Al ser llevada a los estudios de internet, hay una gran variedad de posibilidades para recolectar datos, ya sea en persona, vía correo tradicional o electrónico, desde una página web o por aplicación en dispositivos electrónicos, etc. Pero Borgman (2015) advierte que, según la herramienta de recopilación usada, las variables pueden afectar la captura del fenómeno social. Por ejemplo, usar una página web o aplicación móvil puede potenciar la capacidad de colectar y se puede llegar a más personas, pero puede perder precisión y calidad en las respuestas obtenidas. En contraste, una encuesta en persona suele ser de corte local, pero en compensación se puede llegar a obtener respuestas sólidas que acercan las prácticas reales frente a las representaciones en las respuestas de los entrevistados.

Muchas instituciones y centros ya sean públicas o privadas continúan recurriendo a metodologías de recopilación de datos por entrevista para estudiar a los usuarios de internet, como por ejemplo Kantar IBOPE Media (2021), Reuters Institute (2017), Pew Center (2016), Marktest (2015), Observatório da Comunicação (2014), Dubai School of Government (2012), etc. Sin embargo, a pesar de que se usan herramientas digitales como sitios web donde almacenar las encuestas, esta estrategia lo único que hace es migrar los métodos tradicionales, con sus aciertos y limitaciones (Evans & Mathur, 2018), al mundo digital. Lo que realmente no genera una nueva discusión, a pesar de las herramientas. Así, problemas longevos en ciencias sociales, como la distancia entre las prácticas reales y las representaciones deseadas plasmadas por el encuestado en los cuestionarios (LaPiere, 1934), permanecen sin ser superados.

Una alternativa, cada vez más sólida, es la propuesta hecha por Richard Rogers (2013, 2019), quien propone reconocer internet no solo como un objeto de estudio, sino también como una fuente de datos, métodos y técnica, todo al mismo tiempo. En este sentido, la idea es que la investigación sea guiada por el comportamiento que

queda registrado en el objeto nativamente digital utilizado por el usuario (botones de me gusta u otras reacciones, comentarios, enlaces de sitios web, etc.) y no por la percepción y opinión directa del usuario. Rogers (2013) identifica el enlace, la página web, el motor de búsqueda, la web o las webs (nacionales), las *wikis* y las plataformas de redes sociales como principales objetos enteramente digitales de internet. “Estas metodologías suponen la incorporación de nuevos conceptos, habilidades y técnicas de investigación” (Sued, 2019, p. 21).

Para conseguir extraer información de estos objetos, se tienen en cuenta dos técnicas, ambas provenientes de propuestas automatizadas computacionales. En primer lugar, el *scraping*, que consiste en extraer directamente de la portada de la página web los elementos que se quieren estudiar, a menudo usando lenguajes de programación con R y Python. En segundo lugar, el abordaje a través del Interfaz de Programación de Aplicaciones (API, por sus siglas en inglés) de la plataforma, como Twitter, Facebook o YouTube, y motores de búsqueda, como Ask, Google o Bing.

La API es, en principio, más exitosa y eficaz para extraer información relacionada al *big data*, pero también es una de las técnicas con mayores problemas de alcanzar una madurez estable, ya que está a merced de las políticas, voluntad y transformaciones de la compañía poseedora de los datos producidos (boyd & Crawford, 2012), tales como Google, Twitter o Facebook. Esto sin contar con factores externos, relacionados con la privacidad de los usuarios (Buchanan, 2011) o la protección a la propiedad intelectual (Collopy & Drye, 2017). Este último, por ejemplo, pone enormes barreras frente a lo que el investigador académico puede estudiar, pues si el autor adquiere datos a través del *scraping*, la información recopilada puede estar sujeta a leyes de propiedad de datos (Borgman, 2015).

Esto crea un desbalance y hasta dependencia de las políticas, términos y agenda de quienes ostentan los datos frente a quienes apenas pueden tener un limitado acceso a ellos (Bruns & Burgess, 2016). Un caso reciente que ejemplifica esta cuestión ocurrió en marzo del año 2018, cuando se dio a conocer a la opinión pública que la empresa *Cambridge Analytica* utilizaba los datos privados de alrededor de 85 millones de usuarios, sin consentimiento, para identificar personalidades de los votantes e influir en su comportamiento (Di Fátima, 2019). Estos datos fueron recolectados a través de una aplicación que se conectaba a la API de Facebook y que permitía la recopilación no solo del usuario, sino de su red personal en la plataforma. A consecuencia de este suceso de *Cambridge Analytica*, las API principalmente de Facebook y de Instagram fueron radicalmente reestructuradas, impidiendo o limitando la extracción de datos a partir de estas plataformas (Bruns, 2018). Las herramientas que hasta entonces servían para la extracción, como *Netwizz* en Facebook (Rieder, 2013) o *Visual Tagnet Explorer* en Instagram (Rieder, 2015), fueron afectadas directamente, limitando el trabajo de algunos investigadores e instituciones en los siguientes años.

De este modo podemos evidenciar los límites propios de los métodos digitales, ya que la comunidad académica que explora la apertura de nuevas metodologías tuvo

que someterse a estas medidas reduciendo su acceso al *big data*. Por otro lado, estas medidas no reestructuraron ni resolvieron las cuestiones éticas propias de las prácticas comerciales del análisis de datos (Bruns, 2018; Whitehead, 2007). Por el contrario, todo apunta a la espera de técnicas más agudas que serán tendencia en escenarios electorales y de mercado, como las *deepfakes* y las *fake news* (Franks & Waldman, 2019). A pesar de sus limitaciones, los métodos digitales consiguen abrir espacios a nuevos debates en torno a los distintos métodos en ciencias sociales (Sued, 2019). De este debate se desprende el siguiente estudio de caso, que se pregunta por la plausibilidad de un panel online de usuarios basado en *web tracking*, como aportación al debate de los métodos digitales.

Al mencionar la idea panel, hacemos referencia a la monitorización de un determinado público por un determinado tiempo. Una metodología que antiguamente era exclusiva de la radio o la televisión, a saber, la industria de medición de audiencias más tradicional. En palabras de Raymond Kent (1994, p. 10), “un panel es una muestra representativa de individuos, hogares y organizaciones que han acordado permitir el registro de sus actividades y opiniones con respecto a una gama de productos, servicios o comportamientos del uso de medios de forma continua o regular”, recopilando datos con instrumentos de cuestionario periódico (Hansen, 2008). Así, los resultados se basan en las respuestas a las preguntas planteadas por la investigación.

También existe la versión digitalizada del panel tradicional, con una muestra representativa de individuos monitorizados, en cuyo caso, se opta por usar encuestas a los usuarios de internet (Callegaro et al., 2014). Pero también existe la posibilidad de un panel apoyado por un *web tracking*, en internet. En esta modalidad, el panel online tiene una similitud con el procedimiento hecho para medir audiencias en los medios tradicionales. A saber, se convoca parte de la audiencia que se considere representativa del mercado, se instala un aparato capaz de registrar el consumo de los contenidos web para, posteriormente, recopilar, organizar y analizar los datos (Schmucker, 2011). Por lo tanto, los resultados se basan en la observación directa de los comportamientos en línea y no en respuestas a las preguntas de un cuestionario.

Un panel online basado en *web tracking* es un método poco explorado en el mundo académico, pero en contraste parece ser muy recurrente en el mundo industrial. Existe, por ejemplo, el modelo de negocio de las redes sociales, así como el de los buscadores. Cada movimiento hecho por el usuario dentro de las plataformas es monitorizado, compilado y analizado, a fin de modelar la caracterización de uso (Kumar & Tomkins, 2010) y mejorar la efectividad de relacionar publicidad con el público objetivo (Castellano & Folch, 2018).

Mucho de lo que esté conectado a internet implementa la misma lógica de captura de datos: el *web tracking*. Schmucker (2011, p. 1) explica que *web tracking* es “una tecnología que se utiliza para recopilar, almacenar y conectar registros de comportamiento de navegación del usuario”. Esta tecnología aparece en diferentes campos, pero se usa principalmente para crear anuncios personalizados, mejorar los servicios de seguridad cibernética y probar la usabilidad de las aplicaciones web (Schmucker, 2011).

Las estrategias metodológicas basadas en un panel online y *web tracking* no son exactamente nuevas. Hace muchos años que iniciativas como *Computer and Mobile Panel* (Nielsen), *Panel Network* (comScore) y *NetPanel* (Marktest) recopilan datos de uso de internet, como la página visitada o el tiempo de navegación. El LLMCP propone el uso de este método para contribuir a la creación de enfoques de los estudios de internet en las ciencias sociales. La novedad aquí es que lo hace para estudios académicos, desde el punto de vista de una institución educativa pública, y no con un objetivo de mercado. No se trata de ganancias, sino de generar conocimiento.

3. Metodología

Este estudio de caso del proyecto *Living Lab on Media Content and Platforms* (LLMCP) sigue las propuestas de Yin (2001) para un modelo exploratorio con un enfoque integrado: las limitaciones y los desafíos. El objetivo central es estudiar el panel online, planteándose la idea de que éste puede contribuir a la discusión de los métodos digitales. Como técnica cualitativa, este estudio de caso se rige por circunstancias específicas: el conocimiento acerca del panel online aún es limitado. Dentro de esta lógica, la elección de llevar a cabo un estudio exploratorio cumple con las circunstancias innovadoras del panel online, permitiendo la lectura en profundidad de experiencias atípicas en las ciencias sociales.

El LLMCP es un proyecto de investigación dirigido por la Escuela Superior de Comunicación Social (ESCS), del Instituto Politécnico de Lisboa (IPL), en Portugal. Su objetivo es estudiar los hábitos de uso de internet a través de un panel online. Según la información publicada en su página web, los datos se utilizan en tres áreas: i) estudios de comportamientos online, ii) estrategias de comunicación de instituciones públicas y iii) planificación de políticas de bienestar.

La recopilación de datos para el estudio de caso se basó en la observación directa del panel online del LLMCP, entre los meses de enero y junio de 2021, con cerca de 50 horas de trabajo de campo. Durante esos seis meses, los datos de las observaciones se registraron en un diario de campo (Yin, 2001). Estos datos se organizaron en un conjunto de categorías de análisis (motivaciones, tipos de información, reclutamientos, herramientas y sistemas), que tratan de responder las preguntas de investigación:

- ¿Cuáles son las motivaciones científicas que justifican la creación del proyecto?
- ¿Cómo se reclutan los participantes para el panel online y forman la muestra?
- ¿Qué herramientas informáticas se utilizan para recopilar datos de navegación?
- ¿Cuáles son los principales atributos de los sistemas de información del panel?
- ¿Qué tipo de información se recopila y cuáles son las características de las bases?

Además de recopilar información cualitativa del panel, es interesante entender cómo el equipo de LLMCP ha abordado los desafíos multidimensionales de los métodos

digitales (Rogers, 2013, 2019). Es igualmente interesante examinar, en contexto de *big data* (boyd & Crawford, 2012), el trabajo entre bastidores y la gestión de datos del proyecto. Por lo tanto, también se analizó el paso a paso del desarrollo del panel online, a lo largo de 2018 y 2019. Para este propósito, la recopilación de información se basó en la lectura del proyecto – financiado por la Fundação para a Ciência e a Tecnologia y Comisión Europea –, el examen de sus sistemas de información y acceso privilegiado a su equipo – ocho investigadores de la Escuela Superior de Comunicación Social (ESCS). En el último caso, las conversaciones espontáneas con los investigadores tuvieron lugar, en ambientes informales, durante el curso de seis meses. Como lo indica Yin (2001, p. 113), las conversaciones partieron de las preguntas de investigación y las respuestas se anotaron en el diario de campo.

Finalmente, en la interpretación de los resultados, se utilizó una lógica de triangulación de datos para responder las preguntas de investigación. La función principal de la triangulación es asegurarse de que el objeto de estudio se vea desde todos sus ángulos (Paranhos et al., 2016). Este estudio se justifica por la característica pionera del LLMCP, aún en una etapa temprana, pero buscando la posibilidad de reproducir métodos en otros contextos académicos (Morse, 2010). Además, es la primera vez que un centro de educación pública en Portugal ha avanzado en una investigación con panel online, donde los datos se recopilan a través del *web tracking* – tecnología común en el mundo de los negocios, pero todavía poco explorada por las ciencias sociales, como demuestra la literatura especializada – porque es caro, el trabajo de campo lleva mucho tiempo, se requiere conocimientos de informática específicos, etc.

4. Resultados y discusión

Una de las motivaciones principales del proyecto LLMCP es la intención de explorar formas de cerrar la brecha entre las prácticas y las representaciones (LaPiere, 1934). A saber, lo que los encuestados dicen que hacen, frente a cómo realmente se comportan. Tal problema puede llegar a estar también presente al estudiar el comportamiento de navegación de una muestra de usuarios, ya sea en formato presencial o virtual. Así, del mismo modo que la industria de medición de audiencias ha encontrado un escenario fértil para sus modelos de negocios en internet, esta propuesta del LLMCP pretende acercar métodos utilizados en la industria a necesidades de la academia. La Figura 1 muestra la página web del proyecto, donde es posible encontrar su formulario de participantes para el panel online.

Figura 1. Formulario web del panel online LLMCP

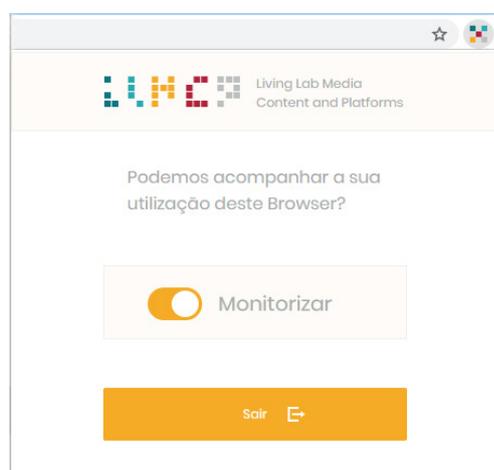


Fuente: Elaboración propia

A través de métodos computacionales, orientados originalmente a intereses de la industria, que siguen el registro del comportamiento del usuario y no sus interpretaciones de la realidad, este proyecto pretende contribuir en la construcción de nuevos enfoques de los estudios de internet. Es decir, la creación de un panel que permita registrar los comportamientos en línea para caracterizar al usuario como sujeto social y no solo como consumidor.

El registro de dichas actividades se realiza a través de una extensión para Google Chrome – el navegador más popular del mundo (Statcounter, 2021) – y los datos se recopilan en tiempo real a través de la tecnología de *web tracking* (Schmucker, 2011). La extensión debe ser instalada por cada usuario que voluntariamente quiera participar de la experiencia. Los participantes son reclutados, como en el caso industrial (Callegaro et al., 2014), con incentivos y con premiaciones, que, en el caso del panel estudiado, consiste en *power banks* al registrarse y tabletas rifadas en casos puntuales. La Figura 2 muestra la extensión del LLMCP, basada en la API de Google Chrome.

Figura 2. Extensión LLMCP para Chrome



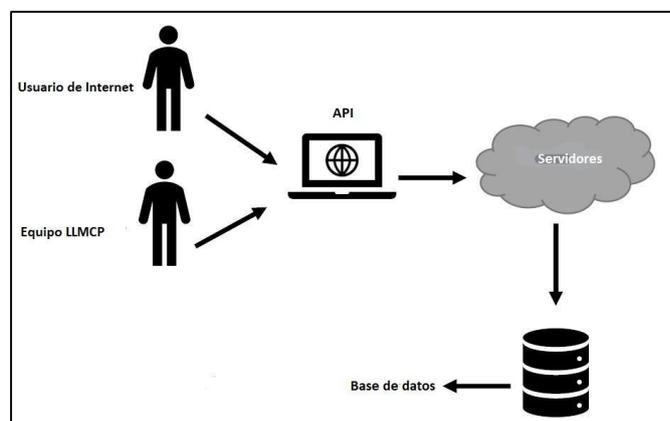
Fuente: elaboración propia

De esta forma, el candidato para el panel se registra en el formulario web con sus datos sociodemográficos (Figura 1). La finalización del registro crea una solicitud REST (Representational State Transfer) para el servidor del proyecto, que procesa la información y se comunica con una base de datos. Este sistema de información garantiza el almacenamiento de las acciones de navegación de los usuarios de la extensión LLMCP (Figura 2).

La extensión trabaja solicitando a la API de Chrome información del historial de navegación, con sus respectivos metadatos por enlace producido y completamente cargado. En otras palabras, la API concede permisos a aplicaciones externas al sistema de Chrome para que estos puedan acceder a partes específicas del total de datos que la compañía recopila a través del uso del navegador. Para dar un ejemplo técnico, a través de la extensión *chrome.history*¹, se pueden obtener datos de los enlaces registrados por el usuario al navegar, o con *chrome.runtime*², se pueden obtener datos relacionados con las ventanas abiertas por el usuario (hora de apertura de una pestaña, geolocalización, etc.).

Además de elementos demográficos, se obtienen el ID del usuario, la hora de entrada y hora de salida de una ventana de navegación, numeración secuencial de las ventanas abiertas, breve descripción de la actividad, enlace navegado, datos geográficos, entre otros. La Figura 3 revela el sistema de información del panel online, que vincula la solicitud REST con la base de datos. Este sistema también permite que el equipo LLMCP acceda a la plataforma a través de un *BackOffice*.

Figura 3. Sistema de información del panel LLMCP



Fuente: Elaboración propia

Siendo que esta extensión entró en operación en noviembre de 2018, en el mes de diciembre de 2019 ya contaba con 167 usuarios reclutados, de los cuales 93 habían registrado su uso por lo menos una vez (56.7%). Se reportaba una actividad de participación semanal en el panel online de entre 21 y 48 usuarios. La variación de

1 Chrome History: <https://developer.chrome.com/docs/extensions/reference/history/>

2 Chrome Runtime: <https://developer.chrome.com/docs/extensions/reference/runtime/>

usuarios se tiene en cuenta a partir de la semana siete, que es cuando la herramienta entra en operación oficial, no siendo necesariamente los mismos usuarios entre semana y semana. En este mismo período, los miembros del panel generaron 1 millón enlaces. En este punto es importante resaltar que los datos del usuario están anonimizados y protegidos, en Google Drive, bajo el *Reglamento General de Protección de Datos*, en vigor desde 2018 en la Unión Europea. La Figura 4 revela un extracto de las bases de datos LLMCP, así como la información recopilada por la extensión y disponible para los investigadores.

Figura 4. Extracto de las bases de datos del panel LLMCP

A	B	C	D	E	F	G	H	I	J	K	L
ID	Género	Nacimiento	Estado Civil	Condições de Trabalho	Dispositivo	Data abertura	Data fecho	Tab	Origem	Atividade	URL
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-27 11:59:23 AM	2018-11-27 05:21:27 PM	10		Diário de Notícias	https://www.dn.pt/
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-27 12:05:27 PM	2018-11-27 05:21:27 PM	13		Facebook	https://www.facebook.com/
8	MALE	1972	MARRIED	OTHER	LapTop	2018-11-27 01:06:39 PM	2018-11-27 01:06:44 PM	21		Netpa Pagina Entrada	https://portal.pt/ptescs/netpa/page
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-27 01:53:33 PM	2018-11-27 06:21:27 PM	42		Público	https://www.publico.pt/
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-27 02:48:41 PM	2018-11-27 02:48:52 PM	102		YouTube	https://www.youtube.com/?gl=BR&h=pt
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-27 03:37:52 PM	2018-11-27 03:37:59 PM	131		Yahoo	https://br.yahoo.com/
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-27 06:35:45 PM	2018-11-27 06:35:48 PM	303		Correio - Outlook	https://outlook.live.com/mail/inbox
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-27 06:35:48 PM	2018-11-27 06:35:51 PM	304		Correio - Outlook	https://outlook.live.com/mail/junkmail
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-27 08:36:01 PM	2018-12-10 03:56:22 PM	423		Facebook	https://www.facebook.com/
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-28 02:34:19 PM	N/A	611		Oriflame Cosméticos	https://pt.oriflame.com/
34	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-28 03:18:46 PM	N/A	833		A BOLA - Homepage	https://www.abola.pt/
34	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-28 03:43:49 PM	N/A	947		Startsurvey	https://surveys.mobrog.com/
8	MALE	1972	MARRIED	OTHER	LapTop	2018-11-28 04:11:42 PM	2018-11-28 04:11:49 PM	1067		NOVA Information Management School	http://www.novaims.unl.pt/
17	FEMALE	1998	SINGLE	STUDENT	LapTop	2018-11-28 04:41:38 PM	N/A	1184	1183	Beagle - Filhote, Preço, Personalidade	https://mescachorro.org/beagle/
34	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-30 12:07:19 PM	2018-11-30 12:07:25 PM	3026		Portal das Finanças	http://www.portaldasfinancas.gov.pt/pt/html/index.html
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-30 12:33:59 PM	2018-12-10 03:56:22 PM	3062		Sinonimos	https://www.sinonimos.com.br/
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-30 03:09:54 PM	2018-12-10 03:56:22 PM	3307		Diário de Notícias	https://www.dn.pt/
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-30 04:00:26 PM	2018-11-30 04:00:29 PM	3440		Yahoo	https://br.yahoo.com/
9	MALE	1983	SINGLE	STUDENT_WORKER	LapTop	2018-11-30 04:56:48 PM	2018-11-30 04:56:49 PM	3536		LLMCP BackOffice	https://www.prd.llmcp.pt/
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-30 05:14:03 PM	2018-11-30 05:18:22 PM	3607		MiPracy - Filmes e Series	https://www7.mspracy.xyz/serie.php?categoria=10
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-30 05:14:15 PM	2018-11-30 05:14:20 PM	3608	3607	MiPracy - Spartacus - Os Deuses da Arena	https://www7.mspracy.xyz/serie.php?mtdb=tt1758429&e=292
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-30 05:14:24 PM	2018-11-30 05:14:24 PM	3609	3607	MiPracy - Spartacus - Os Deuses da Arena	https://www7.mspracy.xyz/serie.php?mtdb=tt1758429&e=484
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-30 05:14:29 PM	2018-11-30 05:14:29 PM	3610	3607	MiPracy - Spartacus - Os Deuses da Arena	https://www7.mspracy.xyz/serie.php?mtdb=tt1758429&e=484
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-30 05:14:29 PM	2018-11-30 05:14:33 PM	3611	3607	MiPracy - Spartacus - Os Deuses da Arena	https://www7.mspracy.xyz/serie.php?mtdb=tt1758429&e=585
16	FEMALE	1999	SINGLE	STUDENT	LapTop	2018-11-30 05:14:33 PM	2018-11-30 05:14:48 PM	3612	3607	MiPracy - Spartacus - Os Deuses da Arena	https://www7.mspracy.xyz/serie.php?mtdb=tt1758429&e=586
46	FEMALE	1990	SINGLE	STUDENT	LapTop	2018-11-30 05:48:19 PM	2018-11-30 05:48:23 PM	3902		Escola Superior de Comunicação Social	https://www.esics.isg.pt/
46	FEMALE	1990	SINGLE	STUDENT	LapTop	2018-11-30 06:14:36 PM	N/A	3960		Gmail	https://mail.google.com/mail/u/0/
35	FEMALE	2000	SINGLE	STUDENT	LapTop	2018-12-01 03:01:22 AM	2019-03-19 10:28:20 AM	4243		Jornais e Revistas - SAPO 24	https://24.sapo.pt/jornais
14	FEMALE	1998	SINGLE	STUDENT	LapTop	2018-12-01 07:11:34 PM	N/A	4621		Google	https://www.google.com/

Fuente: Elaboración propia

Además, esta extensión permite el control del usuario frente a qué momentos de su navegación web quiere que queden registrados y cuáles no, con un simple clic en la aplicación (Figura 2). Quizá una de las dificultades relevantes en esta parte del diseño de la metodología pasa por conseguir que una población homogénea se mantenga activa durante un periodo de tiempo más largo. En este caso, el panel online contó con la ventaja de que la muestra contenía una alta población de estudiantes de la ESCS, lo que permitió dar uso a los datos para estudios aplicados a diversas preguntas de investigación.

Como toda propuesta que pretende ser innovadora en métodos digitales (Rogers, 2013) y *big data* (boyd & Crawford, 2012), las limitaciones y los desafíos se presentan como una parte inherente al momento de pasar de las ideas a la ejecución. A lo largo de los meses de experiencia de la creación del panel online, el equipo de LLMCP identificó una serie de limitaciones – por lo que hasta la fecha se entiende como la frontera que alcanza esta metodología, y desafíos – lo que parece tener una “solución más inmediata” para que el método sea más efectivo. Esta es una reflexión muy importante, precisamente porque el proyecto pretende, en un futuro cercano, permitir la “reproducción científica” del experimento (Morse, 2010). Es decir, permitir que otros centros de investigación utilicen el mismo método.

La extensión LLMCP trabaja con Chrome y solo en el computador *desktop* y *laptop*. Quedan fuera del estudio todas las acciones en otros navegadores web (Firefox, Edge,

Safari, etc.) o dispositivos (teléfonos inteligentes, tabletas, televisión inteligente). La información recopilada es solo de los períodos en que el usuario tiene la extensión activa, pudiendo suspender fácilmente la monitorización. En este caso, no hay forma de determinar en qué períodos el usuario mantuvo la herramienta suspendida. Para solucionar esto hay que invertir recursos tanto financieros como de talento técnico que posibilite la expansión a un universo multi pantalla, común en países más desarrollados (Castells, 2001).

Hasta ahora, el panel también tiene dificultades para registrar la navegación en las redes sociales. Hay líneas fantasmas que sitios como Instagram o Facebook producen en menos de un segundo, imposibilitando que la extensión lo registre adecuadamente en la base de datos. Todavía no queda claro el porqué de estas líneas en el funcionamiento de dichas redes, las mismas no parecen afectar significativamente una representación cercana a la navegación que la máquina registra del usuario. Muy de la mano del punto anterior, la aplicación solo registra líneas de sitios completamente cargados. El usuario puede estar consumiendo contenido en un enlace que aún no cargó en su totalidad y al poco tiempo cerrar la ventana de navegación. La carga del enlace depende de factores externos, comunes en trabajos con *big data* (boyd & Crawford, 2012), como las características del ordenador, la velocidad de internet o la configuración del navegador.

El análisis de las bases de datos del panel online (formato CSV para Excel) revela que hay enlaces que no corresponden a la línea donde están puestos y, la mayoría de las veces, aparece representando la línea anterior. De igual manera en la salida obtenida, en uno de los metadatos que recopila esa extensión, la numeración secuencial de las ventanas abiertas también presenta problemas de ordenamiento temporal para algunos sitios web.

Probablemente la limitación más relevante en esta primera fase es el error en la columna de cierre. La columna de cierre de una ventana con frecuencia no consigue registrar el momento de salida, por lo que genera como respuesta un N/A (Figura 4). Si esta limitación fuese superada, los estudios basados en tiempo serían más precisos, y por lo tanto más cercanos al fenómeno estudiado. Por ejemplo, sería posible calcular con precisión el tiempo que pasan los miembros del panel en diferentes contenidos en internet, caracterizando con mayor precisión el tiempo dedicado a cada acción de navegación (visitar redes sociales, leer noticias, consultar correos electrónicos, comprar en línea, etc.).

También se identificó otro desafío a la hora de interpretar los datos, ya que un usuario puede permanecer con más de una ventana abierta, estar o no consumiendo todas al mismo tiempo, y no tener las diferentes acciones identificadas: YouTube (escuchando música), Messenger (conversando con amigos) y Google Docs (haciendo trabajos escolares). La dificultad está en identificar cuándo un consumo de contenido es activo y cuándo no lo es. Los investigadores del proyecto LLMCP creen que con la observación de “tiempos sobrepuestos se pueda resolver esa pregunta”, aunque los primeros intentos, en septiembre de 2020, “no tuvieron resultados satisfactorios”. Aquí se necesitaría tener

mecanismos para diferenciar cuándo el ordenador realmente está siendo utilizado y cuándo no.

Para las siguientes fases, el LLMCP necesita simplificar su proceso de reclutamiento de miembros del panel online. El análisis del modelo utilizado muestra que el proceso es lento, con muchos pasos entre la selección del candidato, la instalación de la extensión y el monitoreo en sí. Esta iniciativa ayudaría a ampliar la muestra de usuarios de internet, principalmente porque el miedo principal del usuario está relacionado con la privacidad y el formato invasivo de las tecnologías de *web tracking* (Schmucker, 2011). El “temor se reflejan en el trabajo de reclutamiento para el panel”, dicen los investigadores. Por ejemplo, LLMCP invitó a 125 estudiantes a unirse al panel online a lo largo de 2019. Después de una presentación del estudio, se les preguntó sobre su disposición: 62.0% no estaba dispuesto a participar y, para 37.7% de ellos, la razón principal era la privacidad (Di Fátima, Montargil & Miranda, 2019).

5. Conclusiones

Este artículo tuvo como objetivo examinar el método denominado panel online, un recurso creado para apoyar el desarrollo de los métodos digitales. La metodología utilizada fue el caso de estudio del *Living Lab on Media Content and Platforms* (LLMCP), un proyecto de investigación sobre el comportamiento en línea, donde los datos se recopilan a través de la tecnología de *web tracking*.

El estudio del LLMCP indica que las potencialidades de una metodología de este tipo son variadas y pueden representar aportes a los estudios de internet, sobre todo porque se pretende apostar por la observación de los comportamientos en línea, y no en los métodos clásicos basados en encuestas. Así, la idea de un panel online fundamenta su lógica en seguir el objeto digital que produce el usuario (enlaces), y no al usuario en sí mismo. En este sentido, el panel online es en esencia un método digital.

Esto conlleva a que el panel online esté sujeto a las limitaciones discutidas previamente en este artículo. Por ejemplo, la construcción de una aplicación para monitorear el tráfico de los usuarios pasa enteramente por el uso de la API de los navegadores usados. Consecuentemente el panel estará subordinado a los cambios que se produzcan en las políticas e intereses de las organizaciones propietarias de dichos navegadores.

Mientras se intentan superar las limitaciones y los desafíos descritos, el método se encuentra en fase de expansión de sus potencialidades. Vale la pena comenzar a pensar sobre los datos que un método como éste puede ofrecer al investigador en las ciencias sociales y estudios de internet, buscando escenarios en donde el panel online pueda aportar nuevas perspectivas. Por ejemplo, explorar los comportamientos comparados entre poblaciones o comunidades en determinados periodos podrían arrojar evidencia de la información que es consumida en determinados eventos sociales (ejemplo, elecciones, manifestaciones, eventos deportivos, etc.).

Una limitación vista en el estudio de caso del LLMCP radica en el uso de apenas un

navegador (Google Chrome) para el desarrollo del *web tracking*, esto en parte por los altos costos que conlleva el desarrollo y mantenimiento de dichas aplicaciones. Un escenario ideal para el panel online supone interconectar el registro de actividades del usuario, no solo en el *browser*, sino también en diferentes aparatos al mismo tiempo, como los teléfonos inteligentes y tabletas. Esto puede representar un avance en estudios sobre multitarea y pantallas o incluso corroborar teorías sociales sobre las formas de usar la internet (Castells, 2001; Fuchs, 2017).

Finalmente, uno de los desafíos en este tipo de enfoque metodológico sigue siendo el reclutamiento y prolongación de la utilización de la extensión por los participantes. A menudo, los sujetos monitoreados tienden a abandonar el panel al cabo de un tiempo (desde días hasta meses). Aunque el panel del LLMCP llegó a 167 usuarios, la participación semanal variaba, en promedio, entre 21 y 48. Una preocupación aquí es sobre la validación de la muestra y la capacidad que el panel online tendría para generalizar sus resultados a una población más amplia. Diseñar planes de premiación para los participantes puede ayudar a resolver este desafío.

La vida en red requiere el desarrollo de enfoques innovadores basados en sistemas de información robustos. Sin embargo, para comprender sociedades complejas y altamente dinámicas, no es suficiente construir conjuntos de datos con cientos de miles de líneas de navegación (boyd & Crawford, 2012). Es necesario validar todo el proceso de recopilación, procesamiento y análisis de datos. Los resultados deben reflejar con precisión la diversidad y la complejidad de sus comunidades.

Como recomendación para futuros estudios, la experiencia aquí expuesta parece indicar la necesidad de simplificar el sistema de registro para los participantes. Además de los temores sobre privacidad y el formato invasivo de las tecnologías de *web tracking* (Schmucker, 2011), la experiencia LLMCP también revela la dificultad que enfrenta una parte de los voluntarios para completar la solicitud del panel online. Por un lado, la información puede denotar los niveles de alfabetización informática de parte de la muestra. Por otro, es una prueba de que los sistemas más sofisticados terminan construyendo barreras para un estrato del universo estudiado. De cualquier forma, los resultados permiten considerar el panel online como una técnica de recopilación de datos con potencial para el desarrollo de las ciencias sociales.

Referencias

- Bordens, K. S., & Abbott, B. B. (2014). *Research and design methods: A process approach*. McGraw-Hill.
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.
- boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Bruns, A. (2018, 25 de abril). *Facebook shuts the gate after the horse has bolted, and hurts real research in the process*. Internet Policy Review. <https://cutt.ly/Ln4Hm1b>
- Bruns, A., & Burgess, J. (2016). Methodological innovation in precarious spaces: The case of Twitter. In H. Snee, C. Hine, Y. Morey, S. Roberts & H. Watson (Eds.), *Digital methods for social science: An interdisciplinary guide to research innovation* (pp. 17–33). Palgrave Macmillan.

- Buchanan, E. A. (2011). Internet research ethics: Past, present, and future. In M. Consalvo & C. Ess (Eds.), *The handbook of Internet studies* (pp. 83-108). Wiley-Blackwell.
- Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online panel research: History, concepts, applications and a look at the future. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 1-22). John Wiley & Sons.
- Castellano, C. M., & Folch, C. S. (2018). La publicidad en la era de Google. *Trípodos*, 43(2018), 7–9. <https://cutt.ly/Hn75DtG>
- Castells, M. (2001). *La galaxia Internet: Reflexiones sobre Internet, empresa y sociedad*. Plaza & Janés.
- Collopy, D., & Drye, T. (2017, 4 de septiembre). *Share and share alike: The challenges from social media for intellectual property rights*. Intellectual Property Office. <https://cutt.ly/Wn4HONR>
- Di Fátima, B. (2019). *Dias de tormenta*. Geração Editorial.
- Di Fátima, B., Montargil, F., & Miranda, S. (2019). Estudando os comportamentos online: Premissas e desafios no desenvolvimento de um painel de utilizadores da internet. *Texto Livre: Linguagem e Tecnologia*, 12(1), 123–137. <https://doi.org/10.17851/1983-3652.12.1.123-137>
- Dubai School of Government. (2012). Social media in the Arab World: Influencing societal and cultural change? *Arab Social Media Report*, 2(1), 1-29.
- Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research*, 28(4), 854–887. <https://doi.org/10.1108/IntR-03-2018-0089>
- Franks, M. A., & Waldman, A. E. (2019). Sex, lies, and videotape: Deep fakes and free. *Maryland Law Review*, 78(4), 892–898. <https://cutt.ly/fn4ScMg>
- Fuchs, C. (2017). *Social media: A critical introduction*. SAGE.
- Hansen, J. (2008). Panel surveys. In W. Donsbach & M. Traugott (Eds.), *The Sage handbook of public opinion research* (pp. 330-339). SAGE.
- Hewson, C. M. (2008). Internet-mediated research as an emergent method and its potential role in facilitating mixed methods research. In S. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 543-570). Guilford Press.
- Jones, S. (1999). Studying the net: Intricacies and issues. In S. Jones (Ed.), *Doing internet research: Critical issues and methods for examining the net* (pp. 43-58). SAGE.
- Kantar IBOPE Media. (2021). *Data stories. Mobile first: O super-humano multitarefa*. Kantar IBOPE Media. <https://cutt.ly/yn4K6Oy>
- Kent, R. A. (1994). *Measuring media audience*. Routledge.
- Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. SAGE.
- Kumar, R., & Tomkins, A. (2010, 26-30 de Abril). *A characterization of online browsing behavior* [Paper presentation]. 19th International Conference on World Wide Web (WWW'10), Raleigh, NC, Estados Unidos. <https://doi.org/10.1145/1772690.1772748>
- LaPiere, R. T. (1934). Attitudes vs. Actions. *Social Forces*, 13(2), 230–237. <https://doi.org/10.2307/2570339>
- Markttest. (2015). *Um dia das nossas vidas na Internet: estudo de hábitos digitais dos portugueses*. Nova Expressão. <https://cutt.ly/jn78YWv>
- Morse, J. (2010). Procedures and practice of mixed method design: Maintaining, control, rigor and complexity. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research* (pp. 339-352). SAGE.
- Observatório da Comunicação. (2014). *A internet em Portugal: Sociedade em rede 2014*. Observatório da Comunicação. <https://cutt.ly/dn4J2eD>
- Paranhos, R., Filho, D., Rocha, E., Júnior, J., & Freitas, D. (2016). Uma introdução aos métodos mistos. *Sociologias*, 18(42), 384–411. <https://doi.org/10.1590/15174522-018004221>
- Pew Center. (2016). *Social media update 2016: Facebook usage and engagement is on the rise, while adoption of other platforms holds steady*. Pew Research Center. <https://cutt.ly/dn4JOtH>
- Reuters Institute. (2017). *Digital news report*, 6º Ed. University of Oxford. <https://cutt.ly/Ln4Klgn>
- Rieder, B. (2013, 2-4 de Mayo). *Studying Facebook via data extraction: The Netvizz application* [Paper presentation]. 5th Annual ACM Web Science Conference (WebSci'13), Paris, Île-de-France, Francia. <https://doi.org/10.1145/2464464.2464475>
- Rieder, B. (2015, 21 de septiembre). *Visual Tagnet Explorer: Version 1.1*. Github. <https://cutt.ly/9n4H7lp>
- Rogers, R. (2013). *Digital methods*. MIT Press.
- Rogers, R. (2019). *Doing digital methods*. SAGE.
- Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), 885–899. <https://doi.org/10.1177/0038038507080443>
- Savage, M., & Burrows, R. (2009). Some further reflections on the coming crisis of empirical sociology. *Sociology*, 43(4), 762–772. <https://doi.org/10.1177/0038038509105420>

- Schmucker, N. (2011). *Web tracking: SNET2 seminar paper*. Berlin University of Technology. <https://cutt.ly/on4Jsrb>
- Statcounter. (2021). *Desktop, tablet & console browser market share worldwide*. Statcounter GlobalStats. <https://cutt.ly/pn4IUUsN>
- Sued, G. (2019). Para una traducción de los métodos digitales a los estudios latinoamericanos de la comunicación. *Virtualis*, 10(19), 20–41. <https://cutt.ly/in74h1z>
- Whitehead, L. C. (2007). Methodological and ethical issues in internet-mediated research in the field of health: An integrated review of the literature. *Social Science & Medicine*, 65(4), 782–791. <https://doi.org/10.1016/j.socscimed.2007.03.005>
- Yin, R. K. (2001). *Estudo de caso: Planejamento e métodos*. Bookman.